Placing *Come* and *Go*: Locating the Lexical Item

Catherine Worlock Pope, BA, MA

Thesis submitted to the
University of Nottingham
for the degree of
Doctor of Philosophy

July 2015

In memory of David
and the promise I made him

Abstract

By examining language simultaneously along the paradigmatic and syntagmatic axes, Sinclair (2004a) identified the lexical item as an object of the discourse comprising an obligatory core and semantic prosody, and optional collocates, colligates and semantic preferences. This research investigates Sinclair's theoretical model by locating the lexical items that are associated with the complementary verbs *come* and *go* in the spoken and written discourses in a selection of the International Corpora of English (ICE). The corpora selected are ICE-Canada, -GB, -India and –Jamaica.

This research is innovative in that it adapts Sinclair's methodology to examine high frequency lexical items across different discourses and different World Englishes It establishes that there is a significantly greater difference in frequency of the lexical items associated with *come* and *go* within the different discourses of the ICE corpora in comparison to between the ICE corpora. It replaces the core with the node, it introduces structural preference and discourse preference as co-selection components of the lexical item, and it substitutes semantic force for the term semantic prosody as defined by Sinclair: the 'reason why [the item] is chosen' (Sinclair 2004a: 144). Thus the lexical item comprises an obligatory node and semantic force, and optional collocates, colligates, structural preferences, semantic preferences and discourse preferences.

As a consequence of these theoretical and methodological adaptations, this research shows that semantic forces with the associated co-selection components can function in tandem and that semantic forces, again with the associated co-selection components, can function in layers. The research concludes that the lexical item is not an identifiable object in the discourse, but it is the syntagmatic realisations of a paradigmatic choice.

Acknowledgements

There have been many wonderful people who have supported me over the past few years so it is with some reservation that I am singling out just a few to acknowledge here: Michael Hoey and Mike Scott for their inspirational lectures, their abounding enthusiasm both for their subject and for their students, without which I would never have thought I might undertake this research; Michaela Mahlberg for her help and guidance in getting started; and Michael Stubbs for his discussions on philosophy and his valuable advice, without which I might have abandoned this research along the way.

There are my supervisors, Svenja Adolphs and Ron Carter; who have continually given me constructive suggestions and positive encouragement. In addition, Ron has also guided me skilfully through the whole doctorate process and supported me when "life" has unfortunately interfered with my research. Without his kindness, his understanding, his patience, and his invaluable input I could not have completed this research. There is Lydia Wallman, in the School of English office, for always knowing the answer to my queries, and for always sorting out my problems. And there is Kevin Harvey, my internal examiner, for his positive approach, considered feedback and help through the final stages of my doctorate.

Of my friends I must single out Vincent Docherty for always reminding me that it is what one does not do rather than what was does in life that one invariably regrets; Rosemarie and Eric Hutchinson, for sharing their experiences with me; and Richard Thorne, without his backing, encouragement and practical support this research might never have been finished. Finally I must mention my children, Fredrik Fuglesang and Anna Lisa Fuglesang, who have always been there cheering me on.

Contents

Abstract **3**
Acknowledgements **4**
Chapter 1 The Introduction **8**
1.1 Research question(s) **9**
1.2 Paradigmatic and syntagmatic axes **10**
1.3 Lexical Bundles **11**
1.4 Concgrams **11**
1.5 Pattern Grammar **12**
1.6 Lexical Grammar **13**
1.7 Lexical Priming **14**
1.8 Linear Unit Grammar **15**
1.9 The Data **15**
In the chapters that follow …. **18**
Chapter 2 The lexical item **21**
2.1 The Whole **22**
2.2 The Parts **25**
2.2.1 Collocation 25
2.2.2 Colligation 26
2.2.3 Semantic Preference 28
2.2.4 Semantic Prosody 29
2.2.5 Core 32
In summary **33**
Chapter 3 Come and Go **35**
3.1 Multi-word verbs **38**
3.2 Deixis **39**
In summary **40**
Chapter 4 The Corpora **42**
4.1 Spoken v. written language **42**
4.2 The International Corpus of English **45**
In summary **49**
Chapter 5 Organising the data **51**
5.1 Frequencies **56**
5.2 Statistics **57**
5.2.1 Data 57
5.2.2 Non- parametric v. parametric 57
5.2.3 Between and within ICE corpora 57
5.2.3 Pairwise Comparisons 59
5.2.4 Further Test 62
In summary **64**
Chapter 6 Examining the data **65**
6.1 The co-selection components **66**
6.2 Deictic co-selection components of come and go **67**
6.3 Ante and Post co-selection collocations **67**
6.4 Location **68**
6.5 Familiar idioms **68**
6.6 Post pre-set collocates **69**
6.6.1 go and 69
6.6.2 come back 70
6.6.3 go back 71

6.6.4 come from 72
6.6.5 come in and go in 74
6.6.6 go into 77
6.6.7 come on 77
6.6.8 go on 79
6.6.9 come out 80
6.6.10 go out 81
6.6.11 go through 81
6.6.12 come to 82
6.6.13 go to and go to the 84
6.6.14 come up 85
6.7 Ante pre-set collocates 86
6.7.1 they come and they go 86
6.7.2 to come and to go 87
6.7.3 have to go 88
6.7.4 you come and you go 89
6.8 Discourse managers 90
6.9 Sports reporting 93
6.10 Replacement speech/thought verbs 96
In summary 97

Chapter 7 Narrowing the focus 99
7.1 to go to 99
7.2 come and and and go 104
7.3 come and go 110
In summary 114

Chapter 8 The lexical item revisited 115
8.1 The Parts 115
8.1.1 Collocation 115
8.1.2 Colligation 116
8.1.3 Structural Preference 117
8.1.4 Semantic Preference 117
8.1.5 Discourse Preference 118
8.1.6 Semantic Force 119
8.1.7 Core 120
8.2 The Whole 120
In summary 122

Chapter 9 Conclusion 123
9.1 Research conclusions 123
9.2 Research question(s) 125
9.3 Empirical problem 125
9.4 Further research 125
9.5 Links 126

References 129
Appendix I: Frequency/million words of come and come-grams. 136
Appendix II: Frequency/million words of go and go-grams 141
Appendix III: Between and Within the ICE p-values 148
Appendix IV: Pairwise p-values for come- and go-grams that have been shown to be significantly different between the ICE. 150
Appendix V: Pairwise p-values for come- and go-grams that have been shown to be significantly different within the ICE. 151
Appendix VI: ICE file designations 153
Appendix VII: Concordances of to go to 154

Canada/Private **154**
GB/Private **155**
India/Private **155**
Jamaica/Private **156**

Appendix VIII: Concordances of come and **159**
Canada **159**
GB **160**
India **161**
Jamaica **161**

Appendix IX: Concordances of and go **163**
Canada **163**
GB **163**
India **164**
Jamaica **165**

Appendix X: Co-selection components of the come- and go-grams. **166**

Chapter 1 The Introduction

> *we have to work on the assumption that meaning is created on both axes; for want of more accurate information we may assume that they contain equal meaning potential. There is no reason why one should have a priority in meaning potential over the other.*
>
> *Sinclair 2004a: 170*

It is the simultaneous access to both the syntagmatic and paradigmatic axes of language afforded by computers that, I believe, should be considered to be one of the more important linguistic innovations of this past century. The ability to examine language along both axes at the same time by the generation of concordances of real language associated with a specific word or phrase has produced a seismic shift in understanding how meaning is created in language. It has confirmed that grammar and meaning are inextricably intertwined and not, as traditionally supposed, separate entities, where the investigation of meaning was limited to the paradigmatic axis and the investigation of grammar, the syntagmatic. It has demonstrated that this traditional division of language should not be considered as intrinsic to the nature of language but 'more a consequence of the inadequacy of the means of studying languages in the pre-computer age' (Sinclair 2004a: 165). It has revealed that paradigmatic choice has often been over-estimated and syntagmatic constraints on linear sequences correspondingly underestimated' (Stubbs 2009a: 116). It has shown that meaning in language would appear to be inseparable with form in language. And, as a result, it has established that the unit of meaning is not necessarily a single word but a group of words such as a phrasal unit or a lexical item.

This research is founded on this lexico-grammatical model of language in which meaning is determined within the constraints of both the paradigmatic and syntagmatic axes. Investigation of corpora with the aid of computers to sift and sort the words into concordances has allowed the linguist to examine both axes in tandem. In essence, a concordance is a collection of all the occurrences of a particular word or words in their corresponding textual environment – 'in its simplest form, it is an index' (Sinclair 1991: 32). The use of concordances has empirically shown that not only do words keep company with other words but they like to frequent the same type of places. It is this investigation into what type of company words keep and where they like to hang out that is the starting point of the development of the variety of theoretical and investigative strands that 'all take an integrated approach to lexis and grammar' (Römer 2009: 160). These strands all find 'form and meaning inseparable' where a unit of meaning is normally 'not the word in isolation but a construction or phrasal unit' (Römer 2009: 148).

It has become evident that 'it is not the words which tell you the meaning of the phrase, but the phrase which tells you the meaning of the individual words in it' (Stubbs 2002: 14), which is 'why technology, corpus study and phraseology are intimately related (Stubbs 2009b: 15). The traditional study of 'individual, isolated, invented sentences' precluded the discovery of repeated phraseological units, but the corpus linguistic approach of studying sifted and sorted language shows how 'pervasive' are these units (Stubbs 2009b: 15). It has revealed that 'language is highly patterned' (Römer 2009: 141). This has led to the establishment of both investigative and theoretical models examining meaning creation in language: from an investigative perspective, lexical bundles (Biber *et al* 1999) and concgrams (Cheng, Greaves and

Warren 2006); and, from a theoretical perspective, the lexico-grammars – pattern grammar (Hunston and Francis 2000), lexical grammar (Sinclair 2004a), and lexical priming (Hoey 2005). In addition, the importance of the syntagmatic axis and the importance of real, preferably spoken, language in language modelling was the foundation to Linear Unit Grammar (LUG) – a grammar that has a 'syntagmatic orientation' (Sinclair and Mauranen 2006: xviii).

The research seeks to investigate Sinclair's theoretical model of the lexical item. It takes as its starting point the delineation by Sinclair of the five components of co-selection of the lexical item; the core, the collocations, the colligations, the semantic preferences and the semantic prosody (Sinclair 2004: 141). It draws on this dynamic to investigate the lexical items that might be associated with two high frequency verbs, *come* and *go*, across different spoken and written registers and different World Englishes in four of the International Corpora of English (ICE). These are ICE-Canada, -GB, -India and –Jamaica.

Previous research has identified the existence of lexical items using mid to low frequency words (e.g. Sinclair 2004a - *budge*, Stubbs 2007a - *cause*) but there would appear to have been no research of this kind that looks at high frequency words. However, it should be noted that Sinclair undertook a pilot study of the high frequency word *of* using the same methodology he later employed to identify lexical items (Sinclair 1991: 84). In addition there is no research that takes into account both discourse and English differences. I show that while Sinclair identifies *potential* lexical items he does not necessarily identify the full extent of, all the constituents of and the restrictions of context of these items. I will propose modifications to his methodology that are more suitable to the quantity and type of data under investigation based on his observation in relation to his investigation of *of* that 'the small samples showed hardly any consistency in the relative frequencies' of the instances (*ibid*). I will argue for a reconsideration of the constituents of the lexical item. I will demonstrate that the lexical item is the syntagmatic realisations of a paradigmatic choice. I will propose that as a template to investigate language the lexical item has great strengths but I am not so convinced that it, in itself, is the answer to how language means. I will maintain that it is a step in the right direction and we need to build on this by finding other ways to look at language (with computers) that incorporate the concept of the lexical item. And, in so doing, I will contend that we need to abandon the concept of the lexical item as an object that can be located in the discourse, and re-evaluate our approach to the paradigmatic and syntagmatic axes.

I begin by stating my research questions. In the next section I summarise how there came to be a growing realisation of the importance of the syntagmatic axis in language modelling, which is then followed by an outline of the data under investigation. Finally, I give a brief synopsis of the chapters that follow.

1.1 Research question(s)

I believe that it is in the understanding of the extent, the constituents and the contexts of the lexical item that we might come closer to understanding how language means. With this in mind my main research question is

> **Where are lexical items located across World Englishes and discourses?**

which can be further separated into three sub-questions:

1. **What are the extents of lexical items – where do they begin and end?**
2. **Are lexical items restricted to specific discourses?**

1.2 Paradigmatic and syntagmatic axes

The traditional disconnected approach to language held that the meaning of words was a 'collection of isolated facts' and the 'process by which words are joined together to form sentences […] the province of grammar' (Sweet 1925: 7). The investigation of language on the syntagmatic axis identified slots where different words that are 'associated together in the memory' (paradigms) (Saussure 2013: 145) or are confined to 'the dictionary or lexicon' could be inserted (Sweet 1925: 7). Syntagmatic relations held 'between two of more terms co-present in a sequence' identifiable in the discourse, while associative relations (or paradigmatic relations) held 'between terms constituting a mnemonic group' and were absent from the discourse (Saussure 2013: 145).

Some theoreticians, however, questioned this disconnected approach. They suggested that meaning in language was inextricably bound with the context in which the utterance occurred. The 'bonds of mere linguistics' must be burst and the general circumstances in which the language is used must be taken into consideration, 'the conception of *context* has to be broadened' and 'the *situation* in which words are uttered can never be passed over as irrelevant' (Malinowski 1923: 306); 'the meaning of a word must be always gathered, not from passive contemplation of this word, but from an analysis of functions, with reference to the given culture' (Malinowski 1923: 306, 309); 'the complete meaning of a word is always contextual', in that one can only take seriously a study of a word if it is done in its complete context (Firth 1957: 7); the word 'is entangled, shot through with shared thoughts, points of view, alien value judgment and accents' (Bakhtin 1981: 277); 'the confusions which occupy us arise when language is, as it were, idling, not when it is doing work' as 'nothing is more wrong-headed than to call meaning something a mental activity' (Wittgenstein 2009: 88, 181).

Firth suggested that the meaning of words might be limited by what they co-occur with on the syntagmatic axis. He identified 'meaning by collocation' which is 'an abstraction at the syntagmatic level' (Firth 1968: 176). It should be noted that his concept of collocation is different from its current usage. He defines it as part of 'a mutually congruent series of levels […] beginning with the context of situation and proceeding through *collocation*, syntax, including *colligation*, to phonology and phonetics, even experimental phonetics' (*ibid*). However, it was not until linguists had the ability to sort large amounts of real or 'used' language (Brazil 1995), into concordance lines around key/node words that 'the magnificent vision offered by Firth's admittedly untidy model for language [could] be thoroughly explored' (Tognini-Bonelli 2001: 164).

The simultaneous examination of the paradigmatic and syntagmatic axes of large quantities of 'used' language showed that meaning is indeed constrained by co-occurrence on the syntagmatic axis. Meaning, it would appear, is a complex interaction of 'contextual relations, […] phonetics, grammar, lexicography and semantics' and at the centre of this is the context of situation, the contextualization where the past, present and future of a person's biography and the history and culture of society meet (Firth 1957: 18, 27). However, if meaning is constrained by environment where does meaning actually reside? Is it with the word, or is it with more than the word?

In the following sections I review the investigative and theoretical studies into meaning, including lexical grammar, undertaken within the corpus linguistic paradigm. Although they are different in their approaches they are all linked in that they study used language empirically using corpora and make use of the syntagmatic

dimension. In addition, all but LUG consider frequency of occurrence as crucial; they emphasise 'the pervasiveness of co-selection features and collocations; and they all state that grammatical constructions and phraseological items must play a 'more central role in linguistic theory and description' (Römer 2009: 148).

1.3 Lexical Bundles

The study of lexical bundles is premised on the hypothesis that high frequency patterns are neither accidental nor explanatory but 'corpus-based frequency evidence provides descriptive facts that require explanation' (Biber, Conrad and Cortes 2004: 400). Lexical bundles are multiword units such as *n*-grams that occur in multiple texts (normally greater than or equal to 5) to guard 'against idiosyncratic uses by individual speakers or authors', and that are greater than or equal to an arbitrary frequency value (Biber and Barbieri 2007: 269, Biber 2009: 282). Biber (2009: 283) suggests that 'lexical bundles of any length can by analysed' but Cortes (2004: 400) states that it should only be 'sequences of three of more words'. Nevertheless, the majority of research undertaken has been on bundles of four or more words (e.g. Biber and Barbieri 2007, Breeze 2013). The most common frequency cut off selected is 40/million (e.g. Biber *et al* 2004, Biber 2009, Breeze 2013), but Cortes (2004) uses 20/million, and Biber *et al* (1999: 992-3), 10/million for up to five-word lexical bundles and 5/million for six- and seven-word bundles.

As lexical bundles are identified solely by frequency criteria (e.g. 40/million), 'they might be expected to be arbitrary strings of words that have no linguistic status' (Biber 2006: 172). However, while these bundles would be overlooked by the more traditional linguist as they often straddle structural boundaries, they can be readily interpreted 'in both structural and functional terms' as 'building blocks of discourse' that can be associated with basic communicative functions'. (Biber 2006: 172, 174). Three primary functions can be distinguished:

> Stance bundles express attitudes or assessments of certainty that frame some other proposition. Discourse organizers reflect relationships between prior and coming discourse. Referential bundles make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or to single out some particular attribute of the entity as especially important
>
> Biber, Conrad and Cortes 2004: 384

The research into lexical bundles has also shown that they are both 'more prevalent in conversation than in academic writing' and that 'the most common individual lexical bundles also occur with higher frequencies in conversation than in writing' (Biber 2009: 295).

It has been suggested that as the lexical bundles occur at such a high frequency across texts, it is likely that they are 'stored in memory as unanalyzed chunks' (Nesi and Basturkmen 2006: 286). This could be considered to be of some interest as the lexical bundles 'tend to bridge syntactic boundaries and do not generally have idiomatic meaning' so are 'not very salient, either to the listener/reader' (*ibid*).

1.4 Concgrams

Concgrams are 'sets of words that co-occur regardless of constituency variation (e.g. AB and A*B), positional variation (e.g. AB and BA), or both' (Cheng *et al* 2009:

236). The software to identify them has been developed in order to address the limitations of *n*-grams (a contiguous group of *n* words, where *n* is equal to 1 or more) and skipgrams (non-contiguous *n*-grams). *N*-grams can neither handle constituency or positional variation, and skipgrams can only handle constituency not positional variation. For example, if the *n*-gram AB is *work hard*, the skipgram A*B could be *work very hard* but only a concgram would identify BA - *hard work*, or B*A – *hard at work* (*ibid*).

Frequency is used to identify the canonical form of a particular concgram, and the meaning is then determined. This is then used as a benchmark against which all the other configurations are ranked and a meaning shift unit (MSU) is thus identified – 'a paraphrasable family with a canonical form and patterns of co-selection (*ibid*). A MSU is any combination of words that produces a shift in meaning in comparison to other potential combinations even if 'this is only relatively subtle' (Cheng *et al* 2008: 237). Studies have shown that some MSU are only non-contiguous; that intervening words have a tendency to express semantic prosody; and that meaning 'hardly changes no matter whether the forms are singular or plural' (Cheng *et al* 2008: 240).

According to Cheng *et al*, one of the main advantages of this type of analytical approach is that the notion of the node word becomes less dominant and, thus, less attention is paid to it

> Years of studying KWIC [key work in context] displays have perhaps unintentionally created, in the minds of some users, a hierarchical approach which regards the node as the centre of attention and the words associated with the node as being in a subordinate relationship to it. It is worth [stating ...] that although these are convenient terms to use, the term 'node' does not imply a hierarchy between it and its 'collocate', and that 'node' words that have 'collocates' are themselves collocates if the collocate is studied as the node'
>
> Cheng, Greaves and Warren 2006: 414

However, as even a 5 million word corpus will generate 'very long lists of co-occurring words' concgrammers will need to 'establish reliable ways to automatically process the lists of concgrams to identify those which contain associated words' (Cheng *et al* 2009: 41).

1.5 Pattern Grammar

Hunston and Francis (2000: 3) define patterns as phraseologies that can be 'frequently associated with (a sense of) a word'. They show that 'each pattern occurs with a restricted set of lexical items' and, conversely, 'each lexical item occurs with a restricted set of patterns' (*ibid*). There is a close association with meaning and patterns, in that different senses of words typically occur in different patterns, and words that share patterns have a tendency 'to share an aspect of meaning' (*ibid*). This allows the patterns to be sorted into identifiable meaning groups. Even so, this is only done on the basis of 'the intuition of the person looking at the list' so others 'may well come up with a different set of meaning groups' (Hunston and Francis 2000: 83).

The pattern analysis 'stands by itself' in that no attempt is made to relate 'the elements […] to other, more abstract categories' such as Object or Complement (Hunston and Francis 2000: 176). The patterns are 'restricted to those patterns that distinguish one lexical item from another', such as **V n *from* n** (verb noun *from*

noun); patterns that are typical to word classes, such as **DET n** (determiner noun), are omitted (Hunston and Francis 2000: 203).

On the basis of their analysis, Hunston and Francis argue that the pattern to which a word is associated is a better guide to word class of the word than either meaning or form – 'we create classes for them, based on their behaviour' (Hunston and Francis 2000: 179, 197). On the other hand, this can be problematical as it leads to the question of how many classes of words are needed to understand 'the huge range of behaviour that words have' – too few classes result in a bad fit of words to class, and too many and the map will be 'as large as the area of land it represents' (Hunston and Francis 2000: 197).

They suggest that pattern grammar can be seen in terms of both the traditional constituent-within-constituent hierarchical grammar and the increment-by increment linear grammar (Hunston and Francis 2000: 208, Brazil 1995: 4). In terms of a hierarchical grammar, patterns can be seen to be layered with patterns embedded with in other patterns. For example, if one considers the sentence 'Pte Joseph Byers was the first Kitchener volunteer to be executed' the noun group in **V n** could be further analysed as ***the*** **ORD n to-inf**; and, in turn, the **to-inf** could be ***be*** **V-ed** (Hunston and Francis 2000: 204).

A linear approach to grammar takes into account that language, especially spoken language, occurs in time and is 'in pursuit of a purpose' with one word following the next (Brazil 1995: 26). The speaker is thus able to 'make the best judgments they can manage as to present communicative needs' (Brazil 1995: 28). From the linear perspective, two types of pattern configurations – 'ways in which patterns may follow on from each other' - can be distinguished: *pattern flows* and *pattern strings* (Hunston and Francis 2000: 215). The former occur whenever a word that is part of one pattern 'has a pattern of its own', in other words, the patterns overlap; and the latter, when patterns do not overlap. What, they suggest, occurs is that each word is potentially part of a pattern, and thus the use of a particular word has the potential to prospect other words that will fulfil the pattern. And, if, in fulfilling the pattern, another word is used that again is potentially part of a pattern, this word in turn prospects its own new pattern. A pattern flow will prospect a new pattern before the original pattern is completed, but a pattern string will be contiguous to the next pattern string. It should be noted that, in both circumstances, 'the prospection of a pattern ends (is fulfilled) as soon as the minimum requirement of the pattern is met' (Hunston and Francis 2000: 208-213). They hypothesise that while pattern flows and strings can be found in all types of discourse, the former would appear to be more typical to academic and political argument, and the latter to narrative (Hunston and Francis 2000: 24, 218)

Finally, Hunston and Francis emphasise that pattern grammar is from the tradition that considers language to be a social phenomenon, remaining neutral on 'how language is learned or stored' (Hunston and Francis 2000: 292).

1.6 Lexical Grammar

Sinclair begins his description of lexical grammar by observing that when the phraseological tendency, 'where words tend to go together and make meanings by their combinations', is at work 'words enter into meaningful relations with other words around them' which can compromise 'the independence of the word […] in some way' (Sinclair 2004a: 29, 25, 27). As a result, the word as a unit of meaning is compromised for 'many, if not most, meanings require the presence of more than one word for their normal realizations' and the 'patterns of co-selection among words […] have a direct connection with meaning' (Sinclair 2004a: 133). He proposes the lexical item 'as an abstract category distinct from the word' as a unit of meaning which

'reconciles the paradigmatic and syntagmatic dimensions' observable in concordances (Sinclair 2004a: 133, 144). The lexical item is identified 'using the same descriptive categories to describe both dimensions' (Sinclair 2004a: 148). They consist of five levels of co-selection of which the first and last are obligatory. They are the *core* – 'the evidence of the occurrence of the item as a whole', *collocation, colligation, semantic preference* and *semantic prosody* (Sinclair 2004a: 141).

Collocation is the relation between the core and individual word-forms that co-occur frequently with it; colligation is the relation between the core and grammatical choices that co-occur with it; and, semantic preference is the relation between the core and a lexical field which signals frequent topics in the immediate co-text (adapted from Stubbs 2009b: 22). The three central levels relate to 'each other in increasing abstraction' (Sinclair 2004a: 142): collocation 'is precisely located in the physical text'; colligation requires a word class to be assigned to each word examined; and, semantic preference 'requires us to notice similarity of meaning regardless of words class' (Sinclair 2004a: 142). Semantic prosody is the 'determiner of the meaning as a whole' in that it is the 'reason why [the item] is chosen' and it is 'a subtle element of attitudinal, often pragmatic meaning' (Sinclair 2004a: 142, 144-145). It is 'the junction of form and function' (Sinclair 2004a: 174).

1.7 Lexical Priming

Lexical priming explains lexico-grammar in terms of 'the cumulative effects of an individual's encounters' with words, and thus takes as its starting point words rather than lexical items (Hoey 2005: 8). As each word is primed for use it becomes 'cumulatively loaded with the contexts and co-texts in which it is encountered', and our knowledge of it reflects these co-occurrences and as a result 'regular word sequences are constructed' which are also primed (Hoey 2005: 13). These word sequences, in turn, become 'loaded with the contexts and co-texts in which they occur' – they *nest* (Hoey 2005: 13). Nesting occurs 'when the product of a priming becomes itself primed in ways that do not apply to the individual words making up the combination' (Hoey 2005: 8).

Hoey considers that priming 'contextualises theoretically and psychologically Sinclair's insights about the lexicon' (Hoey 2005: 158). However, he suggests slightly different levels of co-selection that also include textual dimensions. Some of these levels of co-selection have equivalence to Sinclair co-selection categories. Every word is primed for the individual user to have *collocation, semantic association, pragmatic association, colligation, textual collocation, textual semantic association,* and *textual colligation* and this is only reflected in corpora indirectly (Hoey 2005: 13, 158). Hoey's collocation and colligation may be considered to be equivalent to Sinclair's same terms; 'semantic preference and semantic association may be seen as interchangeable'; and Hoey avoids the term semantic prosody on account of the confusion with the term (Hoey 2005: 23). As he points out there is the terminology of Louw (1993), who states that 'certain features of a word's meaning are to be found already present in its surrounds', and that of Sinclair, who states that it is 'the meaningful outcome of the complex of collocational and other choices made across a stretch of language' (Hoey 2005: 23, 24). The term Hoey uses is pragmatic association, and while this is not the equivalent to Sinclair's semantic prosody, it 'overlaps with it' (Hoey 2005: 157); it 'occurs when a word or word sequence is associated with a set of features that all serve the same or similar pragmatic functions' (Hoey 2005: 26). There is nothing that is equivalent to the textual dimensions of the words or word sequences in Sinclair's lexical grammar.

Lexical priming also differs in one other important respect that is relevant to this research. Hoey is 'less confident that the lexical item can replace the word as an

analytical starting point' as 'there is […] no obvious boundary to the posited notion of the 'lexical item' especially in view of [Sinclair's] claims relating to textual dimensions of co-selections (Hoey 2005: 158). He suggests that the phenomena of priming and nesting can account for the choice of the word as the 'analytical starting point' (Hoey 2005: 160).

1.8 Linear Unit Grammar

LUG is a linear grammar that 'organises [any type of] text into tractable units for further analysis' by either a conventional or innovative form, and, in so doing, it shows how 'a latent hierarchy can be discerned in the linear string of word forms' (Sinclair and Mauranen 2006: xv). While the majority of language 'descriptions concentrate on one language variety, whether they say so or not, and the descriptions often perform poorly with any variety other than the one chosen', the LUG can be used for any variety (*ibid*). And unlike the more traditional grammars where what is highlighted is that which is not present – the paradigms - it has a syntagmatic rather than paradigmatic orientation (Sinclair and Mauranen 2006: 6, xviii).

It is based on the idea that language unfolds increment by increment, where each of these increments is initially classified as a chunk (Sinclair and Mauranen 2006: 6). A chunk is a 'pre-theoretical term' that supposes that 'to a user of language any text fall into smallish chunks' with 'variation in the perception of where each chunk starts and stops'(Sinclair and Mauranen 2006: xx, 6). Once the text has been chunked, it is then re-classified using 'a small set of descriptive categories […] with clear working definitions […] and rules for their occurrence and combination' (Sinclair and Mauranen 2006: 8). Although the grammar has been constructed to be an either/or at each stage of the process, these categories can be basically divided into three different types of elements – interactive organisational element (OI), text-oriented elements (OT) and message-oriented elements (M) which are then subdivided where required into various types of message-oriented element such as message-fragment (MF) and message-revision (MR) (Sinclair and Mauranen 2006). Sinclair and Mauranen suggest that the OI, MF and MR elements are then removed from the text while noting their role and what is left is tidied up for further grammatical classification with other grammars.

Conversely, I would argue that it is this facility of LUG to easily show the interactive interpersonal nature (OI elements) of language that is its strength. Constituent grammars require the prior knowledge of the constituent parts for identification, but not LUG. The very act of chunking the language serves to reveal the interactive elements of it, so while I agree that it can be used in order to aid classification of any text with other grammars, I would advocate an approach that builds on the strengths of this grammar in disclosing the interactive nature of language, especially spoken language.

1.9 The Data

If meaning is contextual, meaning must be situated within the discourse where the discourse is 'intentional and meaningful social action' (Stubbs 2007b: 145). If we are going to discover how language means we must start by examining discourse, or to be a little less ambitious, an aspect of discourse. Hence, this research considers how the complementary words *come* and *go* function as a part of lexical items in the spoken and the written language in the corpora of four of the ICE. I believe that it is my choice of the word, my comparison between and within the corpora of the lexical

items, and my identification of the lexical item as a whole rather than its constituent parts that makes this research innovative.

Why have I chosen these particular aspects of the discourse? While I am of the opinion that the corpus-driven/corpus-based dichotomy in corpus linguistics is sometimes somewhat over-stated and the debates that it engenders sometimes somewhat over-heated (see Worlock Pope 2010, Barlow 2011, Stubbs 2013, Gray and Biber 2013), the reason I have chosen these two words is to all intents and purposes a corpus-driven decision. These two words exhibit a significantly higher frequency in spoken compared to written language, and the frequent appearance in corpora 'is good evidence of what is typical and routine in language use' (Stubbs 2002: 221), and it is that which is typical and routine in language use that is 'essential […] to pay more attention to' (Römer 2004: 185).

Previous research has examined words of lower frequencies, and have not differentiated across the discourses within or between corpora. I would suggest that the main reason for not differentiating between the different discourses is the requirement for sufficient concordance lines to examine. One either requires a very large corpus or one has to examine high frequency words. For example, *utterly* (Louw 1993), *budge* and *naked eye* (Sinclair 2004), and *cause* (Stubbs 2007a) are all word forms (or phrases) that have undergone a lexical grammatical examination. Table 1.1 below shows the frequency/million of each of these in all the four ICE corpora taken together (approximately 4 million words in total). The table also gives the frequency of *come* and *go* for comparison purposes.

	word/million
utterly	7.5
budge	2.0
naked eye	0.0
cause	150.5
come	1090.0
go	1490.5

Table 1.1: Total frequency word/million in ICE-Canada, -GB, -India and –Jamaica.

Sinclair (2003) suggests that a minimum of 30 concordance lines is required in order to identify the constituents of the lexical item. In order to compare a word such as *budge* across different discourses a minimum of 15 million words per discourse would be required. This is not such a problem with written language – the British National Corpus (BNC) has 90 million words of written, but it is a huge problem with spoken. The spoken component of the BNC is 10 million words, and Cambridge and Nottingham Corpus of Discourse of English (CANCODE) is 5 million. *Cause* would be more manageable – requiring 200,000 words to probably give sufficient concordance lines but, again, if one wants to examine sub-discourses of the spoken and written components of a corpus it has the potential to become problematical.

My decision to use the ICE corpora in this research was influenced by five factors – (1) availability, (2) comparability, (3) the high frequency of the word forms to be investigated, (4) the dearth of lexico-grammatical research into global Englishes

(Mukherjee and Gries 2010: 525) and (5) the spoken language component – its quantity and the importance of spoken language research. It should be noted that there are disadvantages to the use of the ICE corpora. The corpora that were available for use for this research were compiled during the 1990s so are now approximately 20 years old so it is likely that there will have been some shift in the language use that is identified by this research. Additionally, while the ICE corpora include spoken and written language there is no computer-mediated discourse (CMD), such as chat room conversations, so further research will be needed in this area. However, in terms of this research I am comparing the *come* and *go*-grams between and within the ICE corpora so while changes in the co-selection components might have occurred over the past 20 years in the World Englishes examined and within CMD, it should not affect the conclusions I reach in this research regarding the lexical item.

Firstly, all corpus research, as should all research, should be able to be replicated by other researchers. However, for a variety of legitimate reasons such as copyright and ethics (see McEnery and Hardie (2012) for a detailed discussion of the restraints involved) corpora are not necessarily readily available to all researchers for use with a variety of software – commercial or bespoke – but the ICE corpora are. For example, CANCODE is normally only available to researchers within the School of English at the University of Nottingham, and access to the Bank of English[1], part of the COBUILD corpus, is available for research, but it is not possible to use different varieties of software, such as *WordSmith 6* (Scott 2015), with the corpus. The BNC is readily available and can be used with commercial or bespoke software, but the corpus consists of British English only so its use would have narrowed the research conclusions to only this English.

Secondly, the ICE corpora follow 'a common corpus design and common methodology' (Nelson 2004: 225) allowing the comparison of lexical items across more than one English. Thirdly, it is only possible to use these corpora for this type of research because of the high frequency of the word forms in that there are enough examples of *come* and *go* in the ICE corpora to be able to identify sufficient unique lexical items for analysis.

My fourth factor for choosing to use the ICE corpora is predicated on the emergence of English as a global language, the lingua franca of the Twenty-First Century: I believe it is important to undertake studies into this lingua franca. Its use as a global language has given rise to many varieties, a range of uses, and a greater number of non-native speakers than native speakers (see Kachru 1992; Crystal 2003) many of whom are also more able than their native counterparts (McCarthy 2001: 339). It can be a national language, used as a normal means of everyday communication and predominately a first language (L1); a 'link-language', a neutral language that supports inter-ethnic communication that is predominately the second language of its users (L2) (Mukherjee and Gries 2010: 525); and, it is an international language of business and trade – an additional language for the user that has an economic value.

The final factor, and in my opinion the most important, is the spoken language component. While there are difficulties with regard to the production of spoken corpora, it is important that spoken language is used for research. There is a time cost and thus a money cost to the compilation of spoken corpora – recorded data takes time and manpower to transcribe, so the amount of spoken corpora in comparison to written that is accessible is little. Consequently, one of the strengths of the ICE corpora is that they contain, in relation of their size, a large spoken component with a ratio of 3: 2, spoken to written.

---

[1] http://www.birmingham.ac.uk/research/activity/corpus/resources.aspx (last accessed 22 April 2015

It is a common theme amongst linguists that it is the study of spoken, rather than written language that will eventually allow a full understanding of how language works. Sweet and Saussure suggest that it is the investigation of spoken language in particular that is important. Sweet (1925: 203) states that 'the study of a language should always be based – as far as possible – on the spoken language of the period which is being dealt with'. Saussure (2013: 28, 35) refers to the 'tyranny of the written form' asserting that '[t]he object of study in linguistics is not a combination of the written word and spoken word' but 'the spoken word alone constitutes the object'. Jespersen (1924: 21) felt that 'words and forms were often treated as if they were things or natural objects' and that this was a conception that may 'to a great extent' been fostered through a fundamentally false 'preoccupation with written or printed words'. More recently, Stubbs (2002: xviii) bemoans the over-representation of 'mass media language' and the under-representation of 'spoken language'; Kachru (2008: 5) calls for 'good corpora of spoken material […] before diatypic variation can be understood; and, Cermak (2009: 115) cast doubts 'on the state of our linguistic results, as they are based on written language almost exclusively. As Sinclair says,

> Most corpora keep well away from the problems of spoken language – with some honourable exceptions – and, for a corpus which in any way purports to reflect a 'state of the language', this is most unfortunate. Many language scholars and teachers believe that the spoken form of the language is a better guide to the fundamental organization of the language than the written form; and many writers comment on the differences. In my own experience, there is no substitute for impromptu speech, and a decision I took in 1961 to assemble a corpus of conversation is one of the luckiest I ever made. Even at that time, I was assured that an automatic transcription of speech was 'just around the corner'. It still is.
>
> Sinclair 1991: 15-16

And, it still is. Still we wait, over 20 years later, for accurate automatic transcription of everyday speech; and still we see too much research that is based on written not spoken language. This research aims to begin to redress this balance and to this end …

In the chapters that follow ….

In Chapter 2 I introduce the theoretical basis to this research. I begin by examining the history of the lexical item in terms of Sinclair's idiom and open choice principles (see Sinclair 1991 and 2004a). I then discuss the lexical item as a whole; how it is identified and how this can be seen as problematical as it ultimately relies on the interpretation of the linguist. I then reflect on the co-selection categories in terms of sequence and order where sequence is observable in the data but order is not. I end this section by suggesting that Sinclair's language model could be seen to have connections with Carter (2004a), Pennycook (2012) and Wray (2008). In the second half of the chapter I examine each co-selection category of the lexical item individually. I review what has been written about them and, taking this into account, I deliberate on their strengths and their weaknesses. In particular I divide the collocations into those that are used within the node for the generation of additional concordances – *pre-set collocates*, and those that are identified in relation to the node. I question whether, in relation to colligation, it is best practice to define word classes in terms of the grammatical structures in which they are found. I suggest that there is

little consensus in defining semantic preference and argue for as wide a definition as possible. I reflect on the underlying problem of two conflicting definitions of semantic prosody that has dogged the literature and I express my concern that, perhaps, there is a further problem with semantic prosody. What is the relationship between semantic prosody with the core and/or the node? This brings me on to my final contention that the core and the node should be considered to have different identities.

In Chapters 3 and 4, I consider the data under investigation. In the former I introduce the words under investigation, *come* and *go*, and in the latter I introduce the corpora. I begin Chapter 3 by establishing the minimal assumption on which this research is based. There is a minimal assumption that the significant higher frequency of a small collection of verbs, that include *come* and *go*, in spoken compared to written English merits explanation. I continue by describing the generality and the distinctiveness of *come* and *go*. They can be considered general in that they are frequently part of multi-word verbs, and they are distinctive in that they are often utilised for deictic purposes. I suggest that both multi-word verbs and deixis can be re-defined in terms of the co-selection components of the lexical item. I finish the chapter by discussing in more detail multi-word verbs, deixis and deictic shift theory.

I begin Chapter 4 by delineating the similarities and differences between spoken and written language revealed by recent corpus investigation. This also includes a brief account of the problems of working with spoken language. I continue the chapter by describing the ICE corpora, their history and their make up, and I also suggest an alternative approach to the data that removes the requirement to identify texts by genre or register when analysing corpora. This approach divides the corpora parts – spoken, written, printed, private etc. into colonies (see Hoey 2001). I end the chapter by describing the processes by which the corpora were prepared for the research.

The results of the research are analysed in Chapter 5, 6 and 7. The first chapter in this group takes a quantitative approach to the data; the second, a qualitative; and in the third chapter I narrow the focus in light of the conclusion reached in the previous chapter. In Chapter 5, I examine the different frequencies of *come* and *go* across the four ICE corpora in order to organise the data into manageable quantities and I suggest that the best way to examine the two word forms is as part of *n*-grams above a frequency of 40/million words. In order to differentiate between those *n*-grams associated with *come*, and those associated with *go*, I use the terms *come*- and *go*-grams. *Come*-grams are *n*-grams that include the word *come*, and likewise, *go*-grams are *n*-grams that include the word *go*. Having established that I will examine *come* and *go* as part of *come*- and *go*-grams, I compare their frequencies between and within the ICE corpora. This analysis shows that there is a greater difference within the different colonies of the ICE corpora than there is between the ICE corpora themselves. However, what it does not show is whether there are co-selection component differences between or within the ICE corpora. This is addressed in the next chapter.

With the qualitative approach taken in Chapter 6, I analyse the *come*- and *go*-grams that occur in all the ICE corpora in terms of their principal co-selection components. I introduce two additional co-selection components, *structural preference* and *discourse preference* and I do away with the term semantic prosody, replacing it with the term *semantic force*. Structural preference is the predilection for a node to associate with grammatical structures leaving colligation as the predilection for a node to associate with word classes. Discourse preference is the inclination for a particular set of co-selection components to be associated with a particular type of discourse. I go on to observe that, in relation to the *come*- and *go*-grams under investigation, that all the co-selection components can be seen in the spoken colonies but not all the written colonies. While there are some that are particular to the spoken colonies there are none that are particular to the written colonies.

The subsequent analysis of the *come*- and *go*-grams is divided into familiar idioms, *post* and *ante* pre-set collocates, discourse managers, live sports reporting and replacement speech or thought verbs. The familiar idioms occur infrequently in the data and, as such, cannot be examined in any depth. The post and ante pre-set collocates consist of the analysis of individual *come*- and *go*-grams. The discourse managers are those *come*- and *go*-grams that are explicitly used to manage the discourse. I describe the deictic shift that occurs with *come*- and *go*-grams in live sports reporting in terms of their co-selection components in the penultimate section, and I end with a description of the usage of *go*-grams as replacement speech verbs – again in terms of their co-selection components. In the Chapter I suggest that the semantic forces associated with the post and ante pre-set collocates (eg: *come back* and *to come*) would seem to work in tandem with each other when they are examined as a whole (eg: *to come back*). This idea of semantic forces adding to each other is further explored in Chapter 7.

I begin Chapter 7 by examining closely the instances of *to go* and *go to* and then I look at, also in detail, *come and, and go* and *come and go.* I show that the semantic forces do work in tandem with each other, and I also show that they can be "layered". In other words, depending of the initial choice of node, different co-selection components can be identified and, thus, it can be seen that there can be different semantic forces associated with a particular *come*- or *go*-gram.

In Chapter 8 I return to the lexical item. I initially re-consider the co-selection components and then I re-consider the item as a whole in light of my research. I suggest that there might be evidence of prospection with the collocation preferences I have identified. In terms of colligation, I suggest that colligation with lexical items tends to be outside the node, while colligation with grammatical items, inside. I propose that structural preference should include more than the traditional grammatical structures such as hesitation and ellipsis. The definition of semantic preference is extended to be as broad as possible and I suggest that it is this preference that frequently informs the choice of semantic force. Discourse preference is utilised as the analysis of the *come*- and *go*-grams showed that there were some that were only found in a particular type of discourse. I maintain that the semantic force is associated with the choice of the node, and that the semantic force is either related to the message conveyed, the interaction between the participants in an exchange, or the organisation of the message. Finally, I argue that it is the node not the core that is relevant when considering the lexical item.

In the last section of this chapter I discuss the lexical item as a whole. It consists of the co-selection components identified in relation to a particular node. I then proceed to discuss how this differs from Sinclair's definition, and what it means in practical terms, and how it might be linked to the other theoretical paradigms introduced in Chapter 2.

In the final chapter I summarise my research conclusions and return to the research questions, discussing them in relation to the research results. I review my research by considering the lack of empirical data. I suggest further research areas leading from this research. I consider how my research might be viewed in terms of the other theories developed within the corpus linguistic paradigm. I ask if we should be looking for a new approach to language meaning as it would appear that we still under-estimate the syntagmatic axis and over-estimate the paradigmatic axis. I end by advocating a more linear approach to language with the theoretical integration of LUG and lexical items.

Chapter 2 The lexical item

The word lives, as it were, on the boundary between its own context and another, alien context

Bahktin 1981: 284

The lexical item is rooted in the hypothesis that much of language is constructed according to the idiom principle. Sinclair proposed this principle, in conjunction with the corresponding open choice principle, as a result of his investigations into lexis and grammar in the 1960s. This was the starting point for the work described in the *English Collocational Studies: the OSTI Report* (*Sinclair et al* 1970/2004).

The report was originally circulated at the time amongst interested academics, and was finally published in 2004. It describes the initial empirical analysis undertaken into a corpus of spoken language – a small corpus by today's standards – using computer software. The report includes the following observations: (a) 'a unit of language representing a particular area of meaning [with …] a unique pattern of co-occurrence with other lexical items […] cannot always be identified with an orthographic word'; (b) 'there is a possibility that a word which is more strongly "grammatical" than "lexical" will be a member of a "closed class", be highly frequent and also have a low ability to predict its own environment'; and (c) 'words may vary in the degree of lexicality they display according to the register of language in which they are being used' (*Sinclair et al* 1970/2004: 9, 58, 68).

It is the first of these observations that would appear to be a precursor to the idiom principle, which, in turn, evolved into the lexical grammar and the modelling of the lexical item. As a result of his research, and as 'a natural extension to his work on discourse', Sinclair argued that as computers permitted the linguist to look at language in a completely new way – simultaneously across paradigmatic and syntagmatic axes – they, the linguist, should 'refrain from imposing analytical categories from the outside until [they] have had a chance to look very closely at the physical evidence' (Sinclair 1991: xviii, 29). He advocated moving towards 'a theory that reconciles the paradigmatic and the syntagmatic dimensions and allows the description of the language to remain sensitive to both' (Sinclair 2004a: 174). As he says 'the distinction between grammar and lexis is a very basic model of language [so] there would be no motivation to reconsider it unless new evidence gave rise to concern about its accuracy' (Sinclair 2004a: 165). However, by examining the concordances of words he established that at any point in the development of a text it is constructed either using the *open-choice* or the *idiom* principle. There was now new evidence that gave rise to concern about the traditional distinction of syntax and lexis, or, to be more precise, the influence of the paradigmatic and syntagmatic dimensions on each other.

At this stage Sinclair envisaged the two principles as discrete entities – 'there should be no shading of one into another; the switch from one model to the other will be sharp' (Sinclair 1991: 114), but he later suggests that they should be considered to be part of a continuum – 'two conflicting principles of organization which between them produce a rich continuum' that moves from the *terminological tendency* to *the phraseological tendency* (Sinclair 2004a: 29). With the open choice principle, at each point in the text there are a large amount of options available to the language user to choose from with the only constraint being grammaticalness: at each point 'virtually

any word can occur' as long as 'local restraints' are satisfied (Sinclair 1991: 109). In contrast, with the idiom principle, the 'language user has available a large number of semi-preconstructed phrases that constitute single choices' (Sinclair 1991: 110).

Sinclair suggests that when 'any portion of text […] appears to be constructed on the idiom principle' it would be 'unhelpful to attempt to analyse [it] grammatically' because when the idiom principle is in operation it would appear that '[m]any phrases have an indeterminate extent' allowing 'internal lexical variation', 'inter-lexical syntactic variation' and 'some variation in word order' (Sinclair 1991: 113). He goes on to say that '[m]any uses of words and phrases attract other words in strong collocation', 'show a tendency to co-occur with certain grammatical choices' and 'in a certain semantic environment' (Sinclair 1991: 113). These are the aspects of the idiom principle that are the foundation to the components of co-selection of the lexical item – the core, the collocations, the semantic preferences and the semantic prosody.

Sinclair suggests that a lexical item consists of several words, 'with a great deal of internal variation' which 'disappears when the description invokes an appropriate category of abstraction' (Sinclair 2004a: 35). The citation of the full form of the lexical item removes the ambiguity of language in that each item, whether a single word or a number of words, are normally monosemous (Sinclair 2004b: 20, Teubert 2005: 5). Sinclair envisages a text as a string of lexical items, 'each statistically independent of each of those on either side' with the internal categories assuming 'a central rather than a peripheral role in language description' (Sinclair 2004a: 39).

In the following sections, I first consider the lexical item as a whole. I describe how it is identified, I address a number of points that are related to its identification and, as a model for explaining meaning in language, how it might, or might not, relate to other linguistic theories. I then turn to each of the components of the lexical item, discussing them individually. I begin with collocation, then colligation, semantic preference and semantic prosody. I end with the core as I believe that this presents the greatest problem with the model. Finally, I supply a brief summary of the chapter.

2.1 The Whole

The customary way to identify these co-selection components of collocation, colligation, semantic preference and semantic prosody (but not the core) is to sort concordance lines using the node word – the key word in context (KWIC) – under investigation. This then allows the researcher to examine both the syntagmatic axis – horizontally along each concordance line, and the paradigmatic axis, - vertically, 'scanning for repeated patterns present in the co-text of the node' (Tognini-Bonelli 2001: 2). Interestingly, the syntagmatic axis has, 'as in the Saussurian model […,] what is co-present in the linear string' but the paradigmatic axis has, unlike the Saussurian model what is accessible in the mind of the individual, what is actually present in other texts (Stubbs 2013: 18). Although it should be noted that, while the use of KWIC gives the ability to examine both the syntagmatic and paradigmatic axes in tandem, in so doing the texts in the corpus are torn apart. The node word is removed from everything but a small amount of the text in which it is situated thus divorcing it from both the context of situation and the contextualisation of the speaker's/writer's narrative with that of society (Firth 1957: 27).

Sinclair (2003, 2004a) advocates taking approximately 30 random concordance lines at a time (a screen full) and, having edited the concordance to remove any unwanted material such as duplicates and proper names that are the same as the node word, sorting the concordances to highlight possible patterns. Once the

patterns in the first 30 concordances have been studied and the major patterns ascertained, another random sample of different lines should be generated and then examined, and then, possibly another and another, until no new patterns emerge from the data. He also recommends that the concordances are first sorted according to the words either to the left or the right of the node as the strongest patterns have a tendency to be closest to the node. He suggests that the patterns to look for are repeated words (collocates), repeated word classes and/or grammatical structures (colligates), and repeated word groups of similar meaning (semantic preferences). And, having accounted for any such patterns, he then suggests that it should be possible to ascertain how the patterns function within the text (the semantic prosody). In essence, this type of investigation is initially empirically based – it first identifies and quantifies repeated occurrences or words and word classes, but in its final stages it relies on the interpretation of the linguist in determining semantic preferences and prosodies.

It should be mentioned that as corpus linguistics is an empirical paradigm the reliance on the interpretation of the linguist is a significant problem. It is a 'non-statistical technique' where 'it is the linguist's intuitive scanning of the concordance lines that yields up notable examples and patterns, not an algorithm or recoverable procedure' (McEnery and Hardie 2012). However, I would suggest that if one sees the process as a template to compare and gain a better understanding of language across discourses and varieties, with a view to then developing new algorithms or recoverable procedures, this problem could be said to be mitigated.

Stubbs suggests that all that can be directly observed in the raw data is sequence – 'frequency and distribution', anything that is further proposed should be considered to be order, where

> sequence is a feature of raw data. It is concrete and linear – linear in time for spoken language and in space for written language. It is observable, and with the help of technology, we can observe the frequency of things occurring in sequence. In a rough sense, we can then make inductive generalizations about these things. However, the generalizations involve order. Sequence is one exponent of order, but order is abstract, multi-dimensional and not directly observable. It is a theoretical construct, which relies on interpretation and deduction.
>
> Stubbs 2013: 14

The components of the lexical item can be seen in terms of moving from sequence to order: from collocation to semantic prosody. Collocations are observable within the raw data, and colligations, while abstract are, to a certain extent, still observable. To identify semantic preference 'an intuitive understanding of semantic fields and the topic of the text' is required; and to identify semantic prosody an overview of the communicative purpose of the particular lexical item must be formulated (Stubbs 2013: 24). I would suggest, it is only after the identification of the collocation, colligation, semantic preference and semantic prosody that the core is actually identified. Is the core, therefore, a feature of order or a feature of sequence?

Additionally, Stubbs seeks to relate the Sinclairian model to 'proposals by other theorists' (Stubbs 2006: 25). He suggests that the different co-selection components integrate 'lexis, syntax, semantics and pragmatics' (Stubbs 2006:27). Collocation is lexis in a linear sequence, colligation relates to syntax, semantic preference indicates semantic fields and text topic, and semantic prosody, 'generalisations about the speaker's evaluations and attitudes' (Stubbs 2013: 10). He does not mention the core. He also suggests that the distinction between semantic preference and prosody can be likened to that between locution and illocution associated with speech acts (see Austin 1975). He also argues for an additional sixth co-selection component of the lexical item of *discourse management* suggesting that

semantic prosody actually has two aspects (Stubbs 2013: 10). One is illocutionary force and the other discourse management (*ibid*). However, I am inclined to not include a sixth category as it probably only becomes necessary if one has the desire to create equivalence between Sinclair's model and speech acts. If semantic prosody is defined as the reason for using the lexical item, then the reason for using the lexical item could well be to manage the discourse. In other words, if the reason for using a lexical item is to make a complaint, or to give emphasis to the narrative focus, Stubbs would classify the former as semantic prosody and the latter, discourse management (Stubbs 2013: 10). However, I am suggesting that if the semantic prosody is, as Sinclair (2004) states, the reason for using the item, both reasons can be classified as semantic prosody

Stubbs takes this relationship to speech act theory further by suggesting that, by moving from a description of the lexical item to an explanation of 'how cultural norms are reproduced by frequent phrasal units' (Stubbs 2013: 26), an empirical link could be made to Searle's concept of 'the creation of a social and institutional ontology by linguistically representing certain facts as existing, thus creating the facts' (Searle 2010: 87). While I would not disagree that this would be a great achievement, I am more disposed to argue for links with Carter (2004a), Pennycook (2010) and Wray (2008).

Carter argues that language can ultimately be seen as creative, 'it can be a matter of re-vision as well as vision, of re-membering as well as dis-membering and or re-creation as well as creation' (Carter 2004a: 48). But, he also argues that parallel to these creative compositions exist 'a range of non-creative, formulaic expressions' that 'stabilise and routinise' communication (Carter 2004a: 133). He also suggests that some of the formulaic expressions are 'sufficiently flexible in form' that they are open to creative reconfigurations (Carter 2004a: 129). This, I would suggest, could be seen as comparable to Sinclair's idea of language being organised according to either the open choice or the idiom principle. The open choice could be said to represent the creativity of language, and the idiom, the stabilisation and routinisation of language as evidenced by the lexical item, bearing in mind that Sinclair suggests the lexical item can have considerable variability within its constitution.

Likewise, Pennycook appears to take a corresponding view point with regard to language. He contends that language is a local practice, a social activity that is embedded in locality, where both repetition and creativity in discourse are the 'norm rather than the exception' (Pennycook 2012: 4, 40). Again, the creativity of language could be said to be represented by the open choice principle, and the repetition, the idiom principle. Pennycook also suggests that 'grammar is not a set of norms that we adhere to or break, but rather, the repeated sedimentation of form as a result of ongoing discourse' (Pennycook 2012: 41). This idea, that grammar is produced with the repeated sedimentation of form, links the paradigmatic with the syntagmatic dimension, as, of course, does lexical grammar.

Wray approaches language modelling from a psychological perspective. She proposes that language consists of Morpheme Equivalent Units (MEUs) which she defines as

> a word or word string, whether incomplete or including gaps for inserted variable items, that is processed like a morpheme, that is, without recourse to any form-meaning matching of any sub-parts it may have
>
> Wray 2008: 12

This, I would suggest, corresponds to Sinclair's assertion that the lexical item can vary internally and is 'normally monosemous' or is equivalent to a morpheme, although it should be noted that Sinclair does not make any claims as to the processing of language (Sinclair 2004a: 55, 20). However, where Wray and Sinclair

differ is that she has reservations with regard to Sinclair's concept of a language continuum conceptualised by the terminology tendency and the phraseological tendency. She questions whether the continuum model is adequate in terms of the processing of language 'to account for both creative flexibility and the idiomaticity arising from preferred ways of saying things' (Wray 2008: 14). She suggests that language users have 'multiple part-mappings of the same information in the lexicon' (Wray 2008: 15). Phrases can be both stored as wholes, and stored as the individual parts. The language user processes language on a Needs Only Analysis (NOA) basis where input is first checked against known lexical units and only if some variation is identified does the user undertake additional analysis of the unit (Wray 2008: 17). In this way, the pressure on working memory in real-time can be minimised (Wray 2008: 69). She anticipates that MEUs should be 'shared across a speech community' and should be able to be 'reliably identified' by that community once linguists have agreed on the diagnostic criteria (Wray 2008: 107).

2.2 The Parts

2.2.1 Collocation

Collocation is, at its very basic, the propensity for words to associate together. The customary way of determining collocates of a word is by generating concordances and identifying them with reference to the node word either by frequency of occurrence, or by calculating the statistical significance of the co-occurrence using appropriate software. Once they have been identified they can either be used as part of the description of the word or words under investigation, or they can be used to create additional concordances that include these collocates. As there is no specific fixed relationship between a word and its collocations, bar the fact that they occur in the same text, it is probable that collocation has a 'limited value for linguistic theory' (Barnbrook, Mason and Krishnamurthy 2013: 172). What collocation does do is to give 'a good initial impression of the meanings of a word' by 'condensing the data available from concordance lines' producing significant breakthroughs in such paradigms as corpus based lexicography (*ibid*).

With regard to the generation of collocations: for those who advocate a statistical approach to determining collocations, there is some scepticism regarding the methods available. There are at least 30 plus association measures but there would appear to be little work done on validating any of the methods 'against findings from corpora-external data' (Gilquin and Gries 2009: 17). As Gries maintains, there are 'corpus linguists who pretty much argue for trying different ways to modify existing measures and pick whatever yields results that intuitively (!) appear best' (Gries 2010: 6). There are also a number of weaknesses inherent in the current association measures in that they 'hide much of the interesting variability in the data', including: the 'directionality effects'; the homogeneity of associations across corpora or parts of corpora; and, there are problems associated with extending the measures for multi-word units (Gries 2013: 159). In addition the statistical measures also depend on a normal distribution, rather than the Zipfian distribution of data that is inherent in corpora. A Zipfian distribution is 'a very skewed distribution' (Kilgarriff 2002: 112) in that there 'is a constant linkage between word frequency and word rank' – where the word occurs when the words in a text are ordered based on their frequency of occurrence (Scott and Tribble 2006: 27). A word list will contain 'a very small number of very highly used items, and a long declining tail of items which occur infrequently' (*ibid*). In order to address these issues, Gries (2013) proposes a new measure for collocation ΔP. This measure is more sensitive than the more

traditional measures as 'it can tease apart which collocates in a collocation' have the strongest or weakest attractions or repulsions in comparison with the other collocates; 'it is not a significance test' so does not require a normal distribution; and it also 'provides directionality information' (Gries 2013: 152).

Whether this new measure also addresses the additional concern that some association measures have a tendency to disfavour 'combinations that incorporate high-frequency words' thus often excluding 'function words from consideration' is unclear (Biber 2009: 287). This tendency suggests an underlying assumption 'that collocations are combinations of content words, while a lexicogrammatical combination of function word plus content word is a different phenomenon' (*ibid*). This research maintains that there is no difference, it is as important to consider the function word collocates as it is to consider the content word collocates.

Sinclair is also sceptical about association measures; he says

> Over the years I have become more and more suspicious of these tests. I may still use t-score for my day-to-day research in the absence of anything more plausible, but I have lost most of my original confidence in it and in other statistical procedures. If something like the co-occurrence of two or more words is statistically significant, this tells me that there is but a small chance of it being accidental. But I don't expect it to be accidental anyway. One of the worrying aspects from the very beginning, what really made me suspicious, was the frequent finding that the actual co-occurrence of words in texts is many times the prediction that is made on a statistical basis. Not just slightly over the estimate, but hundreds or thousands of times more frequent than the expected. Statistical prediction based on chance seems just irrelevant
>
> *Sinclair et al* 1970/2004: xxii

However, whatever the potential problems associated with the identification of collocates, what collocation studies have shown is that there is a good deal of vocabulary that is 'to a greater or lesser degree fossilized into restricted patterns' thus the emphasis has been shifted from single words to multi-words as holders of meaning (McCarthy and Carter 2006: 8).

With regard to the application of collocation: it can be considered to be frequent co-occurrence with or without significance of two or more words so any process that identifies words that co-occur frequently could, I would suggest, be considered to be a process that is generating collocates. Finally, I would suggest that when the concordance lines are generated where collocates are being used as part of the node it would be helpful to term then as *preset collocations* in order to distinguish them from collocation as part of the description of the function of an item under investigation. In investigating the lexical item, this research draws on both concepts.

2.2.2 Colligation

Colligation is the co-occurrence of word classes or grammatical structures with the node word (Sinclair 2003: 175). However, word classes are abstract entities that have been pre-defined in traditional linguistics and, as such, work against Sinclair's observation that 'the corpus seems to be signalling […] the need for a major overhaul of the notion of word class' (Sinclair 2004: 173). Is it legitimate, then, to use them in the description of the lexical item?

Word classes themselves are broad, essentially abstract, categories, that depending on the theorist defining them, can be classified differently. For example, Jespersen (1924: 72) uses the category *noun* 'for the large class of which substantives

and adjectives are subdivisions' – substantives would be more commonly classified as nouns, and adjectives and substantives are, more commonly, considered to be separate word classes. (Admittedly, by classifying them as part of a larger class of nouns, Jespersen goes someway to solving the problem of substantives/nouns being used as quasi-adjectives, so there is some merit in so doing.)

The investigation of spoken corpora has resulted in a suggested overhaul of word class categories. It has led to the identification of an additional word class or classes. Carter and McCarthy (2006) limit this additional word class to *discourse markers*, and relate it to the different word classes of nouns, verbs, adjectives etc., but Biber *et al* (1999) state the word class includes all *inserts* (incorporating discourse markers) and defines it as an additional type of major word class to be included with *function* and *lexical* words.

According to Carter and McCarthy (2006: 208), a discourse marker is a type of *pragmatic marker*, an item that operates 'outside the structural limits of the clause and which encodes speakers' intentions and interpersonal meanings'. Pragmatic markers also include *stance markers* (indicative of 'speakers' stance or attitude' about the message), *hedges* (allowing a speaker to be 'less assertive in formulating' messages), and *interjections* (indications of 'affective responses and reactions to the discourse') *(ibid).* Discourse markers can be words or phrases and are a 'lexical rather than a grammatical category'; their function is to 'link segments of the discourse to one another' reflecting 'choices of monitoring, organisation and management', and they also can be used to 'indicate degrees of formality and people's feeling towards the interaction' (Carter and McCarthy 2006: 209, 212). However, Carter and McCarthy suggest that as lexical entities, there is a problem in categorising them 'in terms of the conventional word classes' such as nouns and verbs, and they suggest that they should be considered to be 'a class in their own right' (Carter and McCarthy 2006: 209).

The position of Biber *et al* (1999) could be considered to be more radical. They define inserts as stand-alone words that are unable to 'enter into syntactic relations with other structures' and are 'versatile in taking on different conversational roles' (Biber *et al* 1999: 1082). The inserts include *discourse markers* (signalling 'a transition in the evolving progress of the conversation' and the 'interactive relationship between speaker and hearer, and message), *interjections* (exclamatory and 'expressive of the speaker's emotion'), *greetings and farewells, attention signals* ('attracting the attention of the addressees'), *response forms* ('brief and routinized responses to a previous remark by a different speaker') , *hesitators* ('pause fillers'), *polite speech-act formulae* (routinized 'conventional speech acts, such as thanking, apologizing, requesting and congratulating') and *expletives* (taboo and moderate) (Biber *et al* 1999: 1082-1093). And they state that inserts should be considered as an additional class to the two major word classes – function and lexical words.

Sinclair makes the point that 'the commonest five words do not fall in with any normal word classes' (Sinclair 1999: 157). These words are *the, a, of, to* and *and.* He suggests that instead of assuming that these words can be forced to fit into traditional word classes, corpus evidence suggests that they are each in a word class of their own that shares 'some of the defining features of one or more classes, but showing either unique usage patterns or a unique combination of them' (Sinclair 1999: 166).

Recent corpus based language investigations would also suggest that the traditional word class categories are in need of some revision. Hunston and Francis state that words are not classified because they ''have' classes as something intrinsic to them' but they are classified because of the behaviours they exhibit' (Hunston and Francis 2000: 197). Word classes have sets of patterns that are associated with them, and that these patterns 'are the most consistent way of determining class' in contrast to determination by common morphological or semantic features (Hunston and Francis 2000: 179). I would argue that if word class is dependant on the grammatical

patterns (or structures) in which a word is found, then if one is identifying colligation as co-occurrence of word class or grammatical structure the whole definition process is circular. Possibly one is defining something on the basis of what one is seeking to define?

2.2.3 Semantic Preference

There would appear to be little consensus as to what semantic preference actually entails as it has been defined and re-defined a numbers of times subsequent to the initial proposal by Sinclair. I will take Sinclair's definition as a starting point – the co-occurrence of 'words of a particular meaning' regardless of word classes with the node word (Sinclair 2003: 178). I will then introduce and discuss a selection of other definitions relative to this definition.

Stubbs suggests that semantic preference 'concerns propositional content' and is an indication of the topic of the text (Stubbs 2013: 10). This characterisation is probably a little broader than that of Sinclair who suggests that semantic preference is the predilection of the node word to be hanging around with words of a particular meaning and, as such, says nothing with regard to text. To include the notion of topic, would I suggest, extend semantic preference to recognise semantic sets associated with the particular text. The allusion to the propositional content relates to the position that Stubbs takes regarding the equivalence of the co-selection components of the lexical item with speech act theory. As I have indicated, I am inclined to associate lexical grammar with the theoretical stances of Carter (2004a), Pennycook (2010) and Wray (2008), although I would not disagree that semantic preference within this context is an indicator of propositional content of the text in question. Also, as this research will show that lexical items can be register specific, it could be said that the semantic preference exhibited by a node word can be dependant on the topic of a text.

On the other hand, Hunston's suggestion that semantic preference (or 'attitudinal preference) should be used 'with items expressing a particular evaluative meaning' (Hunston 2007: 266), would appear to narrow the definition to those words of a particular meaning that have an evaluative function. However if evaluation is a 'broad cover term for the expression of the speaker or writer's attitude or stance towards, view point on, or feelings about the entities or propositions that he or she is talking about' (Hunston and Thompson 2000: 5), semantic preference possibly shifts to a more pragmatic rather then semantic phenomenon.

Bednarek suggests that collocation should be sub-divided into positive and negative collocation, and semantic collocation, and argues that semantic preference should be used 'as a cover term' for these frequently co-occurring and similar ('differing only in degrees of 'generality'') collocational phenomena (Bednarek 2008: 121). As such, semantic preference can be considered to be '(relatively objectively) observable by looking at corpus evidence' (Bednarek 2008: 122). I think that this definition is interesting in that it perceives collocation as something more than just words in text; it can have a grammatical aspect. I am inclined to keep this notion under consideration, but still favour the idea that the lexical item has potentially five co-selection components as it allows the linguist to distinguish between those that relate to sequence, and those that relate to order (Stubbs 2013).

Hoey introduces the term semantic association instead of semantic preference, although he regards them as interchangeable (Hoey 2005: 23). He defines semantic association as the propensity for 'every word' to be primed for someone 'to occur with particular semantic sets' (Hoey 2005: 13, 15). This definition extends the Sinclair definition by including the explanation as to how semantic preference can be explained in terms of the producer. It is not the node word that is primed for semantic preference, but the person is primed for a particular semantic preference with a particular node word. This research does not seek to hypothesise about the production of language by a speaker or writer.

To Xiao and McEnery (2006; 107), semantic preference has a distinct collocational meaning that is 'a feature of the collocates' (Xiao and McEnery 2006: 107). They see semantic preference as distinct but interdependent with semantic prosody, which, they state, has a distinct collocational meaning that is 'a feature of the node word' (*ibid*). Again, this would appear to narrow the definition supplied by Sinclair, in that they suggest that semantic preference is specific to the collocations that have been identified, rather than the co-occurrence of words that, taken together, have the same meaning. In other words, the collocations of the node word are identified and then the semantic preference is ascertained from these collocations, nothing more.

This research will use the original definition of semantic preference given by Sinclair, but will take into account that there could be inherent discourse and variety differences, and that collocation might be considered to be sub-divisions of semantic preference. Semantic preference is the predilection of words that mean much the same, regardless of word class, to associate on a regular basis with a node word.

2.2.4 Semantic Prosody

There would seem to be considerable confusion as to what semantic prosody is as researchers would appear to be bringing into play an amalgamation of 'the conflicting positions' on semantic prosody 'without any apparent recognition of this conflict' (Stewart 2010: 3). I would suggest that the initial confusion stems from the differing account proffered by Louw (1993) and Sinclair (2004a). As Stewart states

> Although semantic prosody has been assigned certain features which would appear to be common to almost all accounts of it, it is nonetheless the case that the first two exponents of semantic prosody, Louw and Sinclair, described it in very different ways. Most subsequent contributions on the subject contain features of each of these descriptions, and some may be crudely divided into those influenced primarily by Louw, and those influenced primarily by Sinclair. It is normal that as a concept develops, it will be approached and discussed in several ways, but the impression is that single contributions do not give sufficient stress to the degree of difference between these main approaches. As a consequence, the appellation 'semantic prosody' has become something of an umbrella term whose breadth may deceive those anxious to find out more on the subject.
>
> Stewart 2010: 159

A variety of solutions have been put forward in order to resolve the problem, which, I would suggest, have in turn created more confusion. These include contributions from Louw (2000), Whitsitt (2007), Hunston (2007), Bednarek (2008), and Stewart (2010). Also, Hoey (2005) has introduced another concept that is akin, but different, to semantic prosody that he terms pragmatic association. This too would seem to have introduced even more misunderstanding, for example, as found in Ellis *et al* (2009), and Morley and Partington (2009).

The concept of semantic prosody was first described by Louw, although he credits Sinclair with suggesting the term (Louw 1993: 230). The word prosody is used in the same way as 'Firth used the word to refer to phonological colouring' that transcends 'segmental boundaries' (*ibid*) and 'semantic because it deals with meaning' (Sinclair 2003: 117)). Louw demonstrates both the role of semantic prosody in the use of irony and insincerity by speakers or writers, and how it can be used to grade suasive writing (Louw 1993: 230). He states that semantic prosody can

be identified from the 'habitual collocates' of words, or sequence of words, which colour their meaning so it can no longer be considered in isolation when used without these collocations (Louw 1993: 234). He suggested that semantic prosody is a reflection of 'either pejorative or ameliorative changes' as a result of historical refinement through language change (Louw 1993: 238): in other words, as a result of language change words can become associated with either negative (pejorative) or positive meaning (ameliorative). Once the semantic prosody of a word, or sequence of words, 'predominates sufficiently strongly' the word, or sequence, can be used to create ironic meaning by prosodic clash – deliberate usage running counter to the semantic prosody (Louw 1993: 234). He then suggested that where this type of usage is not deliberate, it is not just a slip of the tongue but an indication of insincerity in the speaker or writer (Louw 1993: 239). In establishing that semantic prosodies occur as a result of pejoration or amelioration he goes on to claim that suasive writing can then be graded according to the 'good' or 'bad' prosodies it contains. He also suggests that semantic prosodies tend to occur together - they 'hunt in packs' (Louw 1993: 239). In essence, he defines semantic prosody as a collocational phenomenon that shades the meaning of a particular node word to be either "good" or "bad", and where there is one semantic prosody there is likely to be another close by.

This is somewhat different to the Sinclairian approach. Sinclair views semantic prosody as 'something close to the 'function' of the item – it shows how the rest of the item is to be interpreted functionally' to the extent that 'without it, the string of words just 'means' – it is not put to use in a viable communication', and, without it, it would be difficult to integrate an item with its surroundings (Sinclair 2004: 34). It is the semantic prosody of an item that dictates why the item was chosen, 'over and above the semantic preferences that also characterize it' (Sinclair 2004: 145). Its meaning is attitudinal and often pragmatic (*ibid*). He believes that one of the most important contributions that corpus investigation has provided is 'the recognition that semantic prosody is a constant feature of text' (Sinclair 2003: 178).

In a later paper Louw seeks to resolve any possible confusion over the definition of semantic prosody. He re-emphasises the importance of collocation in identifying semantic prosody: he says it is 'established through the proximity of a consistent series of collocates' and should be recovered 'computationally from large language corpora rather than intuitively'. He re-states that often semantic prosodies are 'positive' or 'negative' and he also suggests that the 'negative prosodies are much more frequent' (Louw 2000). He re-asserts that irony is created 'through the deliberate injection of a form which clashes with the prosody's consistent series of collocates' and where this is inadvertent it is an indication of the insincerity of the producer (Louw 2000). However, he also states that they arise 'from ***fractured*** [emphasis in the original] contexts of situation', where, either, delexicalisation has, in effect, deleted the 'human participants from the context of situation' by replacing them by 'desirable or undesirable human characteristics', or, where something has occurred – 'caused or spontaneous' – that meant that the context of situation is incomplete or fractured (Louw 2000). I would suggest that this additional refinement to Louw's definition of semantic prosody moves it even further away from Sinclair's definition. Louw perceives semantic prosody as an aspect of the collocations of the node word that imbues it with, predominately, fracture and negativity: and Sinclair perceives semantic prosody as a major function of text production that indicates why a particular item is chosen in conjunction with the node word, which is not necessarily associated with collocation and/or semantic preference. It is not really surprising that attempts have been made to re-define semantic prosody.

Whitsitt says that semantic prosody, at the time he was writing, 'is defined in at least three, distinctly different ways', that of Louw, that of Sinclair, and a third that he suggests 'is very widespread' and 'treats semantic prosody as if it were a synonym of connotation' (Whitsitt (283, 285). He is critical of the term itself, taking the view that the idea of a phonological colouring on which Sinclair/Louw based the term

semantic prosody is not carried beyond the immediate environment in which it occurs – the vowels in *Amen* only have a nasal quality when in proximity to m and n, so neither should semantic prosody (Whitsitt 2005: 291). He also makes the point (perhaps a little melodramatically but it is worth quoting) in regard to the definition supplied by Louw (1993, 2000) that it appears that

> [the node word that is] imbued with meaning, is empty; or, to put it another way, let us return to the story of semantic prosody, which is that of a word-form which is inexplicably empty, or perhaps not so much empty as weak and innocent, and suddenly finds itself (could be "herself") inexplicably thrown into a world of bad company, which is made up of unpleasant words which are, likewise inexplicably, full of themselves, and cannot help themselves from pouring their negativity into any empty form which is near them. In this world, proximity clearly leads to promiscuity, but the flow is always one-way, form strong, full, bad words, into the weak, empty, innocent forms, which are incapable of resisting the force of bad company, to which they too will soon belong, and from which they can never leave, ever again.
>
> Whitsitt 2005: 292

What Whitsitt does not do is to suggest how the dichotomy between Sinclair and Louw might be resolved. The different approaches that have been taken are discussed at length by both Hunston (2007) and Bednarek (2008) as they attempt to settle on one or other of the definitions.

Hunston advocates restricting the term 'to Sinclair's use of it to refer to the discourse function of a unit of meaning' which cannot necessarily be precisely articulated and is not necessarily 'negative' or 'positive' (Hunston 2007: 266). She also advocates the use of semantic preference, or attitudinal preference, to 'refer to the frequent co-occurrence of a lexical item expressing a particular evaluative meaning'(*ibid*) . While Bednarek states that semantic prosody refers 'to the *complex attitudinal and/or functional meaning* [her italics] of lexical items' and she believes that it is 'crucial to uphold [Sinclair's] distinction between semantic preference and semantic prosody' where semantic preference is associated with collocation – semantic or negative and positive collocation (Bednarek 2008: 131, 132, 121). However, although both these linguists are adamant that Sinclair's terminology should be the one that is utilised, others such as Stewart (2010) disagree. Confusion still would appear to exist additionally confounded by Hoey's semantic association and his pragmatic association (Ellis *et al* 2009, Morley and Partington 2009).

Hoey (2005: 23) stresses that there would appear to be two accounts of semantic prosody in operation – Louw's and Sinclair's – and rather than opting for one or the other he introduces two new terms, semantic association and pragmatic association. Semantic association is exactly the same as semantic preference, and pragmatic association, overlaps with, but is not the same as, semantic prosody (Hoey 2005: 23, 157). It occurs when there is an association of a word, or nested sequence with similar pragmatic meanings (Hoey 2005: 26).

Ellis *et al* (2009: 89) examines 'the psycholinguistic reality in language users of the phenomena of collocation and semantic prosody'. They define semantic prosody in a similar way to Louw – tied to collocation (Louw 1993, 2000). They say it is 'the general tendency of certain words to co-occur with either negative or positive expression', but they use 'a famous example' of Sinclair's – *set in*, in order to elucidate their point (Ellis *et al* 2009: 90). They also explain that Hoey classifies this as semantic association (*ibid*). However, while Hoey would classify this as semantic association, he is at pains to point out that it is his semantic association and Sinclair's semantic preference, not semantic prosody, that are interchangeable (Hoey 2005: 23).

It is strange that Morley and Partington claim that 'corpus linguistics seem to be reaching a general agreement in appreciating the good-bad, positive-negative distinction' as an integral part of evaluative/semantic/discourse prosody (Morley and Partington 2009: 143; see also Partington *et al* 2013: 58), as the evidence might suggest otherwise. They suggest that semantic prosody is a paradigmatic phenomenon that has a diachronic dimension. It relates to the 'innate human need', and possibly to that of other biological organisms, to evaluate things as good or bad (Morley and Partington 2009: 141). They say that it is essential to survival (*ibid*). It would appear to have an 'extraordinary unifying explanatory power regarding the function of communication' because it maintains evaluative harmony in text by co-selection of items (where item is either a single orthographic word or a multi-word unit) that have the same evaluative or attitudinal force (Morley and Partington 2009: 143, 145). They say that semantic prosody is in the DNA of an item, is part of connotational meaning that is 'expressed over stretches of discourse' that is 'shaped by semantic preference'; they say it can be viewed from three different perspectives – lexical-priming, textual discourse, and statistical discourse; and they say it can be 'switched off or overridden or exploited by users' (Morley and Partington 2009: 149, 151, 142, 146). Diachronically, they suggest that there is an interaction between items and semantic prosody that alters the 'priming instruction-suggestions of an item' over time (Morley and Partington 2009: 151).

It is surprising how Morley and Partington portray the position Hoey takes with regard to lexical priming (Hoey 2005). He states that semantic prosody overlaps with but is not the same as pragmatic association (Hoey 2005: 23). This undermines both their claim that semantic prosodies can be 'switched off, overridden, or exploited' and the claim for semantic prosodies to be part of the diachronic process of language change (Morley and Partington 2009: 146). They support their arguments with lexical priming theory, but in so doing make claims for lexical priming that do not appear to be supported by Hoey.

Stewart proposes, contrary to Hunston and Bednarek, that 'discourse prosody' should be used 'to denote the 'Sinclair interpretation', and 'semantic prosody' to denote the 'Louw interpretation' (Stewart 2010: 162). He suggests that the confusion is such that semantic prosody either 'can represent the reason for making the utterance' or it is 'an 'aura', 'halo', 'shade' or 'hue' of meaning' (Stewart 2010: 54). He argues that semantic prosody has emerged as a

> a phenomenon/feature/meaning which extends/stretches/ranges/is spread/is dispersed either (i) over/across a(n extended) unit of meaning/unit of language/discourse unit, or (ii) over /across more than one unit/several units.
>
> Stewart 2010: 51

He questions the nature of the units under discussion and what might be intended by the metaphors such as stretching or spreading (*ibid*). He is also concerned that the 'typical presentation of the concordance' privileges the word not the unit of meaning suggesting that this is one of the greatest outstanding problems for semantic prosody: what is semantic prosody's relationship with the node/core (Stewart 2010: 121, 163)? This research will show that this is indeed the case.

2.2.5 Core

Semantic prosody and core are obligatory co-selection components of the lexical item, and collocation, colligation, semantic preference are considered to be optional (Sinclair 2004a:141). All but the core are identified using concordance lines that have been generated around a node word or words, and as such are identified in association

with the node. However it is not entirely clear how the core is to be identified in that it constitutes 'the evidence of the item as a whole' (*ibid*). Is the core and the node one and the same thing as some might suggest or are they different (see Stewart 2013: 163, Xiao and McEnery 2006: 107)?

The node is the word or words that are fed into the computer software in order to generate a concordance for examination of KWIC. This type of examination has, possibly, placed the node as central to the classification of the lexical item which, in turn may have unintentionally created 'a hierarchical approach which regards the node as the centre of attention and the words associated with the node as being in a subordinate relationship to it' (Cheng *et al* 2006: 414). This would suggest that the node is essentially part of the toolkit for identifying a lexical item and once the lexical item has been identified it becomes superfluous as the core is 'the evidence of the occurrence of the item as a whole' (Sinclair 2004a: 141). The core cannot include collocation, colligation, semantic preference or semantic prosody as it is a co-component of these co-selection components. It can include 'one or more words that are either invariable or subject to certain grammatical variations' (Sinclair 2003: 173). The variations that are permitted are grammatical inflection or 'membership of a specified grammatical class or a lexicalisation of this class' (*ibid*).

In effect, it would appear that Sinclair might be suggesting that the core is an extension of the node allowing for variation of the inflection (plural, tense etc) and/or substitution of grammatical class words with the equivalent grammatical word or a lexical word. However, he also states that whereas it was traditionally presumed that different forms of a lemma shared the same meaning 'we are now beginning to discover that in some cases, if they did not share the similar spelling, we might not wish to regard them as being instances of the same lemma' (Sinclair 2004a: 17). This seems to contradict the idea of the core allowing for variation of the inflection.

In essence, by generating concordances of a particular node, a potential lexical item can be identified, but the node is not necessarily a fixed feature of the lexical item. Having identified a potential lexical item, it is the semantic prosody that will identify subsequent forms of item, and thus identify the core which might include inflectional and grammatical variation. This would suggest that the node and the core do indeed have separate identities where the node is a feature of sequence and the core by the nature of its identification, order.

In summary

The main theoretical points that have further relevance to this research are summarised below The chapters that follow describe the research I have undertaken and I discuss further these theoretical points in light of my research findings in Chapter 8.

- Concordances allow the examination of language across the syntagmatic and paradigmatic axes simultaneously.
- Sinclair states that the lexical item consists of five co-selection components, two obligatory and three optional. The core and semantic prosody are obligatory, and collocation, colligation and semantic preference, optional.
- The co-selection components go from sequence to order – collocation to semantic prosody. Sequence can be directly observed in the data, but order is a theoretical and its identification relies on the interpretation of the linguist.
- Theoretical links between Carter (2004a), Pennycook (2010) and Wray (2008) have been suggested.

- For the purposes of this research two types of collocation have been identified – pre-set collocation and co-selection collocation. The former is a feature of the node, and the latter is identified as a co-selection component of a particular node.
- Sinclair defines colligation as the co-occurrence of word classes and or the patterns associated with the node.
- Sinclair suggests that semantic preference is the propensity for the node to be associated with a word or words that have the same meaning regardless of word class. Other linguists have extended this to include topic – the propensity for the node to be associated with semantic sets. Additionally, it has also been suggested that collocation should be considered to be a sub-division of semantic preference.
- Semantic prosody is problematical as there are two conflicting definitions. One arising from Louw (1993) and the other Sinclair (2004). This research follows Sinclair's definition that the semantic prosody of an item is the reason why the item was chosen.
- It would appear that the core of the lexical item is a highly problematical premise.

Chapter 3 Come and Go

the place to which one goes is a place where I am not […]. The place to which one comes is a place where I am or where you are
Fillmore 1966: 223

This research incorporates both corpus-driven and corpus-based approaches. However the foundation of this research is a corpus-driven observation that I made when examining high frequency verbs in the BNC. Corpus-driven research starts with minimal assumptions and develops new models from the findings extracted from the corpora, in contrast to corpus-based research which incorporates pre-existing hypotheses and often aims to validate established models of language (Mahlberg 2005: 17, Teubert 2004: 112). The minimal assumption that this research is based on is that 'frequency data identifies patterns that must be explained' (Biber *et al* 2004: 376) so the occurrence of lexical verbs at a significantly higher frequency in spoken rather than written discourse would suggest that the patterns they occur in need to be further investigated using corpora.

Verb	Frequency/million	Log Likelihood (G^2)
got	932	117320
know	1233	104930
think	916	71946
mean	411	53431
get	995	46650
go	881	35449
say	679	24125
want	572	20109
going	658	16769
put	596	15152
come	695	13389
see	1186	13371
thank	122	12397
like	344	12162
wan~	26	9356
look	433	9310
let's	83	8097
saying	180	6386
talking	128	5463
getting	203	4647

Table 3.1: The G^2 scores and overall frequencies in the BNC of the top 20 lexical verbs that are significantly more frequent in spoken rather than written language.

Word frequencies taken from the British National Corpus (BNC) (Leech et al 2001) suggests that there are a large number of lexical verbs that are significantly more frequent in spoken rather than written language.when examined using 'the log-

likelihood ratio or G^2 statistic' (Leech et al 2001: 16). Table 3.1 shows the top 20 lexical verbs with the greatest significant difference. It should be noted that a G^2 score of above 6.6 shows a significance of $p<0.01$, so the scores shown in the table are high. The top five lexical verbs with a greater significance in written rather than spoken are *held* - with a G^2 score of 2381 and frequency of 276/million, *became* – 2381 and 304/million, *made* – 2336 and 304/million, *found* – 2304 and 487/million, and *seemed* – 2144 and 238/million. It is interesting that the significant differences are not as high as those associated with spoken language, and they are all past forms. While I would suggest that each of the verbs, those significantly more frequent in spoken and those significantly more frequent in written, would warrant further investigation of the type undertaken here, I have chosen to examine the, often considered, complementary pair of verbs, *come* and *go*.

I had initially planned to examine *know, think, mean, go, say, want, put, come, see* and *like.* I excluded *get, got* and *going* as *get* and *got* can be used to form the passive by functioning as an auxiliary verb (e.g. Biber *et al* 1999: 376, 475) and *going to* is a 'common way of marking future time in conversation (and fictional dialog)' (Biber *et al* 1999: 490). I felt that in having established pre-existing grammatical functions they were not solely lexical verbs. As there would appear to be some question as to whether *thank* can be considered to be a stand-alone verb I decided to omit this verb from the investigation. *Thank* with *you* behaves 'pragmatically and lexically as an analysable formula' that is used predominately as an insert (Biber *et al* 1999: 1083).

Table 3.2 shows the frequency of occurrence of each of the verbs in the spoken texts of each of the ICE corpora, and Table 3.3 shows the same in the written texts. It should be noted that as the ICE corpora all contain the same number of words there is no need to normalise the data for comparison purposes.

Table 3.4 gives 'log-likelihood ratio or G^2 statistic' (Leech *et al* 2001: 16) which indicates the significance of the difference between the frequencies of the verbs in the spoken compared to written components of each ICE corpora. It should be noted that the statistic is influenced by the size of the corpora so direct comparisons with the BNC cannot be made. However the results do show that, as with the BNC, all the verbs have a significantly higher frequency in the spoken component in comparison to the written component in all the ICE corpora.

	ICE-Can	***ICE-GB***	***ICE-India***	***ICE-Jam***
know	4751	2525	2307	4127
think	1982	2443	1558	1800
mean	1102	1638	683	799
go	1441	996	1089	1456
say	927	1047	1266	1460
want	1149	738	613	1022
put	626	516	386	409
come	615	674	1059	1078
see	927	1142	1552	1112
like	3683	1742	2003	2454

Table 3.2: Frequency of verb-forms in the spoken ICE corpora.

	ICE-Can	*ICE-GB*	*ICE-India*	*ICE-Jam*
know	258	308	223	340
think	226	206	85	152
mean	70	77	64	86
go	275	230	251	238
say	211	204	142	159
want	176	145	107	132
put	147	168	111	171
come	216	226	253	253
see	361	442	195	357
like	602	464	609	405

Table 3.3: Frequency of verb-forms in the written ICE corpora.

	ICE-Can	*ICE-GB*	*ICE-India*	*ICE-Jam*
know	3293.72	1195.97	1256.63	2434.64
think	980.79	1425.58	718,74	1049.65
mean	723.89	1186.17	378.20	409.57
go	465.75	255.41	279.98	548.62
say	242.40	330.81	632.94	743.17
want	458.32	231.02	217.14	465.34
put	156.87	72.43	69.86	27.78
come	71.86	88.36	259.00	270.34
see	80.51	101.13	720.44	161.07
like	1387.80	360.50	325.46	916.60

Table 3.4: Log-likelihood ratio spoken in comparison to written of the verb-forms

As my research progressed it soon became apparent that if I wanted to examine the verbs, and their corresponding *n*-grams at the delicacy I wanted to achieve I would need to choose to examine either one of the higher frequency verbs, or two lower frequency verbs from the list. It seemed to me that it would be better to examine two verbs rather than one, so, as *come* and *go* could be considered to be a pair and are often taken together, I chose to continue my research with them.

Come and *go* are often considered to be complementary as they can be considered to be

> 'the two central verbs of motion [… in that] they are the most general, the least explicit, and, from a certain point of view, the most distinctive; the least explicit in the sense that their meanings can be stated without reference to the character of the movement […]; the most distinctive in the sense that their meanings *cannot* be stated without reference to the direction of the movement
>
> Behre 1973: 11.

The generality of *come* and *go* allows them to combine with particles to create multi-word verbs that 'behave to some extent either lexically or syntactically as a single verb' where 'a single meaning selection straddles a major structural boundary' (Quirk *et al* 1985: 1163, Sinclair 2004a: 26). For example, in terms of this research this could be re-expressed as *come* and *go* have collocates that when examined together exhibit distinctive colligations, semantic preferences and semantic prosodies. In the BNC the 20 most prolific verbs that include *come* and *go* combine with 'eight particles (*out, up, on, back, down, in, over,* and *off*) – a total of 160 combinations' to create more than half (50.4%) of the multi-word verbs found by using tagging software that identifies AVP (adverb or prepositional-adverb) particles (Gardner and Davies 2007: 249, 349, 342).

The distinctiveness of *come* and *go* allows them to play a part in the 'deictic anchorage' of the utterance (Fillmore 1971/5: 222) where the 'tripartite relationship between the linguistic system, the encoder's subjectivity and contextual factors is foregrounded grammatically or lexically' (Green 1995: 11). I would argue that the deictic role of *come* and *go* can be seen in terms of the co-selection components of the lexical item. Deictically, *come* typically indicates movement 'from the listener's location to the speaker's location or vice versa', movement from a third-party location to either the listener's or speaker's location, or movement accompanying either the speaker or listener to a location (Carter and McCarthy 2006: 69, 70). From a lexical item perspective, the **locations specific and non-specific (third-party) to the participants** are the semantic preferences, and the **movement associated with the participants or towards their location**, the semantic prosody. Likewise, if deictically *go* is typically an indication of movement from either the speaker's or the listener's location to a third-party location (*ibid*), then again the semantic preferences are the **locations that are specific and non-specific to the participants**, and the semantic prosody the **movement from specific location towards a non-specific location**. In other words the semantic prosody identifies the deictic centre – where the utterance is anchored. However, deictically when the movement involves only third-parties, either word may be used 'depending on whether the speaker sees things from the agent's or the recipient's viewpoint' (*ibid*). In this case, the semantic preferences are the **locations specific and non-specific to the agent and recipient**, and the semantic prosody, by indicating the direction of movement, identifies whether it is from the agent's or the recipient's perspective.

In Sections 3.1 and 3.2, I describe the traditional approach to multi-word verbs and to deixis, as well as giving an overview of deictic shift theory – where the deictic centre is shifted from the perspective of the speaker to another perspective within a narrative world. However, as I have explained above and re-iterate in the summary that ends that chapter, I would argue that both can be easily described in terms of the co-selection components of the lexical item. I would suggest that there is an advantage in describing both multi-word verbs and deixis with the same categories as it gives an inclusive rather than exclusive description of the machinations of language. In more traditional linguistic descriptions, deixis tends to be considered as an additional rather than integral phenomenon.

3.1 Multi-word verbs

Multi-word verbs can be grouped according to the particles associated with the verbs. There is the phrasal verb (verb + adverbial, e.g. *come back*) – transitive and/or intransitive, the prepositional verb (verb + preposition, e.g. *go into*), or phrasal-prepositional verb (verb + adverbial + preposition, e.g. *come up with*) (Quirk 1985: 1161, Biber *et al* 1999: 403, Carter and McCarthy 2006: 431), and there are multi-word constructions that include nouns and adjectives, such as '*to give rise to', 'to be a*

match for', 'to be fond of' and '*to steer clear of'* (Bolinger 1971: 5, Biber *et al* 1999: 403). Bolinger (1971) suggests another category of particle, the *adprep*, which I believe to be relevant to this research. These are 'particles that oscillate between preposition and adverb' (Bolinger 1971: 26). Adpreps are adverbs and prepositions at one and the same time: they are 'collapsed compound' prepositions (Bolinger 1971: 28). It has two connections: the first with the verb where it is split from the prepositional object, and the second with the prepositional object where it is now split from the verb (*ibid*).

In contrast to free combinations, where the lexical verb and the particle have a separate grammatical and semantic status, these multi-word verbs are identified using both structural and/or semantic criteria (Biber *et al* 1999: 403). However, it can be difficult to 'make an absolute distinction between free combinations and […] multi-word verbs', thus it is better to consider them as on a cline from relatively free to relatively fixed (*ibid*) and presume that all verb + particles combinations are potential multi-word verbs until proven otherwise (Darwin and Grey 1999: 75). Additionally, some that function as multi-word verbs can also function as prepositional verbs, as well as free combinations (Biber *et al* 1999: 408).

Semantically, while the meaning of the multi-word verbs is different from the independent meanings of each word of the group, each word contributes 'something recognizable to the meaning of the whole' (Sinclair 2004a: 20): in a sense, they have been re-lexalised as a single unit. Often a multi-word verb can be identified by replacing the combination with a single word verb, but sometimes, especially in informal situations, this can sound like a pretentious circumlocution (Darwin and Grey 1999: 66).

Structurally, there are a number of "tests" that can be undertaken to identify the different types of multi-word verb combinations, but none are can be considered watertight. For example, one should be able to passivise transitive phrasal verbs but where the particles are adpreps it allows nearly all combinations with *go*, but, excludes nearly all combinations with *come*, for example

- he will *go* into the subject carefully*[2]/the subject will be gone into carefully*
- he will *come* into a fortune*~~/a fortune will be *come* into~~[3]*

(adapted from Bolinger 1971: 7)

There would appear to be a lack of agreement about what verb + particle combinations should or should not be included, in that multi-word verbs that are included by one scholar may easily be excluded by another (Darwin and Grey 1999: 75). This research will not seek to classify multi-word verbs associated with *come* and *go*, but it will discuss them in terms of the co-selection components of the lexical item

3.2 Deixis

While *come* and *go* are mainly discussed in terms of *spatial deixis*, this research would suggest that there is some evidence that they are also used in *temporal deixis* and *discourse deixis*. In addition, there is evidence that, in live radio sports reporting, there are shifts in the deictic centres (also know as *origos* or *zero-points*) akin to that identified in *deictic shift theory* (e.g. Stockwell 2002, Tsur 2008, Segal 1995/2009, Zubin and Hewitt 1995/2009, MacIntyre 2007). I begin this section with a description of deixis and I follow with an overview of deictic shift theory. This research will, due

[2] * denotes examples invented not taken from 'used' language
[3] crossing out denotes examples that are normally considered unacceptable

to space and time constraints, seek only to show that there is evidence of deictic shift occurring across all Englishes examined within the live radio sports reporting medium – further research will be needed to study this phenomenon in depth.

Deixis is primarily that feature of language that anchors meaning to the context of the spoken utterance. The deictic terms 'are interpreted in relation to where the speaker is situated', and the location of the speaker is known as the deictic centre (McIntyre 2007: 123). However, it is has also been argued that this can be extended into the written language, applying 'equally well in literary or fictional situations' (Stockwell 2002: 43). The core types of deixis are time deixis – 'to the time of utterance, and to the times before and after the time of utterance'; place deixis 'to the location of the speaker at the time of utterance'; and, person deixis - 'to the identity of the speaker and the intended audience' (Fillmore 1966: 220). Other deixis categories are also recognised such as referential, syntactic and *origo*-deixis: 'deictics whose function is to refer' such as pronouns and the definite article; deictics related to 'participant voice'; and, deictics that operate 'within a certain syntactic frame' (Green 1995: 21). Or, there are social and empathetic deixis – the former indicating the closeness or lack of closeness of relationship between the participants, and the latter the psychological closeness or lack of closeness (MacIntyre 2007: 123-124). There is also discourse, or, in relation to written texts, textual deixis (Fillmore 1971/5: 259, Stockwell 2002: 46). Discourse/textual deixis is used to 'indicate or otherwise refer to some portion or aspect of the ongoing discourse' (Fillmore 1971/5: 289), they are used to indicate to the listener/reader what has or is going to occur such as, in written texts, 'explicit 'signposting' such as chapter titles and paragraphing' (Stockwell 2002: 46).

Basically deictic terms place the speaker at a deictic centre within the context of the exchange. However, when one extends this to literary or fictional situations it can be argued that there is a shift of the deictic centre. The deictic co-ordinates are not interpreted with reference to the speaker, but they are interpreted with reference 'to a deictic centre somewhere within the fictional world' in that we are projecting 'a deictic centre that is different to our own' (MacIntyre 2007: 124).

Deictic Shift Theory emerged from the interdisciplinary work undertaken into cognitive science into deictic approach to narrative (Duchan *et al* 1995/2009: xii-xiii). The theory suggests that a reader interprets a text by taking 'a cognitive stance within the world of the narrative' (Segal 1995/2009: 14). The deictic centre shifts from the 'real-world situation' – the here and now – of the reader and/or author to a particular location within the narrative that is indicated by the use of deictic terms (*ibid*). In cognitive terms, 'the reader tracks the shifted deixis in the text' as if they had been placed in the centre, with the deictic centre of the teller fading into the background: 'the deictic structure presupposes its own story world, and not the current interactional context of the teller and audience' (Zubin and Hewitt 1995/2009: 131). The readers are seeing things 'from the perspective of the character or narrator inside the text-world' (Stockwell 2002: 47). Frequently, in the process of reading a number of deictic centre shifts will be required to interpret the narrative (MacIntyre 2007: 124), and this depends on deictic shift cues within the text, such as 'spatial and temporal locative expressions' (Stockwell 2002: 49). This research will suggest that the deictic shift can also be seen in live sports reporting.

In summary

In the following, I summarise the points pertinent to this research in this chapter. I further analyse the frequencies of the *come* and *go* in Chapter 5, and multi-word verbs, deixis and deictic shift are all identified in terms of the co-selection components in Chapter 6.

- *Come* and *go* are among a small selection of verbs that occur at a significant higher frequency in spoken language in comparison to written language.
- *Come* and *go* are considered to be complementary in that they are both distinctive but also general.
- They are general in that the combine with particles to create new verbs.
- They are distinctive in that they play a major role in identifying the deictic centre of an utterance.
- Deixis can be seen in terms of the co-selection components of the lexical item.
- In literature or fictional situations there is often a shift in the deictic centre to a location within the narrative.

Chapter 4 The Corpora

> *The complexity of the written language is its density of substance, solid like that of a diamond formed under pressure. By contrast, the complexity of spoken language is its intricacy of movement, liquid like that of a rapidly running river*
> *Halliday 1985: 87*

> *To begin with, English is an international language in the Commonwealth, the Colonies and in America. International in the sense that English serves the American way of life and might be called American, it serves the Indian way of life and has recently been declared an Indian language within the framework of the federal constitution. In another sense, it is international not only in Europe but in Asia and Africa, and serves various African ways of life and is increasingly the all-Asian language of politics. Secondly, and I say 'secondly' advisedly, English is the key to what is described in a common cliché as 'the British way of life'*
> *Firth 1968: 97*

The corpora that make up the ICE corpora are unique in that they have a larger proportion of spoken language than written language – a ratio of 3:2 – that has been collected to a common, and thus comparable, design. The corpora are parallel corpora with the texts comprising the same categories containing the same amounts of language. I begin the chapter by contrasting spoken and written language – the difference, the capture of spoken language in what, essentially, is the written medium, and the grammars of spoken language. I end the chapter with an account of the ICE corpora. I first describe the history behind the project to amass a collection of World Englishes, and then I explain why I have chosen the particular ICE corpora for this research. I continue by describing their constituent parts and I argue for viewing these subdivisions as colonies rather than registers or genres. I end by outlining how I prepared the ICE corpora for the research.

4.1 Spoken v. written language

Writing evolved from speech as 'a very far fetched and derivative function of language' (Malinowski 1923: 312). Writing evolved as a result of societies requiring more than just a spoken record: a record that could be consulted that was other than 'people's first-person experiences'; a record that was 'dislocated from the reality to which it referred'; a record that was authoritative 'because only those in charge could command the arcane art of writing' (Teubert 2010: 152). Writing has a social prestige as it is 'associated with learning, religion, government and trade', but 'writing is unnatural in the sense that it must be deliberately taught and learned' (Halliday 1985: vii, Chafe 2006: 56). Speech came first, by many millennia, and speech comes first in the life of an individual, but linguistic theory evolved from the written word not the spoken word (Halliday 1985: vii). There is still a requirement to convert the spoken word into the written word in order to investigate language. This requirement would appear to have created 'a mismatch between linguistic theory which recognizes speaking as primary to writing and much linguistic practice which, in effect, is biased toward the written variety' (Rühlemann 2006: 405).

Corpus linguists currently study spoken language through the medium of the written word. This can be enhanced with annotation, such as phonetic or prosodic information, and, with the advent of multi-modal corpora, a video or a sound track can be sequenced with the transcription. It is nevertheless the transcript, with or without enhancement, which is at the heart of any investigation. No enhancement can replace the actuality of being present at the exchange: the physical surroundings, the circumstances leading to the exchange, the behaviour of the individuals – their facial expressions and gestures, the relationships involved, are all, to varying extents, lost to the researcher (Teubert 2010: 150). It is impossible in principle to separate what is said from the setting in which it takes place, […thus] no annotation system can do justice to the non-linguistic extensions of the discourse' (Teubert 2010: 167).

The linguist must accept that the empirical investigation of spoken discourse is, at present, an investigation of an orthographic representation of the spoken discourse and not 'how texts are meant by their speakers or how they are understood by their hearers' (Teubert 2010: 167). As Wittgenstein cautions, any claim of knowing what is meant by speakers and what is understood by hearers is mistaken. He points out that 'the essential thing about private experience is really not that each person possesses his own specimen, but that nobody knows whether other people also have *this* or something else' (Wittgenstein 2009: 102). However, while the 'the transcribed record of spoken text cannot capture the experience of its regular use' Widdowson 2004: 10), the orthographic traces of the discourse can show us, not 'in abstract terms but in an infinity of examples' the types of linguistic behaviours in which discourse communities participate – how communities negotiate meaning in the here-and-now (Teubert 2010: 170).

The traditional approach to linguistic modelling posits a hierarchical/constituent structure – a structure which derives 'from an act of abstraction away from potential use' (Brazil 1995: 241) which normally takes as its starting point the sentence or clause, and is informed by written language. The advent of spoken corpora has produced spoken language performance grammars – 'the abstracting or modelling grammar by the interaction of data and theory' (Leech 2000: 687) – that are based on this traditional approach, but it has also led to the development of a different type of grammar – linear grammars. Linear grammars are grammars that entertain 'the possibility […] that the rule systems purposeful speakers habitually work with are of a different kind from those that they find in sentence grammars' (Brazil 1995: 13): grammars that are based on the 'unremarkable fact that the events that comprise […] discourse occur one after another' (Brazil 1995: 6). Both Brazil's (1995) *Grammar of Spoken English* and Sinclair and Mauranen's (2006) *Linear Unit Grammar* (LUG) are such grammars.

Initially spoken grammars were divided into those that are advocating a 'totally different model from those traditionally applied to written language' (Approach A), and those whose approach is 'holistic' using 'the same framework of categories, structures, and rules for both spoken and written grammar' (Approach B) (Leech 2000: 688, 689). The former included both the linear grammar approach suggested by Brazil (1995) and the work of 'Carter, Hughes and McCarthy at Nottingham' (Leech 2000: 688); the latter were 'represented by Biber *et al* (1999)' (Leech 2000: 689). Although there are differences in the approaches taken by Biber *et al* (1999) and that of Carter and McCarthy (2006), both the grammars impose a hierarchy and are orientated to the clause, in comparison to the linearity of Brazil's (1995) grammar and the more recent LUG (Sinclair and Mauranen 2006). I would, therefore, suggest that the approaches to spoken grammar could be re-divided accordingly into the more traditional hierarchical constituent-within-constituent and the linear increment-by-increment approach.

Biber *et al* (1999) describe the grammar of spoken language in terms of traditional written grammar while Carter and McCarthy (2006) describe spoken language in its own terms. 'The descriptive framework and terminology' of Biber *et*

al 'closely follows' Quirk *et al* (1985), so that the categories and terms used 'are familiar and unobjectionable to the widest range of grammar users: the 'same "grammar of English" can be applied to both spoken and written language'(Biber *et al* 1999: 7, 1038). Carter and McCarthy, on the other hand, suggest that such a descriptive framework and terminology that is based on written grammars can mean that 'appropriate terms for describing particular features of spoken grammar are not available' and 'what may be considered "non-standard" in writing may well be "standard" in speech' (Carter and McCarthy 2006: 9, 12). As Rühlemann argues those terminologies based on written language are loaded 'with value judgements', and 'new terms and concepts' should be used that 'adequately reflect the conditions and constraints which structure speech in interaction' (Rühlemann 2006: 404).

For example, both Biber *et al* (1999) and Carter and McCarthy (2006) identify *situational ellipsis* as a feature of spoken language. In seeing it in terms of written language, Biber *et al* states that it occurs because speakers drop 'words with contextually low information value', which suggests that in a 'more orthodox sentence grammar' these words would still be there (Biber *et al* 1999: 1104, 1074). However, according to Carter and McCarthy (2006) ellipsis is as an integral part of spoken language. When any ellipsis occurs 'nothing is "missing" from [the message as it contains] enough for the purposes of communication' so, in terms of situational ellipsis, the speaker does not need to explicitly refer 'to people and things which are in the immediate situation' (Carter and McCarthy 2006: 181). As Hughes suggests

> In the spoken channel language users can afford to be extremely economical in the way they construct utterances. Indeed, were they to express their ideas in the full forms […], they would sound like non-native speakers who tend to cling to full sentences as being 'correct' when their communication would be improved by using less complete clauses
>
> Hughes 1996: 14

Where applicable, this research will employ that terminology that strives to avoid value judgments. However, terminology aside, what is common to both these grammars is that spoken language is typically face-to-face, relying on deictic terms and shared knowledge; it is typically interactive, using, for example, questions, discourse markers and vocatives; it reflects the interpersonal, expressing politeness, emotion and attitude; and it takes place in real time employing such features as filled pauses and repetitions and re-castings (Biber *et al* 1999: 1043-1062; Carter and McCarthy 2006: 164-175). Also, and this is integral to this research as the two registers are being examined by means of the same structural model, 'spoken language and written language are not sharply divided but exist on a continuum' (Carter and McCarthy 2006: 164).

The linear grammars, Brazil's (1995) and Sinclair and Mauranen's (2006), also emphasise the real time constraints and the interactivity of spoken language. The grammars are interrelated. As Sinclair and Mauranen developed LUG, they 'felt themselves moving ever closer to' Brazil's position (Sinclair and Mauranen 2006: viii). To them, speech is a purposeful activity that takes place one word after another unfolding in time; and, it is interactive, with 'the communicative value of any item […] negotiated between participants' (Brazil 1995: 4, 6, 34; Sinclair and Mauranen 2006: 27).

I would suggest that there can be no doubt that the advent of corpora of spoken language has shown that to solely model language using only written language is mistaken. Whether from a hierarchical or from a linear perspective, the investigation of spoken corpora has uncovered and continues to uncover new features of language. This is why the spoken components of the ICE corpora are such a valuable assets for linguistic research.

4.2 The International Corpus of English

The ICE-project was the brainchild of the late Sidney Greenbaum. He envisaged a collection of representative corpora of different "World Englishes". The term embodies the pluricentric position that there are different Englishes that deserve 'consideration and recognition as autonomous or semi-autonomous varieties of the language' rather than the monocentric position that there is only the one English 'with all its geographical and social varieties' (Bolton 2009: 241). These World Englishes are 'essentially displaced and discontinuous encodings' that are 'globally scattered […] and unique to our time': they are something other than dialects – 'something less dependant' (Widdowson 2003: 52).

The project arose from the initial desire to compare spoken British English and American English because the only existing available parallel corpora, the American Brown Corpus and the British Lancaster-Oslo/Bergen Corpus, contained only printed material (Greenbaum 1991: 83). The ICE corpora were loosely based on these two previous corpora, each having been collected from material produced in 1961 and consisting of 500 texts of 2,000 words each (Nelson 2009: 737). It was agreed that countries where English was used not only as the first language, but as an official additional language should be invited to participate (Greenbaum 1991: 84). It was also felt that the value of the enterprise would be considerably enhanced if the spoken also had a matching written component collected over the same time period as this would allow comparison 'within each national variety and across national varieties' (*ibid*). By doing this they also hoped to show that 'a common grammatical "core" unites all varieties' (Nelson 2009: 740). There was at the time of conception a minimum expectation 'that the whole of the ICE [would] be computerised and concordanced for lexical strings by the end of 1995' (Greenbaum 1991: 90).

While there are, at the time of writing, 26 research teams worldwide who have either prepared or are preparing ICE corpora, when this research began in 2010 there were only 9 ICE corpora available for research purposes[4]. These were ICE-Canada, -Jamaica, -India, -Singapore, -Ireland, -East Africa, -Hong Kong, -Philippines and –Great Britain (GB). As with all corpus research (see Chapter 5 for further discussion) there is a balance between having too much data, and having too little. In order to try and achieve that balance I decided to choose four of the ICE corpora which are, as stated earlier, ICE-Canada, -GB, -India and Jamaica. The selection is based on their classification according to Schneider's five-stage 'Dynamic Model of the evolution of Postcolonial Englishes' (Schneider 2007). In order to have as broad a selection as possible I have chosen varieties from different stages of evolution plus ICE-GB. ICE-India is between stages 3 and 4; ICE-Jamaica, stages 4 and 5; and ICE-Canada is at stage 5, as is ICE-GB.

Schneider (2007: 12-14) identifies two previous models that have categorised the different varieties of World Englishes: the first distinguishes English as a native language, as a second language, and as a foreign language. The second separates English into those varieties that belong to the Inner Circle – 'the traditional cultural and linguistic bases of English', Outer Circle – 'the institutionalised non-native varieties', and Expanding Circle – 'varieties that lack official status and are typically restricted in their uses' (Kachru 1992: 356). Schneider suggests that both of these models are problematic in that they are 'superficial and fuzzy' when 'establishing categories of linguistic description and classification' (Schneider 2007: 13). While both models have criteria for categorization they do not 'convincingly […] fit problematic cases […] and neither one has attempted to list all countries in a given category exhaustively' (Schneider 2007: 13). Instead Schneider proposes the

[4] this information, and the following information not specifically referenced, is taken from http://ice-corpora.net, last accessed 30 July 2014.

'Dynamic Model of the Evolution of Postcolonial Englishes' which argues for the emergence of postcolonial Englishes as a five-stage evolutionary progression from foundation (stage 1), through exonormative stabilization (stage 2), nativization (stage 3), endonormative stabilization (stage 4), to differentiation (stage 5). The characteristics of each stage are distinguished along four constitutive parameters: the extra-linguistic background – historical events leading to the socio-political situation; identity constructions resulting from this; sociolinguistic conditions – contact, attitudes to language, specific language usages etc.; and, the structural realisation of these in the grammar, lexis and phonology (Schneider 2007: 31, 33; see Schneider 2007: 56 for a detailed description of each stage of the evolutionary process).

Each ICE corpus consists of approximately 1 million words divided into 300 spoken and 200 written texts of 2,000 words each. The texts for the original ICE corpora were collected over a three year period in the early 1990s (Greenbaum 1991: 86). The spoken and written texts are sub-divided into different categories as indicated in Table 4.1 below. The number of texts in each category is indicated in brackets; for example, there are 180 dialogues of which 80 are public, and of those 80, 10 are broadcast interviews. The authors or speakers are over 18, were born (or moved to) the country and educated in the particular English variety at a minimum of secondary school level (*ibid*). However, it was recognised that there would be a need to be flexible and 'others whose public status […] make their inclusion appropriate would also be admitted' (*ibid*). A variety of age groups are represented as well as both sexes but this does not necessarily equate to the language use in the whole population (*ibid*). The spoken material is transcribed orthographically but there are recordings available of the material to allow other transcription protocols to be used (*ibid*).

SPOKEN (300)	Dialogues (180)	Private (100)	Face-to-face conversations (90) Phonecalls (10)
		Public (80)	Classroom Lessons (20) Broadcast Discussions (20) Broadcast Interviews (10) Parliamentary Debates (10) Legal cross-examinations (10) Business Transactions (10)
	Monologues (120)	Unscripted (70)	Spontaneous commentaries (20) Unscripted Speeches (30) Demonstrations (10) Legal Presentations (10)
		Scripted (50)	Broadcast News (20) Broadcast Talks (20) Non-broadcast Talks (10)
WRITTEN (200)	Non-printed (50)	Student Writing (20)	Student Essays (10) Exam Scripts (10)
		Letters (30)	Social Letters (15) Business Letters (15)
	Printed (150)	Academic writing (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Popular writing (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Reportage (20)	Press news reports (20)
		Instructional writing (20)	Administrative Writing (10) Skills/hobbies (10)
		Persuasive writing (10)	Press editorials (10)
		Creative writing (20)	Novels & short stories (20)

Table 4.1: The number of texts in each category of the ICE corpora.

I will now discuss *register* and *genre* in relation to corpora, and the subdivisions of corpora. Different linguists have different opinions as to what is constituted by and what the parameters are of these two terms: Biber (1988: 22, 13) suggests that registers and genres 'are similar (or different) to differing extents with respect to each dimension' where a dimension is a 'strong co-occurrence patterns of linguistic features'; Thompson (2004: 40, 42) describes register as the 'use of certain recognizable configurations of linguistic resources in certain contexts', and genre 'as register plus purpose'; and, Freadman (2012: 660) argues 'genre is destabilized by uptake' as 'as no discursive event is a pure example of any genre'.

In addition how does one, especially with spoken dialogue, decide on the register/genre? As Adolphs states

> it is important to identify external categories for grouping transcripts in a corpus, especially where levels of formality and other functions are concerned which need to be judged against the wider context of the encounter. This process tends to be much more straightforward when dealing with written texts, as many of genres that are used for written corpus analysis are well established, such as fiction versus non-fiction, letters versus e-mails etc. it can often cover a number of themes in one dialogue
>
> Adolphs 2008: 6

And when it comes to subdividing corpora, is spoken-scripted (monologue) just written that is being read or is it spoken language? In the ICE corpora it is considered to be part of the spoken component but the frequencies of the *come-* and *go-*grams would suggest that it is actually written-to-be-read and would fit better with the written component (see Chapter 5).

In order to address these concerns, I would like to propose that instead of trying to allocate parts of corpora to specific registers/genres that they should be regarded as members of colony levels of the specific corpus or corpora under investigation (Hoey 2001: 72-92). Hoey identifies a group of text – 'Cinderella' texts – that are neglected 'in most text theories' which include 'shopping lists to statutes, bibliographies to Bibles' (Hoey 2001: 73). He suggests that these texts 'might be characterised as *colonies*' and then proceeds to define the properties of these colonies (Hoey 2001: 75).

I. The meaning is not derived from the sequence of the constituent parts. For example no meaning can be derived from the sequence of entries in a dictionary.
II. Constituent parts that are adjacent cannot be considered to be continuous prose
III. There is some sort of 'framing context' in which the constituent parts are organised. In a dictionary this would normally be some pages of explanation of terms, organisation etc.
IV. There is no single identifiable writer.
V. Each constituent part can be used without reference to any other part.
VI. Each constituent part can be 'reprinted or reused in subsequent works'.
VII. Constituent parts can be 'added, removed or altered'.
VIII. Many of the constituent parts 'serve the same function'.
IX. The constituent parts can be alphabetically, numerically or temporally sequenced.

(adapted from Hoey 2001: 88)

If one were to consider each text segment of a corpus to be the equivalent of a constituent part such as an entry in a dictionary, a corpus would appear to exhibit the

same type of properties as that of a colony. Meaning is not derived from sequence; adjacent texts are not continuous prose; a corpus normally (and it is best practice that it does) has some form of meta-data (headers, manual etc) supplied as a framing context; there is normally no single writer – although sometimes in, say, a corpus of Shakespeare's plays there is a single writer but it is not a single work; the texts can be used in isolation, can be reprinted or re-used, and they can be added, removed or altered; the texts serve the same function in that they are being used to examine specific nuances of language or discourse; and, they can be sequenced alphabetically, numerically or temporally. In terms of the ICE corpora, each of them could be considered to be a colony. However, this alone does not necessarily help solve the problem of categorising the subdivisions of corpora in a way that avoids the potential quagmire of definition by register and/or genre.

In order to address this, I would suggest that each subdivision of a corpus could also be considered to be a corpus and so, in a sense, the subdivisions come together to create a larger corpus. The ICE corpora can be divided into the spoken sections and the written sections and each section can be considered to be representative, accordingly, of the spoken variety and the written variety of a particular World English, but taken together they are representative of the World English as a whole. If the main corpus – for example, ICE-GB, -India etc – is delineated as the colony *level 1*, and each time the corpus is subdivided it becomes another level – *level 2*, *level 3* etc, the subdivisions can now be categorised without having to refer to register or genre. Using this system, a corpus can be subdivided in different ways that are dictated by the research parameters; for example if age and/or gender are the main interest of the researcher, it would have been possible to sub divide the ICE corpora on this basis.

This research uses the 'common design' subdivisions of the ICE corpora as colonies. The colonies are grouped as in Table 4.2 with the colonies specific to speech shown in red, and those under investigation in this research bold type.

Colonies	
Level 1	ICE-GB, -India, -Canada, -Jamaica
Level 2	spoken, written
Level 3	dialogue, monologue, **non-printed, printed**
Level 4	**private, public, unscripted, scripted**, non-professional writing, correspondence, academic writing, non-academic writing, reportage, instructional writing, persuasive writing, creative writing
Level 5	direct conversation, telephone calls, class lessons, broadcast discussions, broadcast interviews, parliamentary debates, legal cross-examinations, business transactions, spontaneous commentaries, unscripted speeches, demonstrations, legal presentations, broadcast news, broadcast talks, non-broadcast talks, student essays, examination scripts, social letters, business letters, academic-humanities, -social sciences, -natural sciences, -technology, non-academic-humanities, -social sciences, -natural sciences, -technology, press news reports, administrative writing, skills and hobbies, press editorials, novels and stories

Table 4.2: Colony levels and contents in the ICE corpora.

It should be noted that, in using this approach, the spoken data is examined at Colony Level 4 (private, public, unscripted and scripted), and the written at Level 3 (non-printed and printed), thus the spoken data is being examined more delicately than the written. As the focus of the research tends towards the spoken rather than the written data because the spoken component has more instances of *come* and *go* and *come*- and *go*-grams, I am of the opinion that this is not necessarily a problem.

However, the quantitative data would suggest that there is a difference in the frequencies between the non-printed and printed so a more delicate investigation (non-professional writing, correspondence, academic writing, non-academic writing, reportage, instructional writing, persuasive writing, and creative writing) could be of additional benefit.

The ICE corpora selected for this research are all annotated with *textual markup*, but ICE-GB is additionally annotated with *word class tagging* and *syntactic parsing*. ICE-GB also has dedicated software ICE-CUP to enable the researcher to exploit the annotated data. The *textual mark-up* in the written texts indicate the features of the original layout which include 'sentence and paragraph boundaries, heading, deletions, and typographic features' and the spoken texts 'are marked for pauses, overlapping strings, discourse phenomena such as false starts and hesitation, and speaker turns'. While *WordSmith 6* (Scott 2015) recognises the *textual mark-up* annotation so can ignore it when generating concordances and lexical bundles, it does not recognise the *word class tagging* or *syntactic parsing* annotation.

The data has been both standardised and converted in order that it might be used with *WordSmith 6* (Scott 2015). In generating collocations and *n*-grams to be compared across corpora it is important that the input data is consistent otherwise the output data will be incompatible for comparative purposes. A word-list was generated for each of the ICE corpora to check that there were no differences, or anomalies in the data which would impact on the research. It was discovered that the data contained HTML or XML symbols, for example *&eacute* (é), which were included in the word lists. All these symbols were removed from the corpora. An additional problem was also identified with the ICE-GB of a space before an apostrophe where the apostrophe is in place of letter(s) that begin a word (*'s, 've, 're, 'm, 'd* and *'ll*). The spaces were removed.

WordSmith 6 (Scott 2015) is, by default, set to ignore any text within tags <>, but the ICE corpora also has additional content markup and in some cases it includes words which need to be excluded from the data extraction. *WordSmith 6* (Scott 2007) was set to ignore editorial comment (all words between <&> and </&>), extra corpus text (all words between <X> and </X> , normative insertions (all words between <+> and </+>), and discontinuous word corrections (all words between <)> and </)>). *WordSmith 6* (Scott 2015) does allow the researcher to access the original source text from the concordance lines and the tags can then be examined. This is particularly pertinent with spoken language as various discourse phenomena, such as hesitations and overlaps, can be influential in the interpretation of the data. However, I wanted to be able to examine who is speaking, the hesitations, the overlaps, the anthropophonics (laughs, coughs etc) in the actual concordance lines so the corpora were converted to allow this using *WordSmith 6* Text Converter (Scott 2015). Accordingly, different speakers are represented by the symbol $ plus a letter of the alphabet; anthropophonics and unclear words are indicated by double brackets ((*))*;* short pauses (,), and long pauses (,,). Speech overlaps are contained in square brackets [ ], with the beginning and end in curly brackets { } and where there is more than one overlap it is indicated by numbers, such as {3 [3 …3] … [3… 3] 3}.

In summary

I summarise below the salient points from this chapter. I discuss further the linear approach to grammar in Chapter 8 and Chapter 9. I adopt the term colonies in all the subsequent chapters.

- The ICE corpora were chosen as they contain a ratio of 3:2 spoken to written language.

- In order to investigate language the spoken word must be converted into the written word.
- The advent of corpora of spoken language has served to show that spoken grammar is not necessarily the same as written grammar, although the consensus is that they exist on a continuum.
- This research uses spoken terminology that avoids value judgements.
- Typically spoken language is interactive and in real time.
- Traditional grammars describe language in the form of a hierarchical/constituent structure. Recent grammars describe it in linear terms.
- This research describes the different types of language under investigation as colonies to avoid the confusions surrounding register and genre. It also advocates this framework for other research as it is more explicit with regard to the delicacy at which corpora are investigated.
- The ICE corpora chosen represent a broad spectrum of World Englishes in terms of Schneider's characterisations of English Varieties.
- The ICE corpora have been both standardised and converted for the purposes of this research.

Chapter 5 Organising the data

Thirty year ago when this research started it was considered impossible to process texts of several million words in length. Twenty years ago it was considered marginally possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today it is very popular.

Sinclair 1991:1

Surprisingly, in view of the copious quantity of digital data now available, one of the recurrent problems in much of corpus linguistic investigation is the sparseness of the data. 'There is no finite base' when it comes to words and combinations of words, and no 'finite syntactic base' because as a corpus grows in size so does the vocabulary (Wilks 2005: 217). As corpora grow so do the number of new words and the number of identifiable syntactic rules grow (*ibid*). This would suggest that in order for any description of language to be as comprehensive as possible, the larger the corpus the better. However this, paradoxically, can then lead to there being too much raw data for investigation and choices need to be made about how it should be sorted and what should be chosen for investigation, so the linguist can reach a stage where they can proceed to interpretation.

Stubbs (2002: 66-67) distinguishes between different levels of investigation by classifying the data to be observed as first-, second- and third-order. The first-order data is the raw corpus which, at a realistic size, is too large to analyse for significant patterns without further sorting. This first-order data is the data that is the input for software programmes or statistical analysis and from this the second-order data is extracted. This data consists of concordance lines and word lists. However, as he points out, 'beyond a few hundred concordance lines' (Stubbs 2002: 66) the data again becomes too unwieldy for analysis so a further sort of the data can be undertaken using additional software. The result is third-order data such as collocations, n-grams, concgrams, lexical bundles, patterns, with or with out associated statistical information. This research investigates third-order data, specifically *come*- and *go*-grams and their associated statistical information, rather than that suggested by Sinclair, the second-order data of concordance lines.

	ICE-Can	**ICE-GB**	**ICE-India**	**ICE-Jam**
go	1714	1221	1333	1696
come	827	895	1310	1328

Table 5.1: Raw frequencies of *come* and *go* in the ICE corpora.

	ICE-Can	ICE-GB	ICE-India	ICE-Jam
go	1440	996	1084	1455
come	611	674	1059	1078

Table 5.2: Raw frequencies of *come* and *go* in the spoken colonies of the ICE corpora.

	ICE-Can	ICE-GB	ICE-India	ICE-Jam
go	274	225	249	241
come	216	221	251	250

Table 5.3: Raw frequencies of *come* and *go* in the written colonies of the ICE corpora.

Sinclair (2003/2004) suggests examining 30 concordances – 'a screenful', and then a further 30 etc until all the variety of usages can be identified in order to identify the co-selection components of the lexical item. He suggests the concordances should be sorted until the strongest pattern – whether it be a word, grammar, semantic associations or something else - can be identified. This is then interpreted and the process is repeated with the next strongest pattern (Sinclair 2003: xvi). The following screen shots (Figure 5.1) are reproduced in order to illustrate Sinclair's suggested approach. The screen shots show the output from *WordSmith 6* (Scott 2015) of two randomly generated 30 concordances of *come* from the ICE-India private colony. They are then sorted alphabetically giving two screen shots for each of the random selections. The first of the two is sorted to R1 and the second to L1.

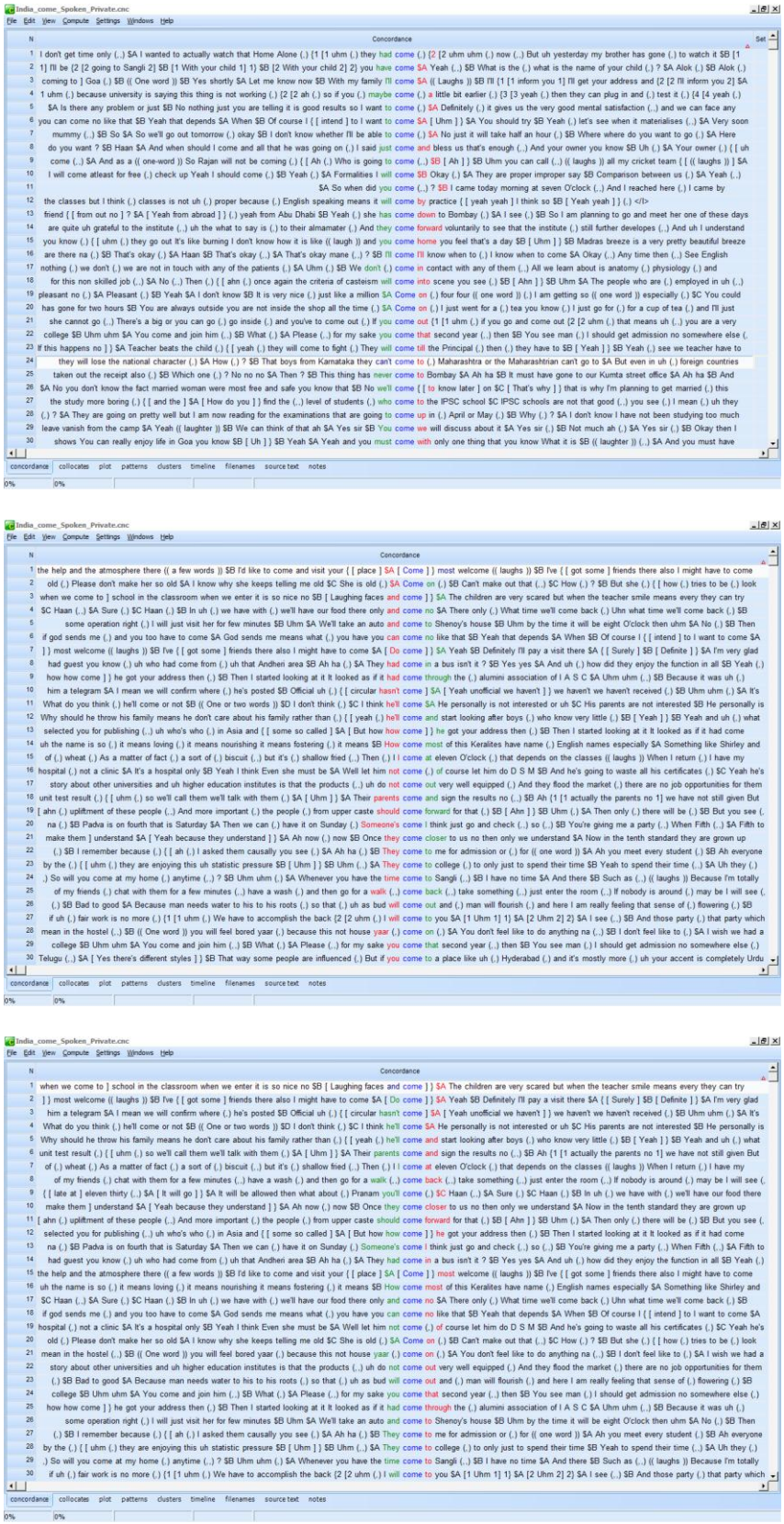


Figure 5.1. Screen shots taken from *WordSmith 6* (Scott 2015) of two randomly generated from the ICE-Indi a private colony 30 concordances sorted alphabetically to R1, then re-sorted to L1.

This approach presents a problem for this research as the two verbs *come* and *go* are polysemous so it is difficult to identify what the strongest patterns and the subsequent patterns associated with them are without some form of further sorting. As mentioned in the introduction, Sinclair identified this as a potential problem when examining *of* as the small samples studied 'showed hardly any consistency in the relative frequencies' of the instances exemplified (Sinclair 1991: 84). This can also be seen in the screen shots in Figure 5.1: in the first concordance sorted to R1 there are 5 instances of *come to* and in the second concordance sorted to R1 there are no instances. This research will show that *to come* is one of the strongest patterns associated with *come* and this is not reflected in the randomly generated concordances. Likewise, in the concordances sorted to L1, they both contain examples of *come out* but neither concordance have *come out with.* This research will show that *come out with* is the strongest pattern associated with *come out.* Again, this is not apparent in the concordances. While Sinclair's methodology, the examination of second-order data, works well with words of medium of low frequency, it is not so good for higher, more polysemous, words. In order to overcome this problem this research examines third-order data.

The data is further sorted into *pre-set collocations* (see Section 2.2.1) with a frequency cut-off level of equal to or above 40 per million. In addition to giving a manageable amount of concordance lines that contain the strongest word patterns this also has the added benefit of reflecting the different frequencies across the colonies as, at this data level, all concordances generated for these pre-set collocations in each colony can be examined. The initial statistical examination of the *come*- and *go*-grams includes all those that occur in the spoken colonies of <u>at least one</u> of the ICE corpora under investigation at a frequency of 40/million. However, as a result of the statistical investigation, the second, more delicate qualitative examination is undertaken on *come*- and *go*-grams that occur at a frequency of 40/million or above in the spoken colonies of <u>all of the ICE</u> corpora.

The overall frequency figures for the verbs are shown in Table 5.1; those for the verbs in the spoken colonies in Table 5.2, and the frequencies of *come* and *go* in the written colonies are shown in Table 5.3. They show that *go* is somewhat more frequent than *come* in the spoken colonies in all the ICE corpora bar ICE-India. In ICE-India the two frequencies are much the same. The frequencies of *come* and *go* in the written colonies are very similar.

	ICE	***Colonies*** **Spoken**				**Written**	
		Private	**Public**	**Scripted**	**Un-scripted**	**Non-Printed**	**Printed**
come	**Can**	211	149	60	191	83	133
	GB	266	190	81	137	69	152
	Ind	540	227	66	226	92	159
	Jam	438	325	68	247	96	154
go	**Can**	660	317	67	396	98	176
	GB	514	219	75	188	79	146
	Ind	678	187	49	170	91	158
	Jam	746	363	61	285	85	156

Table 5.4: Raw frequencies of *come* and *go* across the different colonies of the ICE corpora

As stated, I use pre-set collocations of the *come*- and *go*-grams to generate the concordances. I do this on the basis that Sinclair advocates examining the texts in concordances immediately to the left and right of the node word first and selecting the most frequent to first examine. The pre-set collocates are those words that occur adjacent to *come* and *go* at a frequency equal to or above 40/million. Essentially they consist of the node word plus a collocation (2-gram) or collocations (3-gram, 4-gram, etc) of the node word. By generating first *come* and *go* from the concordance and then using the information generated from them to generate further concordance lines I am pre-setting the collocations of *come* and *go.* The *come*- and *go*-grams can be considered to be *pre-set collocations* of *come* and *go.*

As I indicate above, the *come*- and *go*-grams investigated are those that occur equal or above 40/million times. To a certain extent the frequency cut off I have chosen could be considered to be arbitrary, as I could have equally chosen other cut off points across different colony combinations. However, I do have reasons for the cut off point I have chosen. I have elected to investigate those with a frequency of equal to or above 40/million in the spoken colonies rather than in the individual ICE corpora as the frequencies, when normalised to frequency/million, are greater in the spoken in comparison to those either in just the written colonies or written and spoken colonies combined (see Appendix I and II for normalised frequencies across the colonies and Section 5.1 for an explanation of the tables). This means that more *come*- and *go*-grams are generated that could potentially be of interest in that they could highlight differences between the spoken and written colonies.

The cut off point is the same as the arbitrary cut off point adopted by those theoreticians studying lexical bundles (eg: Biber *et al* 1999). Willis (1990: vi) suggests that the 700 most frequent words in English 'account for around 70% of all English text', the first 1,500 account for 76%, and the first 2,500 account for around 80% (see also O'Keeffe, McCarthy & Carter 2007: 33ff). The cut-off point selected for the extraction of n-gram of 40/million is also equivalent to any n-gram having the same or a higher frequency than 2,500th ranked word of all four ICE corpora taken together. Table 5.5 gives the frequencies/million, -/600,000 and -/400,000 of the 500th, 1000th, 1500th etc ranked words in the ICE corpora used for this study. At the practical level, each of the ICE corpora comprises a million words of which 600,000 are spoken and 400,000 are written which, at the level of 40/million, translates to a minimum frequency of 24 for the spoken corpora and 16 for the written. The amount of concordance lines, at 40/million, available to examine for the less frequent *n*-grams is not ideal, but it was felt that the advantages of this approach are such that the sparseness of the data, while a disadvantage, should be accepted.

Word rank	Frequency /million	Frequency /600,000	Frequency /400,000
500	216	130	86
1000	100	60	40
1500	67	40	27
2000	48	29	19
2500	40	24	16

Table 5.5: Frequencies of word ranks in the ICE corpora.

5.1 Frequencies

Appendices I and II give all the frequencies (normalised to *n*-gram/million where *n* = 1, 2, 3, etc) of all the *come*- and *go*-grams that occur in at least one of the four ICE corpora at a frequency of above 40/million in the colonies spoken and written (Level 2); written-non-printed and –printed (Level 3); and, spoken-private, -public, -scripted and -unscripted (Level 4). These are *they come, to come, you come, come and, come back, come from, come in, come on, come out, come to, come up, will come, come down, has come, have come, I come, come here, come into, and come, had come, not come, we come, come up with, to come and, to come to, to come up, you come to, come back to, come out of, come to know, come to the, and go, I go, they go, to go, we go, you go, go back, go into, go in, go on, go out, go through, go to, have to go, to go to, go to the, can go, just go, go and, go for, to go and, you go to, go back to, will go, go down, go there, going to go, want to go, can't go, gonna go, should go, go ahead, go home, go up, go with, I go to, had to go, to go back, to go in, to go into, when you go* and *go into the.* Table 5.6 is a section taken from Appendix 1 for illustration purposes.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
L2	L1	Node	R1	R2	n-gram	ICE	Extra	no. ICE	Norm Spoken	Norm Written	Norm. Private	Norm. Public	Norm. Scripted	Norm. Un-scripted	Norm. Non-Printed	Norm. Printed
		COME			1	Canada		4	1000	524	1054	902	587	1330	823	427
		COME			1	GB		4	1087	528	1359	1140	755	909	676	480
		COME			1	India		4	1582	614	2543	1363	574	1422	946	510
		COME			1	Jamaica		4	1726	617	2121	1936	650	1696	934	509
	THEY	COME			2	Canada		4	47	12	15	48	10	118	20	10
	THEY	COME			2	GB		4	56	0	36	42	28	119	0	0
	THEY	COME			2	India		4	51	5	89	36	17	44	21	0
	THEY	COME			2	Jamaica		4	110	7	121	119	0	165	0	10
	TO	COME			2	Canada		4	257	119	270	242	186	306	238	80
	TO	COME			2	GB		4	226	162	347	198	168	139	284	123
	TO	COME			2	India		4	270	122	400	282	113	227	350	51
	TO	COME			2	Jamaica		4	328	131	349	363	191	357	292	76
	YOU	COME			2	Canada		4	47	7	60	48	0	63	10	6
	YOU	COME			2	GB		4	50	14	82	60	19	20	49	3
	YOU	COME			2	India		4	115	24	235	48	17	107	51	16
	YOU	COME			2	Jamaica		4	96	15	97	131	48	89	19	13
		COME	AND		2	Canada		4	62	12	70	54	10	97	30	6
		COME	AND		2	GB		4	52	29	72	54	9	53	88	9
		COME	AND		2	India		4	79	17	127	66	9	88	21	16
		COME	AND		2	Jamaica		4	131	30	189	113	48	130	58	20

Table 5.6: Sample taken from Appendix I

Columns 1-5 comprise the *n*-grams: L1 and L2 are the words to the right of the node, and R1 and R2, those to the left. Column 6 is *n* and Column 7, the ICE corpora. Column 9 gives the number of ICE corpora that contain the *come*- and *go*-grams above a frequency of 40/million, and Column 8 indicates (with the word *extra*) those ICE corpora that do not have that particular *come*- (or *go*-gram) above a frequency of 40/million in the spoken colony. Columns 10-17 have the normalised (*n*-gram/million) frequencies of the *n*-grams in each of the colonies.

The figures would appear to show a number of trends:

- The *come*- and *go*-grams, by and large, occur at the greatest frequency in private conversation
- The frequencies for the spoken-scripted colonies seem to fit more with the written colonies, which would suggest that this colony might be better considered as written-to-be-read, rather than spoken.
- There would appear to be a greater differences within the ICE corpora than between the ICE corpora
- *come to know* appears at a high frequency in the ICE-India spoken, and it is barely used in the other ICE corpora. This would appear to be the only instance of the *come*- and *go*-grams occurring in one ICE corpus and not the others at such a high frequency.

The following statistical investigation confirms that the trends identified have statistical significance. It also accounts for the choice made for the second, more

delicate qualitative examination of only the *come*- and *go*-grams that occurred at a frequency of 40/million or above in all of the ICE corpora.

5.2 Statistics

In Sections 5.2.1 and 5.2.2, I first describe the data that I am using, and I then discuss my reasons for choosing non-parametric rather than parametric tests, outlining the possible advantages and disadvantages of this choice. I present and discuss my initial findings in Section 5.2.3. In Section 5.2.4 I examine the pairwise comparisons between and within the ICE corpora. In light of these results I propose a further test that I present and discuss in the Section 5.2.5.

5.2.1 Data

To re-cap: the data is taken from the frequency counts of all the *come*- and *go*-grams in 6 colonies (private, public, scripted, unscripted, non-printed and printed) that occur above or equal to a frequency of 40/million in at least one of ICE-Canada, -GB, -India, or –Jamaica (See Appendices I and II, Columns 9-16). Each *come*- and *go*-gram has 24 counts - one for each ICE corpora in each colony (columns headed 'ICE', 'Private' etc.). There are 75 *come*- and *go*-grams, 32 *come*-grams and 43 *go*-grams. It should be again noted that the counts have all been normalised to *n*-gram/million.

5.2.2 Non- parametric v. parametric

The statistical investigation that follows uses non-parametric tests as the data is not necessarily normally distributed and the number of counts is below 30 for each of the *n*-grams. Non-parametric tests are more robust when identifying significant differences as they are less prone to Type I errors – 'deciding that the null hypothesis is false when it is actually true', but less powerful as they are more likely to make Type II errors – 'deciding that the null hypothesis is true when it is actually false' (Howitt and Cramer 2005: 99). In other words a non-parametric test is less likely to identify a significant difference between data.

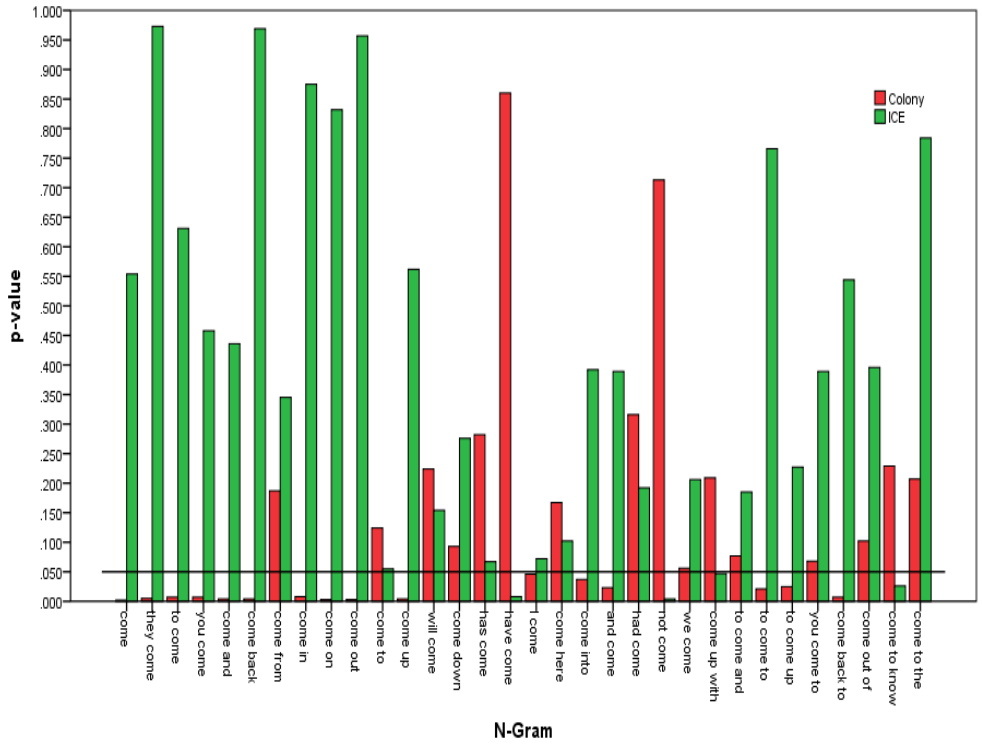
The tests used are the Independent Samples Mann-Whitney *U* Test (within the ICE corpora), the Independent Samples Kruskall-Wallis Test (between the ICE corpora), and the Pearson Chi-Square Test (the further test). I am using a significance level of 95%, where $p \leq 0.050$ (to 3 decimal places). Where applicable degrees of freedom (df) are stated. All the tests are performed using *IBM SPSS Statistics* version 20.

5.2.3 Between and within ICE corpora

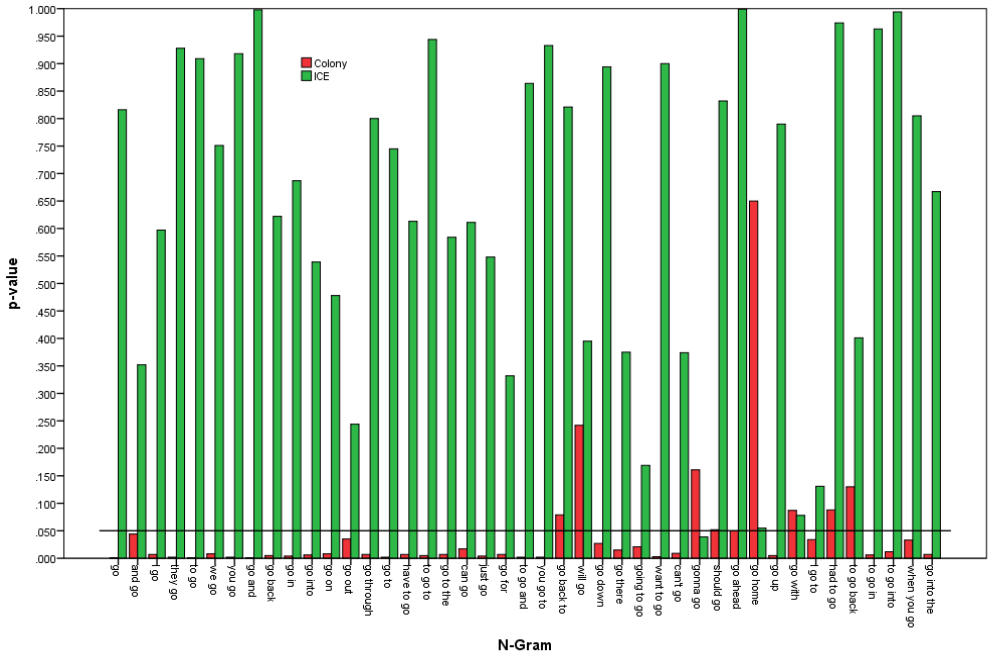
The p-values for distribution of come- and go-grams across colony types were calculated (df 5), with the null hypothesis that the distribution is the same across all the colony types. And, the p-values for the distribution of come- and go-grams across the ICE corpora were calculated (df 3), with the null hypothesis that the distribution is the same across all the ICE corpora. These results are tabulated in Appendix III, columns 2 and 3 respectively. The p-values that are significant ($p \leq 0.050$) are shown in red in the Appendix. They are also depicted as bar charts in Graph 5-1 and Graph 5 2 below.

Graph 5-1 and Graph 5-2 show the values for the come-grams and go-grams respectively. The green bars indicate the significant levels between the ICE corpora, and the red bars, the significant levels within the ICE corpora. The line from the y-axis is at the p-value 0.050 – the significance level of 95%. Any bar does not reach

this line indicates a significant result. The come- and go-grams are ordered on the x-axis according to first the number of ICE corpora in which they are equal to and above 40/million in the spoken component and then alphabetically. Come through to come up are present in all 4 ICE corpora at this level; will come and come down, 3 ICE corpora; has come through to come into, 2 ICE corpora; and, and come through to come to the, 1 ICE corpus: and, go through to go to the, 4 ICE corpora; can go through to go back to, 3 ICE corpora; will go through to want to go, 2 ICE corpora; and, can't go through to go into the, 1 ICE corpus.



Graph 5-1: Distribution of p-values for *come*-grams within (Colony) and between (ICE corpora) in ICE-Canada, -GB,-India and -Jamaica.



Graph 5-2: Distribution of p-values for go-*grams* within (Colony) and between (ICE corpora) in ICE-Canada, -GB,-India and -Jamaica.

I would suggest that the two graphs show that the *come*- and *go*-grams are much more homogeneous between the ICE corpora than within the ICE corpora. The green bars (those showing the p-values for the difference between the ICE corpora) tend to be longer and this length is more randomly distributed across the *come*- and *go*-grams. The red bars (those showing the p-values for the difference within the ICE corpora) tend to be shorter although they would appear to be slightly longer further away from the y-axis.

The original premise of the research was the frequencies of *come* and *go* were significantly greater in the spoken opposed to the written component of the ICE corpora so I would have expected to see a significant difference within the ICE corpora for a proportion of the *come*- and *go*-grams: 50% of the *come*-grams and 79% of the *go*-grams are significantly different across the different colonies. Perhaps the interesting point here is that not all the *come*- and *go*-grams are significantly different? As the *come*- and *go*-grams are ordered according to the number of ICE corpora in which they are present above a level of 40/million this could suggest that there is some correlation here which is explored in Section 5.

5.2.3 Pairwise Comparisons

5.2.3.1 Between the ICE corpora

When it comes to the differences between the ICE corpora; 12.5% (4) of the *come*-grams and 1% (1) of the *go*-grams are significantly different. Appendix IV shows the pairwise comparison of the 5 *come*- and *go*-grams that are significantly different between the ICE corpora. The first column, Column 1 is the specific *come*- or *go*-gram, and Columns 2-7 are the pairwise p-value between the specific ICE corpora: for example, Column 5 is the pairwise p-value calculated for ICE-GB and -Jamaica, and Column 7, ICE-Jamaica and -Canada. The p-values that are significant ($p \leq 0.050$) are shown in red.

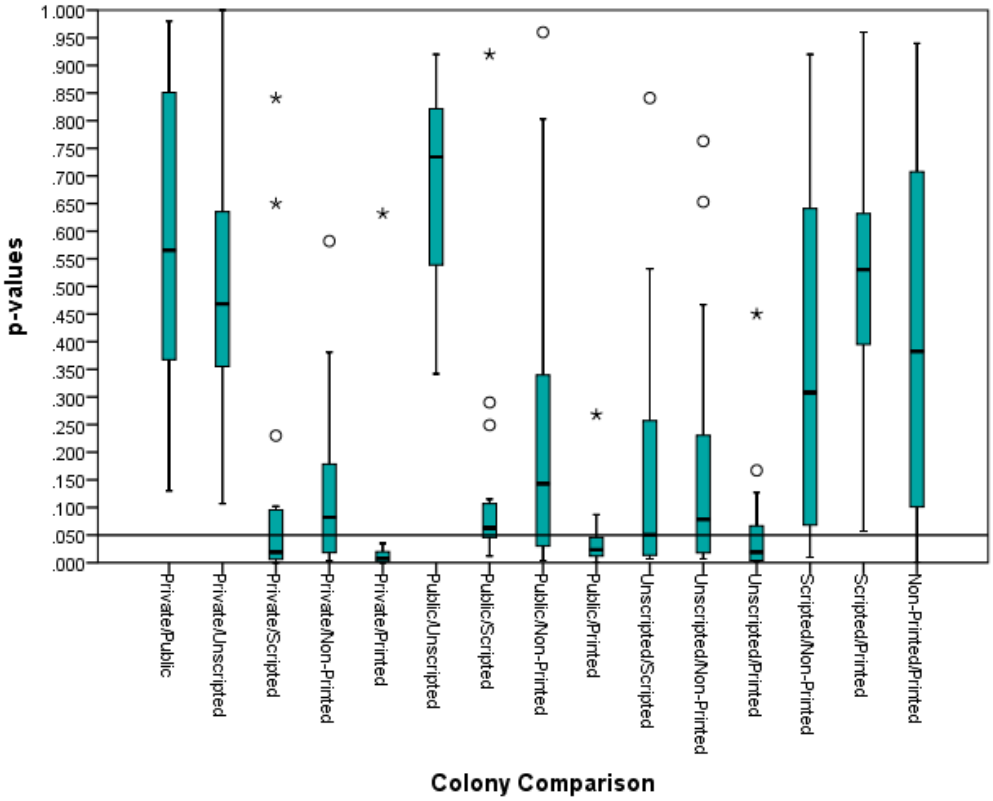
Having examined the frequencies of these *come*- and *go*-grams in the different colonies of each ICE corpus in Appendices I and II, Columns 10-17, I would suggest the following

- The significant difference of *gonna go* between the ICE corpora might be as a result of transcription differences and/or accent differences. It is the one *come*- and *go*-gram under investigation that could be considered specific to spoken language, and it is possible that there were different transcription protocols across the ICE corpora.
- It would appear that *not come* is underused in ICE-Canada, and *come up with* is used more in ICE-Canada and underused in ICE-India. Whether this is specific to these particular corpora or not would need further research with additional corpora.
- There would be appear to be a greater use of *have come* in the ICE corpora that have not reached Stage 5 in Schneider's (2007) the 'Dynamic Model of the evolution of Postcolonial Englishes', where Jamaica is considered to be between Stages 4 and 5, and India between 3 and 4. I would suggest that further research into the uses of *have* between different World Englishes might be advantageous.
- The final *come*-gram, *come to know*, would appear to be specific to the spoken colonies of ICE-India at these frequency levels. It is not used at all the written colonies of any of the ICE corpora, and there are 6 instances each in ICE-Canada and –Jamaica in the Public Colony, and none in ICE-GB.

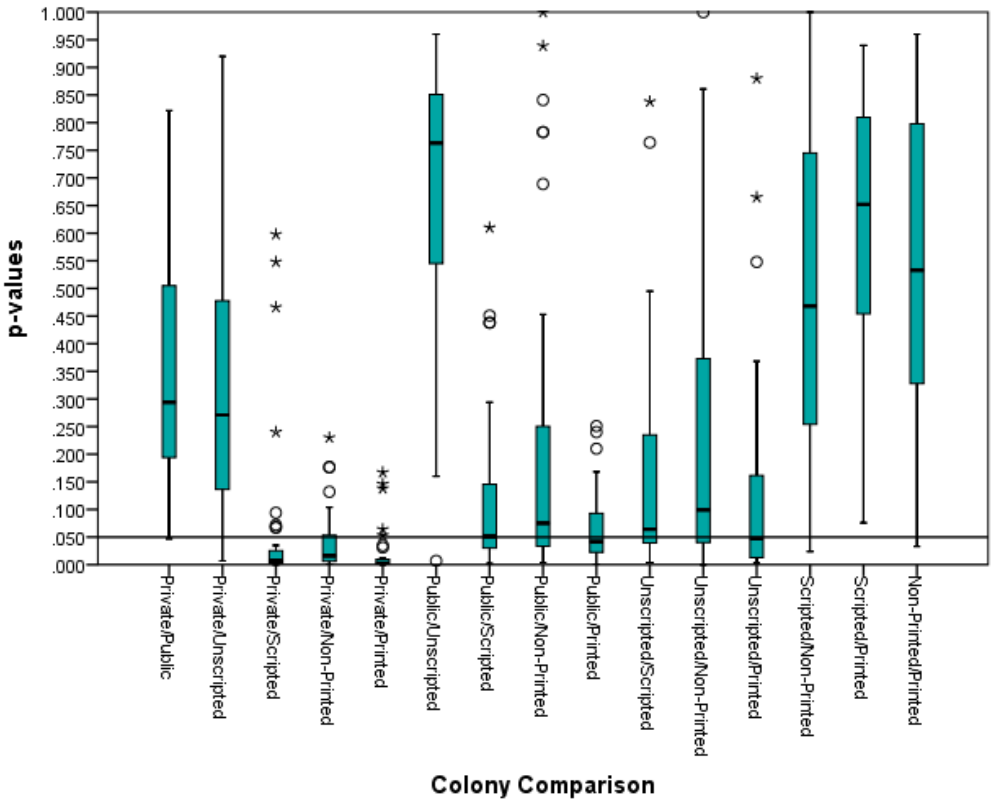
Overall, it would appear that the ICE corpora appear much more homogeneous when examining *come*- and *go*-grams than possibly some research into World Englishes would indicate where the aims have been to delineate difference rather than similarity, such as Schneider's (2007) research into the emergence of postcolonial Englishes.

5.2.3.2 Within the ICE corpora

This section examines the pairwise figures for those *come*- and *go*-grams that are significantly different ***within*** the ICE corpora. 50% of the *come*-grams and 79% of the *go*-grams are significantly different across the different colonies. The figures are shown in Appendix V. The first column, Column 1 is the specific *come*- or *go*-gram, and Columns 2-15 are the pairwise p-value between the specific colonies: for example, Column 5 is the pairwise p-value calculated for Private and Non-Printed Colonies, and Column 11, Unscripted and Scripted. The p-values that are significant ($p \leq 0.050$) are shown in red. These results are shown in the following graphs (Graph 5-3 and Graph 5-4) as box plots. Each box plot represents the pair-wise p-values calculated for all the *come*- and *go*-grams between each different colony of the ICE corpora. The circles and stars are outliers and extreme outliers respectively. Again, the horizontal line is $p = 0.05$ (or a significance of 95%)).



Graph 5-3: Box plots showing the pairwise p-values of the colonies for the *come*-grams.



Graph 5-4: Box plots showing the pairwise p-values of the colonies for the *go*-grams.

A number of observations can be made about the two graphs.

- The distribution of the box plots for each of the pairwise comparisons between the *come*- and *go*-grams look very similar. I would suggest that this need not be the case as there is no reason to expect that the *come*- and *go*-grams would necessarily produce such similar profiles.
- The pair that would appear to be show the least difference in both cases is Public/Unscripted, whereas Private/Public and Private/Unscripted would appear less alike. Private/Scripted shows considerable differences.
- The pair that would appear to show the most difference in both cases is Private/Printed. This I would expect as *come* and *go* are significantly different between the spoken and written components of the ICE corpora and this research is based on the premise that remarkable frequency data needs explanation.
- The medians for Printed/Private, -/Public, and -/Unscripted are all below p = 0.05, and the inter-quartile ranges are relatively smaller. I would suggest that this again emphasises the difference between the spoken and written colonies.
- Scripted would appear to be much more similar to the written colonies. The difference between Scripted/Non-Printed and Scripted/Printed would appear to be less than that between Private/Public and Private/Unscripted. Private/Scripted is the second most different pairwise combination. And, Scripted/Printed would appear to be more homogeneous than Non-Printed/Printed.

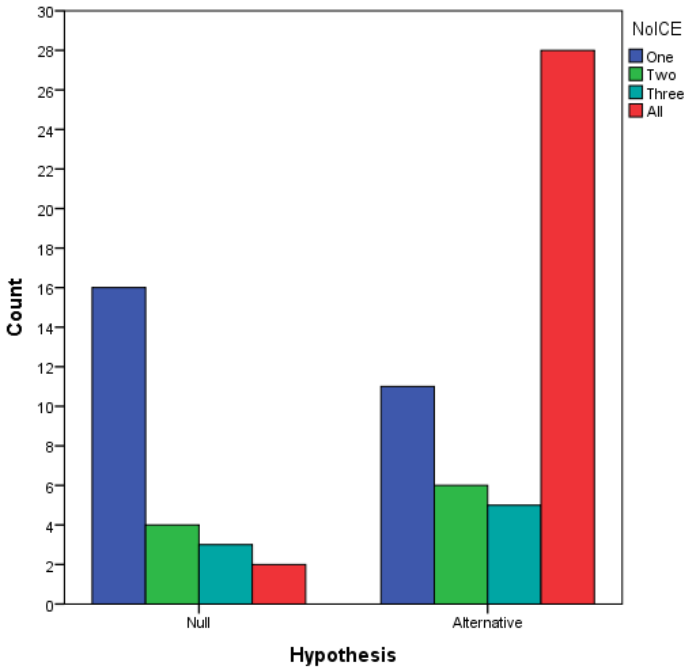
5.2.4 Further Test

The Pearson Chi-Square test was used to see if there was any correlation between the number of ICE corpora in which a *come*- or *go*-grams was present above 40/million in the spoken component as indicated in Column 'No. ICE' in Appendices I and II, and whether there was a significant difference between the colonies shown in Column 2 – 'Colonies' – in Appendix III.

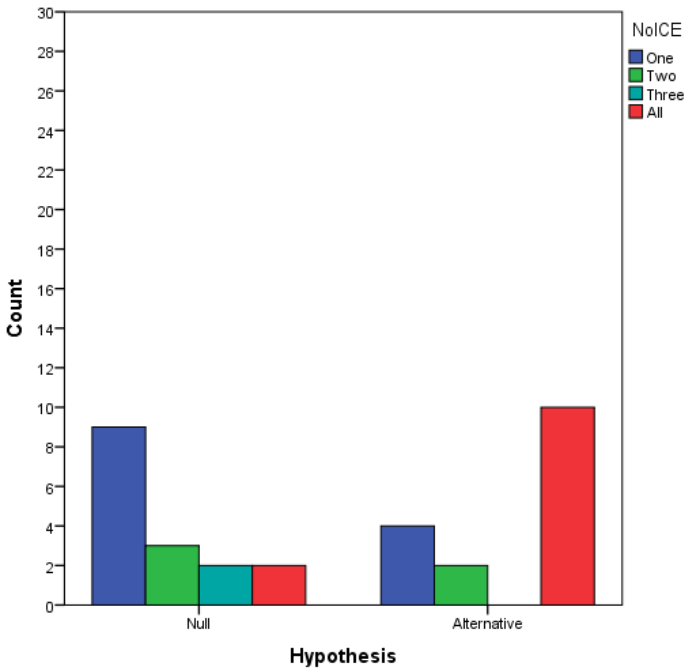
The Colonies p-values were re-coded as either Null (for $p > 0.050$) or Alternative (for $p \leq 0.050$). A significant correlation was found for the *come*- and *go*-grams of p = 0.001 (df 3); for the *come*-grams of p = 0.024 (df 3); and for the *go*-grams of p = 0.008 (df 3). This shows that there is a significant correlation between the number of ICE corpora in which a *come*- or *go*-gram appears and the measure of significant difference between the colonies. The more ICE corpora in which a *come*- and *go*-gram occurs at a frequency equal to above 40/million, the more likely there is to be a significant difference between frequency of occurrence within the ICE corpora.

Graph 5-5 shows the number of ICE corpora a *come*- or *go*-gram is present above 40/million in relation to the significant differences between colonies. Graph 5-6 shows the equivalent for the *come*-grams, and Graph 5-7 for the *go*-grams. Pearsons Chi-Square Test and the graphical depiction clearly indicate that there is a highly significant correlation between the number of ICE corpora in which a *come*- or *go*-gram is present at a frequency of above 40/million in the spoken colony and whether there is a significant difference between the colonies. This suggests that those come- and go-grams that occur in all the ICE corpora are more likely to show a significant difference in frequency between the colonies. On the basis that a significant difference between colonies might be an indication of differences of co-selection components, the *come*- and *go*-grams that occur above 40/million in the spoken colonies in all of the ICE corpora are examined in the chapter that follows. These are *they come, to come, you come, come and, come back, come from, come in, come on, come out, come to, come up, and go, I go, they go, to go, we go, you go, go*

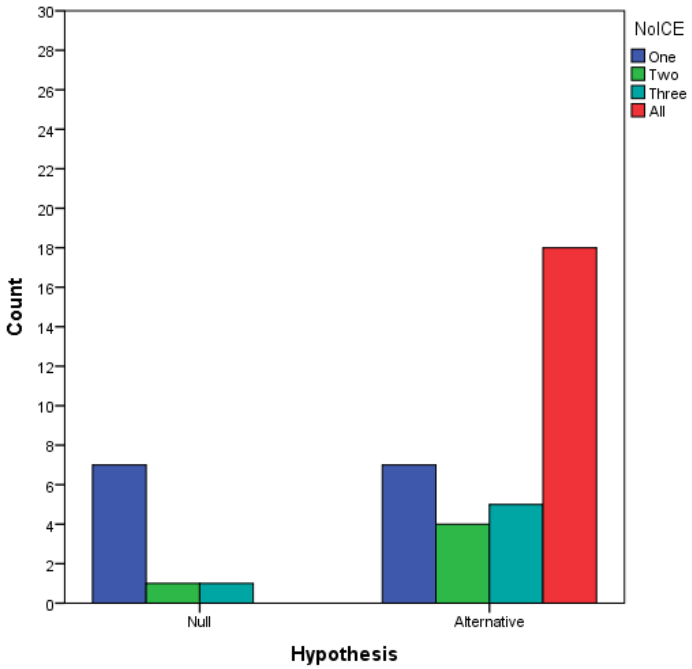
back, go into, go in, go on, go out, go through, go to, have to go, to go to, and *go to the.*



Graph 5-5: The distribution of the number of ICE corpora a *come-* or *go*-gram is present in the spoken colonies above 40/million (One, Two, Three or All of the ICE corpora) and significant differences between colonies where Null indicates a p > 0.050, and Alternative, p ≤ 0.050..



Graph 5-6: The distribution of the number of ICE corpora a *come*-gram is present in the spoken colonies above 40/million (One, Two, Three or All of the ICE corpora) and significant differences between colonies where Null indicates a p > 0.050, and Alternative, p ≤ 0.050..



Graph 5-7: The number of ICE corpora a *go*-gram is present in the spoken colonies above 40/million (One, Two, Three or All of the ICE corpora) and significant differences between colonies where Null indicates a $p > 0.050$, and Alternative, $p \leq 0.050$..

In summary

The results from the statistical investigation undertaken in this chapter are summarised below. These results will be further discussed in Chapter 8 and Chapter 9.

- The quantity of data under investigation in this research is such that the level of investigation should be considered as third-order data.
- The initial analysis of the *come*- and *go*-grams (pre-set collocations) include all those that occur in any one of the ICE corpora equal to or above a frequency of 40/million in the spoken colonies.
- There is a greater homogeneity of frequency between the ICE corpora than within the colonies of the ICE corpora.
- *Come to know* is restricted to the spoken colonies, and is significantly more frequent in ICE-India.
- The *come*- and *go*-grams exhibit similar profiles of difference within the ICE corpora.
- The frequencies of scripted colonies would appear to be closer to the written colonies than those of the spoken.
- There is a significant correlation between the number of ICE corpora in which the *come*- or *go*-grams appears and the measure of significant difference between the colonies.

It must be stressed that frequency differences are not an indication of co-selection component differences. These are examined in the following chapter.

Chapter 6 Examining the data

`But "glory" doesn't mean "a nice knock-down argument,"' Alice objected.
`When I use a word,' Humpty Dumpty said in rather a scornful tone, `it means just what I choose it to mean -- neither more nor less.'
`The question is,' said Alice, `whether you can make words mean so many different things.'
`The question is,' said Humpty Dumpty, `which is to be master - - that's all.'
Alice was too much puzzled to say anything, so after a minute Humpty Dumpty began again. `They've a temper, some of them -- particularly verbs, they're the proudest -- adjectives you can do anything with, but not verbs -- however, I can manage the whole of them! Impenetrability! That's what I say!'
`Would you tell me, please,' said Alice `what that means?'
`Now you talk like a reasonable child,' said Humpty Dumpty, looking very much pleased. `I meant by "impenetrability" that we've had enough of that subject, and it would be just as well if you'd mention what you mean to do next, as I suppose you don't mean to stop here all the rest of your life.'
`That's a great deal to make one word mean,' Alice said in a thoughtful tone.
`When I make a word do a lot of work like that,' said Humpty Dumpty, `I always pay it extra.'
`Oh!' said Alice. She was too much puzzled to make any other remark.

Carroll 2003: 288

As one moves from the examination of the concrete to the abstract co-selection components of the lexical item – collocation to semantic prosody, so one moves from the empirical to the intuitive. Likewise this research moves from the essentially empirical investigation of *come* and *go*, to a principally intuitive investigation. While I would hope that I have not followed in the footsteps of Humpty Dumpty and paid *come* and *go* extra to do as I might wish, the interpretations I make of the patterns I identify are, of course, subjective. It is possible that another might interpret the data in different way. However, many of the patterns would appear to be constant across the ICE corpora, so if the interpretation is wrong the identification of their existence is valuable. In order to emphasise how homogenous these patterns are across the ICE corpora I have included a number of examples for each pattern. Furthermore, there will always be exceptions to the rules, but as this research is rooted in frequent usage, not rare usage I do not necessarily identify them. This research is interested in what is, or appears to be, the normal in the discourse, not what is, or appears to be, the abnormal.

In this chapter, I provide an analysis of the main co-selection components of the lexical items associated with the *come*- and *go*-grams previously selected. These are the *come*- and *go*-grams that appear in all the four ICE corpora equal to and above a frequency of 40/million in the spoken colonies. Whereas Sinclair advocates examining a screenful of concordances (Sinclair 2003: xiv), which gives approximately 20 words either side of the node, I opted to look at approximately 60 words either side; as words vary in length and computer screens vary in size it is not possible to be more precise. In the majority of cases, 20 words would probably have been sufficient, but occasionally, such as the use of *come*- and *go*-grams in sports reporting, and as inserts in speech, the increased number of lines was of benefit. Where I have given examples I have hopefully re-produced sufficient words to make

my case, but as all the ICE corpora used are available for research purposes, the original text can be independently checked. The file designations are listed in Appendix VI.

In all the examples, the different speakers are represented by the symbol $, and in each text they are allocated a letter of the alphabet. Additional information to the text such as when a speaker laughs or there are words that are unclear is indicated by double brackets ((*)). Short pauses are shown as (,), and long pauses (,,). Speech overlaps are contained in square brackets [ ], with the beginning and end in curly brackets { }.Where there is more than one overlap it is indicated by numbers, such as {1 [1 …1] … [1… 1] 1}. The examples given have been copied and pasted from either the concordance lines generated in *WordSmith 6* (Scott 2015) or from the original downloaded corpora, they are not re-produced by re-typing. This means that any typographical mistakes are as the originals. I have not corrected them.

I begin the analysis by introducing two additional co-selection components that are used in the examination of the data, and I also replace the term semantic prosody with semantic force. I then discuss the deictic co-selection components previously identified in Chapter 3. Next I introduce the hypothesis that the examination of the data would suggest that semantic forces appear to act in tandem. I progress with the examination of the data by first establishing that while 'familiar idioms', for want of a better term, can be found in the data as they occur very infrequently across the data they are not pertinent to this research (Sinclair 1999: 157). I then analyse, in terms of the predominate co-selection components identified, the key post (after) pre-set collocates, the *come*- and *go*-grams that begin with *come* or *go* and are followed by the collocation. These are *go and, come back, go back, come from, come in* and *go in, go into, come out, go out, go through, come to, go to* and *go to the,* and *come up.* I follow these with the corresponding selection of the ante (before) pre-set collocates; those that begin with the collocation and end with *come* or *go.* These are *to come, to go, have to go, you come* and *you go.* I then discuss particular types of *come*- and *go*-grams that I have identified in the previous sections; these are discourse managers, sports reporting and replacement speech/thought verbs.

The discourse managers (or discourse deictics) are those *come*- and *go*-grams that explicitly refer back or forward to some aspect of the discourse, these are *come back, come in, come on, come to, come up, I go,* and *go back.* (Discourse managers should not be confused with discourse organisers: discourse organisers implicitly organise the discourse.) The sports reporting section explores the idiosyncratic use of *come* and *go* in live sports reporting. The replacement speech verbs *and go, I go, they go we go,* and *you go,* are used instead of actually speech verbs such as *say* and are chiefly confined to the spoken colonies of ICE-Canada. The chapter ends with a summary of the main conclusions reached. Appendix X consists of a tabular representation of the analysis outlined in this and the following chapter.

6.1 The co-selection components

While there has been considerable deliberation in the literature with regard to collocation and semantic prosody, there has been less with regard to colligation and semantic preference. As a result of my research, I would suggest that the definitions of both colligation and semantic preference need to be adjusted. With this in mind, I have added *structural preference* as a co-selection component of the lexical item, while keeping colligation to indicate word class, structural preference is the preference for particular structures, that both include traditional grammatical structures and other structures that have not necessarily been considered before. These can include such features as hesitation and ellipsis.

I would agree with Stubbs that semantic preference should be interpreted to include the topic (semantic sets) of the discourse, or more importantly a section of the discourse, as it can completely alter the semantic force of the *come*- and *go*-grams. Additionally, I suggest that collocation should be considered to be a sub-category of semantic preference when describing the co-selection components of the lexical item, but as a pre-set collocate it has its own identity. The former is a descriptive category and the latter is a part of the node and has been identified on the basis of frequency of occurrence with a different node – in this case either *come* or *go.*

Again, as a result of my research, I am also introducing the concept *discourse preference* – in certain circumstances the co-selection components are restricted to a particular form of discourse that goes beyond topic such as spoken or written, or a particular ICE corpora.

Finally, as I have undertaken this research I have become more and more uncomfortable with the term semantic prosody. Although it must be said that each time there has been an attempt to change or modify or define the term it would appear to have been unsuccessful, I have decided to replace semantic prosody with the term *semantic force*. Semantic as it is coded within the language, and force as the reason for using it is to effect a particular meaning – 'how […] it is to be taken' (Austin 1975: 73). I would also suggest that the classifications suggested by Sinclair and Mauranen of message (M), interactive organisation (OI), and text organisation (OT) could be also considered to be different types of semantic force. The reason for drawing on a particular set of co-selection components can either be to construct a message, to interact with the other participants of the discourse, or to aid the organisation of the text.

6.2 Deictic co-selection components of come and go

It would appear that there is a semantic preference and semantic force of deictic *come* and deictic *go* that is associated with straightforward physical movement and is often present with the *come*- and *go*-grams investigated. It is the additional co-selection components not these deictic categories that are described in detail in the following sections. The deictic semantic preference for *come* is a **locations specific and non-specific to the participants** – specific in the sense that they are at the location or will be at the location, and that of *go,* **a locations specific and non-specific to the location of the participants** – unspecific in the sense that it is a new, or third party, or vague location**.** This gives a semantic force for *come* of **movement associated with the participants or towards their location**, and for *go,* **movement from a specific location towards a non-specific location**. This then leads on to the observation that the subsequent semantic force of the *come*- and *go*-grams under consideration have their foundations, or are layered with, in the deictic semantic forces associated with *come* and *go.*

6.3 Ante and Post co-selection collocations

The research will show that not only do the semantic forces have their foundations in the deictic semantic forces for *come* and *go,* but the semantic forces of the post and ante *come*- and *go*-grams work in tandem, individually adding to the overall semantic force but not combining to create a new semantic force. In other words for, say, *to come back* the semantic force of the ante *come*-gram *to come* is added to that of the semantic force of the post *come*-gram *come back* giving a semantic force that solely incorporates these two forces for *to come back.* There is no additional semantic force indentified. It is the sum of the two semantic forces rather than a new semantic force.

Unless the linguist had examined the post and ante *come*- and *go*-grams individually, they would not necessarily have identified this. This would suggest that it is the words that are selected for the node that have a bearing on the semantic forces that are identified. While this characteristic is investigated in detail in the chapter that follows, I do include some examples where it is relevant in this chapter. In Chapter 7, I narrow the focus to *to go to, come and, and go* and *come and go* in order to show that the co-selection components of the lexical item are influenced by the choice of node.

Additionally, and I believe that this might also associated with the selection of the node and I will discuss this in Chapter 8, the ante pre-set collocates would appear to consistently have co-selection components that also occur before the node, and the post pre-set collocates to consistently have co-selection components that occur after the node.

6.4 Location

All the co-selection components identified can be found in the spoken colonies, but not all of them can be identified in the written colonies investigated. There would appear to be no co-selection components that are particular to the written colonies, while there are a number that would appear to be particular in the spoken colonies. Of course, this does not mean that there is no specific written colony co-selection components, but the evidence would suggest that it is far less likely to happen. This, I would suggest, might show that innovation in language has a tendency to occur within spoken language not written language. However, it should be noted that the ICE corpora do not include any computer-mediated discourse (CMD) and it is quite possible that innovation in language could also be occurring here. So further research will be needed to ascertain if this is so. If one were to only study written language one might exclude much that is important and only found in spoken or CMD.

6.5 Familiar idioms

Familiar idioms are 'actually very rare' and this research is all about what is very frequent so while one should be aware that they exist in the data they are too infrequent to examine in any constructive way (Sinclair 1999: 157). Amongst the *come*- and *go*-grams examined there is evidence of familiar idioms. These are *come*- and *go*-grams that I, as a native speaker of English, would consider to be familiar to me as idiomatic phrases. This is a somewhat subjective viewpoint but as yet there seems to be no better approach to their identification (see Cermak 2001). Examples of these *come*- and *go*-grams are *hard to come by, come on board, touch and go,* and *go out of our way.* They are illustrated below in Examples 6.1-4. There is one example of *hard to come by* in the ICE-Can public and ICE-GB printed colonies; one example of *come on board* in the ICE-Can public colony; one example of *touch and go* in the ICE-Can printed and ICE-India private colonies; and, one example of *go out of our way* in the ICE-India private colony.

> pour the meatball and tomato sauce mixture into the middle . FOR PHOTOGRAPH TURN TO PAGE 71 . ITALIAN MEATBALLS (2) This variation on the meatball theme was originally made with veal , but both in America and in this country veal can be ***hard to come by*** and turkey breast makes a surprisingly satisfactory substitute
>
> **Example 6.1. ICE-GB:W2D-020**

$A Okay (,) And you then eventually Is it after that you took a position with Percy Grey-Tootoo or Redside { [((word))]
$B [Uh (,)] } no not right away I went with an insurance company which a friend of mine was working for He asked me to ***come on board*** with him as a sales rep (,) and I worked with him for a year or so and it wasn't (,) quite what I to my liking And uh then I went on with uhm (,) Percy Grey-Tootoo

Example 6.2. ICE-CAN:S1B-064

million (U.S.). " We had a very severe cash prob lem," said McIntyre. Restructur ing the business appears to have paid off - Wang reported a very small profit in the first quarter of ended Sept. 30, 1991, and is hoping to build on that, al though, as McIntyre is aware, " it's going to be ***touch and go*** for a while." McIntyre attributes the turn-around to the refinement of Wang's business internally and not to the response of its custom ers. " I would not want to trick anybody into thinking that people have flocked to our door and begun to buy just because we've cleaned up our act," he said.

Example 6.3. ICE-CAN:W2B-033

$A Yeah (,)
$B Become (,) since I've dealt successfully with some of the children See if we love them (,) if { [we pamper a little]
$A [You had to do that na] } that is yeah
$B No if we pamper them little
$A Yeah (,)
$B But (,) sometimes no we have to ***go out of our way*** (,) { [yeah (,) for these childrens especially (,) though others they will think that the teacher is going out of the way means we have to give them (,,) many means whatever (,) at first they have not { [done home work]

Example 6.4. ICE-IND:S1A-085

6.6 Post pre-set collocates

6.6.1 go and

Go and is predominately colligated with a verb that occurs immediately after the *go*-gram. I would also argue that there is a structural preference of **surplus**, by this I mean that the *go*-gram is extra to the exchange and could easily be omitted. I would suggest that its inclusion serves to emphasise the following verb and isolate it from the preceding verb. The semantic force of *go and* is **movement to undertake and amplification of the following action.** In Example 6.5 the speaker describes his reaction to journalistic intrusion by exhorting forcefully to the journalist that he should *go and do some real fucking journalism.* I would suggest that the *go and* amplifies the request. Similarly, the *go and* intensifies the statement that the speaker wishes to *kill everybody* who comes between *good friends* in Example 6.6, and, I would suggest, it also emphasises the fact that speaker *actually* did not pay in Example 6.7.

I'm not saying anything to you about my private life I said and

I don't know what you're talking about anyway He said (,) I said and if you call yourself a journalist I said He said oh I'm a journalist So if you're a journalist ***go and*** do some real fucking journalism then He said I didn't pick this phone up to be insulted I said well fuck off then ((other speakers laugh

Example 6.5. ICE-GB:S1A-052

$A Friends we can tell (,)
$B Yeah whatever
$A And you are supposed to be more closer with friends (,)
$B Surely (,)
$A Uhn (,)
$B But friend also should be truthful (,) isn't it
$A Friend is truthful (,) you don't have to doubt about it (,) I will ***go and*** kill everybody who tries to bring split between the friends (,) good friends (,,)
$B Yeah (,) { [my (,) Kiran I have a class now dental anatomy (,,)
$A [My] }
$A I think you should go for your dental anatomy class because (,,) I don't want you to bunk (,) Dr Kumar's class (,,)
$B Oh ho (,)

Example 6.6. ICE-IND:S1A-054

$A And uh Cam Jeffreys joins me again You know Cam I was down uh at the Old Port last night and the C word was was the common uhm the the C letter I should say uh was the theme as it turned out I didn't actually ***go and*** and pay the three fifty that you have to to get on the pier (,) to go see the show I know Lauren Elliot was down there last night (,) and I was there too late for him so I decided not to go But I walked around (,) and first I saw a street performer uh playing some Cat Stevens music And then th

Example 6.7. ICE-CAN:S1B-039

6.6.2 come back

At its very basic, the semantic preference of c*ome back* is **some previous, usual, or routine state or place** and **a specific point in time**, and the semantic force is **movement to return to a previous/usual/routine state or place at a specific time**. The semantic preference can be implicitly suggested rather than explicitly stated in the surrounding text. In Example 6.8, Gorbachev returns to Russia *a week ago* – although Russia is not explicitly stated but presumed to be his normal place of residence. In Example 6.9, the writer is planning to return *home*.

$A When President Gorbachev returned from Japan a week ago he appeared to be in big trouble His Tokyo summit looked like a waste of time and money He'd abandoned his collapsing country for a whole week to fly a huge retinue of government ministers advisers and experts half way round the world and ***come back*** with nothing (,) no promises of Japanese aid for the Soviet economy (,) no breakthrough in bilateral relations (,) For the Soviet public watching his journeying on television it was an insult While he his wife and his dozens of officials enjoyed lavish guesthouses and expensive Tokyo hotel

Example 6.8. ICE-GB: S2B-040

How are Appa & Amma? What are Appa's immediate plans? (and post-retirement plans?) What about his trip to Switzerland? Do write in detail. I am planning to ***come back*** home in the middle of Oct., get married & go for my exams to UK in 1 st week of Jan ' 94. Accumulate your leave & keep yourself free during Nov-Dec.

Example 6.9: ICE-IND:W1B-010

There is also evidence to suggest that there is a different action, before *come back* to that after the come-gram. It is used as part of a **list** of events that are taking place, or going to take place – a structural preference. It would appear that when it does occur in a list, it performs as a type of pivot. The actions before and after *come back* often appear to be of a different nature. In Example 6.10, they return from the seminar before going to the airport and the students in Example 6.11 attend the college for three years, but when they return they experience problems.

ssion is over (,) are you going (,) uh to that place I mean just (,) uh well from you have come or you want to have a round here
$B No no no I'm not going anywhere (,) Well Shastri and I were planning (,) {1 [1 ah (,) we'll keep the bags here (,) {2 [2 ah (,) so that from the seminar then we'll just ***come back*** (,) {3 [3 okay (,) pick up our bags (,) {4 [4 okay (,) and then (,) you know move to the airport

Example 6.10. IND-ICE: S1A-096

Every year you should try to know at least two lecturers because at the end of the day you might need them for references So you have to chat to get to know them As the students come here for three years and when they ***come back*** and they need something from their lecturer they say they don't know any of the {2 [2 lecturers and lecturers don't know any of them which is bad You have to know at least two of your lecturers every year so at the end you have six lecturers who you can draw on who you can go to and say l

Example 6.11. ICE-JAM: S1A-058

6.6.3 go back

Whereas the semantic preferences of *come back* are **routine and usual places or states at specific times**, the semantic preferences of *go back* appear much less specific. The semantic force is thus **a return to somewhere or something that is slightly vague/non-specific location or point in time**. The speaker, in Example 6.12, suggests that the listener could look at the *poets of the seventeenth century*, and the *legends* in Example 6.13 are rooted in *the epics* – somewhere in the stories of the past, and it is not apparent where the plane actually landed in Example 6.14

fty years now and yet we still seem to think of ourselves in some way as countrymen Why do you think that is
$C We need our countryside more because we are an urban society And there's still this linkage between landscape and the romantic imagination You can ***go back*** to George Herbert and Wiltshire poets of the seventeenth century and uhm nymphs and shepherds of Drayton and uh all that sort of thing and it goes

through to the images of the scenic posters of the Second World War when they were telling us your country is worth fighting for

Example 6.12. ICE-GB: S1B-014

nto oblivion (,) The river which was a kilometre broad gush of deep blue water (,) is now just a sand bed (,) A report from a Trivandrum correspondent (,) John Ulanand (,,)
$C Bharat Pura is Kerala's longest river flowing from the Western Ghats into the Arabian sea (,,) Legends surrounding the river ***go back*** to the epics (,,) It has not only been the key to the prosperity of northern Kerala (,) but also an integral part of the people (,) and their culture for centuries (,,)

Example 6.13. ICE-IND: S2B-017

$A [But uh] }
$A I don't know ((one-word)) Basdeo said that that that uh the the Bajan government had denied landing rights to such a plane { [(,)] Well (,) one would need more detail Did it just turn and ***go back*** to Tripoli or did it land somewhere else Was it really en route to Trinidad If so why did it land in Barbados first
$B [Yes] }

Example 6.14. ICE-JAM: S1A-091

6.6.4 come from

The semantic preferences of the physical locations from which people are moving from can be classified as **notable locations** – a *board meeting* (Example 6.15), places where they inhabit – *London* (Example 6.16), or, and this predominately occurs in ICE-India, where they were born – their **roots** (Example 6.17). In other words, the discourse preference is ICE-India where the semantic preference is **roots**. I suspect that this is because it is culturally more important as might be suggested by the example. Where the semantic preference indicates a physical location the semantic force indicates a **movement from that place to the current location**.

Well it so happens that this week is an extremely busy week for everybody at the university (,) with university meetings (,) And uh (,) as you heard (,) Doctor Chen Yang has ***come from*** a board meeting (,) P V C Green said he dashed over here from another meeting (,) I suspect Professor Nettleford is going to sneak out to a meeting (,)

Example 6.15. ICE-JAM: S2A-027

$B Eighty-seven (,,) ((all-laugh)) My name is Amanda I ***come from*** London
$B I live in London
$A I live in London I am a secretary
$B Yes That's very (( )) (( ))
$A What's I'm

Example 6.16. ICE-GB: S1A-014

$B No doubt
$A They *come from* low status of life (,) uh from villages (,,) they are living standard is very low (,) { [yes (,) and (,,) they are actually exploited (,) by the aristocrat classes (,)
$B [Yes] }
$A Because you see Premchand was a true (,) novelist in its sense He just painted rural pictures in its real

Example 6.17. ICE-IND: S1A-006

However, when the semantic preference is more abstract in the sense of **coming from a source**, the semantic force indicates that **whatever or whoever come from the source appears to make a difference for the better or for the worse**. It is important and it matters. In Example 6.18 there *are profound dangers* resulting form *joining the nation with the state*; and in Example 6.19, it is the diverse nature of the Jamaican people produced by their different origins that produces *this grand language that we call patois.*

Political structure : What has also become clear from the analysis based in identification theory are the profound dangers which *come from* joining the nation with the state . Once a cultural , therefore psychological , community becomes concordant with a political system , that political system takes a quantum leap from being a purely functional power entity into being a psycho-political entity .

Example 6.18: ICE-GB:W2A-017

rectly and ((speaker-A-chuckles)) we have heard some differently and therefore say them differently and end up with what we call Patois((speaker-A-laughs)) It it's it's a combination of English of uhm (,) of many languages from Africa Remember we *come from* many places in Africa so we have little uhm Ghanaian we have some West African we have some East African words and all of those words get pitched into this grand language that we call Patois

Example 6.19. ICE-JAM: S1A-039

There does appear to be an exception to the semantic force that the thing that is coming from the source has a benefit or makes it worse: this has a discourse preference of ICE-GB, and is specific to peoples' origins. This is a semantic force that could be considered to be culturally based – **it is what one contributes and not where one comes from that counts**. It is difficult to be specific about the semantic preferences as they are quite broad but, I would suggest, that they are **peoples' origins and whether they are working in Great Britain** (Example 6.20).

It does not matter who you are or where you *come from* It is what you put in and what you achieve which counts A white-collar job is not superior to another kind of job it's just different And your job need not limit your horizons You're as likely to meet the window cleaner as the bank manager

Example 6.20. ICE-GB: S2B-035

6.6.5 come in and *go in*

Although a number of the *come*- and *go*-grams examined would be classed as multi-word verbs *come in* and *go in* are the most diverse and, I would suggest, worth considering in conjunction with co-selection components. Their very diversity is probably why multi-word verbs are so problematical to the language learner.

The colligation of *in* can be adverbial, adprep, free particle (as part of a free combination), or, in the case of *come in*, ***elliptic***. As an adverbial it is followed by a prepositional phrase; as an adprep it is difficult to tell whether it belongs to the verb or to the following noun phrase – is it an adverbial followed by a noun phrase, or part of the following prepositional phrase; as a free particle it begins and is embedded with the following prepositional phrase; and, as an **elliptic** it appears that there is an *in* missing.

Looking first at *come in*: in Example 6.21, it is not obvious whether the *in* is attached to *come* or forms part of the prepositional phrase with *the water*. This example also shows the extension of the usages of *come* into the third person singular that has occurred in the Jamaican English. Example 6.22 shows *come in* followed by a prepositional phrase *on Tuesday* and the example that follows, Example 6.23, is that of *in* as an adverbial.

```
But now I'm learning that you must breathe {12 [12 through your
mouth 12] So I'm listening to all the tips and you know
everybody come in the water You say ha I don't know how to swim
and everybody give you a little tip (,)
```

Example 6.21. ICE-JAM: S1A-020

```
$B Are you going to be in on Tuesday then or just leave it till
Wednesday
$A Uh (,) I might come in on Tuesday depending if I've got
anything to do It is a D s it is a D six I thought it would be
There you are D' you know how much these cost
$B Yes I do
```

Example 6.22. ICE-GB: S1A-008

```
In his unfinished ' 'Autobiography', Manley wrote, 'Welfare
tapped a deep stream of middle-class interests in Jamaica. We
had no trouble in building a good staff. School teachers were
anxious to come in and were prepared to do so on terms that
meant no financial benefit to them but answered a deep feeling
that rural Jamaica needed special help.'
```

Example 6.23: ICE-JA:W2B-016

Where it would appear that there is an elliptic *in* there is a semantic preference of **objects** and/or **types of objects,** producing a semantic force of **placing the objects into types or orders of types**. In Example 6.24 the text could read 'come in in other colours', and in Example 6.25 the double *in* actually is present.

```
$B Uhm (,,) (( word )) maybe (( two words ))
$A Maybe Or maybe if I got it in a more toned colour or
something (,,)
$B What
```

$A Well does that ***come in*** other colours or (,)
$B Ya I think these are some of the other colours

Example 6.24. ICE-CAN: S2A-037

When you think of the Moslems in India the mind tends to jump straight away to the Moguls They ***come in*** in fifteen twenty-six But there's a period before that which can approximately be called the Sultanate (,) And there are a number of manuscripts which probably belong to this Sultanate (,) period or the Sultanate tradition and this is one of them (,) Uhm it's suggested here that it was made

Example 6.25. ICE-GB: S2A-059

Turning to *go in*: in the first example, Example 6.26, *go* and *your car* both utilise the *in*, and in the second, Example 6.27, *go in* is followed by a prepositional phrase that does not utilise the *in*. There are also occasions where it is not a phrasal verb at all the *in* is free. This can be seen in the sport reporting in Example 6.28, the following prepositional phrase *in the five thousand metres* could be omitted from the text and it would still be understood. In the final two examples, Examples 6.29 and 6.30, the first *in* is an adverbial and the second an adprep. This is a good example of how the addition of one word, *there*, can change the designation of the particle. The speaker could easily add the word *there* to the first example, and remove it from the second – this would, I would suggest, change the emphasis in the utterance but not the meaning. The addition of *there* acts in terms of semantic force to **emphasise the place into which the movement is to take place**.

doxically enough, this speed was attained only because the data were not fully computerised for arriving at the total population figure of India: hand tabulation by millions of census enumerators made this possible. As they say in New York, often the fastest way to get to a place is to walk and not ***go in*** your car. This is true of the Indian census also: hand tabulation is faster then the electronic computers! But one must hasten to add that there are limits to hand tabulation and

Example 6.26; ICE-IND:W2B-013

Uhm Jacqueline Oh I have it here (,,) Okay twenty-first it's Amelie but I think Jackie is going in too with her (,)
$B Ya cos they have to be two
$A Ya but they said it wasn't working out that way But anyway (,) ((sniff)) (,,) Jacqueline is gonna ***go in*** with Amelie on the twenty-first And then you and Mike here (,,) Twenty-second Jacqueline is gonna be all by herself so I might go in there (,,) John and Mike here And the third Gerald and Ja Gerald and Jack

Example 6.27. ICE-CAN: S1B-079

no further change in the competition (,) He still leads from Samuels and Agipon (,) so I don't think that uh jump by Edwards uh just going up at the moment in fact is going to be better No it isn't It's sixteen metres thirteen (,,)

$A Five laps to ***go in*** the five thousand metres We've had the same leader throughout Nigel Adams

Example 6.28. ICE-GB: S2A-007

$A My kids don't have to wear uniforms and I ***go in*** and I get (,) raincoats and (,) t-shirts and sweats and pants (,) and it just makes l and underwear and socks it just makes life easier to buy it all (,) at one shtore

Example 6.29. ICE-CAN: S1B-035.

else to fix it (,) You don't cos you not you don't you don't get any skills They don't teach you any skills to do that {1 [1 (,) 1] because (,) basically (,) my T V something was wrong with it and I had to go (,) to an electrician to {2 [2 fix it 2] because I don't have the skills to fix it I could ***go in*** there and I messed up I would mess it up (,) and then basically I have to buy another one so I had to (,) go to the electrician {3 [3 (,) 3]

Example 6.30. ICE-JAM: S1A-047.

I think the following two examples are also worth introducing to show how problematical multi-word verbs can be. In Example 6.31 *go* is used as an euphemism for lavatory – a non-specific reference – and thus is not part of a phrasal verb, and Example 6.32 shows how the phrasal verb *go in* could be possibly mistaken for *go into.*

and then they go (,) Then he got married to a girl from his own (,) relations (,) that is uh (,,) his mama's daughter (,) cousin sister (,,) And even then (,) she was having a similar kind of torture (,) torture means I cannot explain you to what extent that torture was That girl (,) she had to (,) ***go in*** that uh common (,) lavotory In Bombay (,) in the slum areas there are these common toilets and common bathroom (,) So she had to go there (,) She used to come back window she used to call someone (,) She used to tell them that please give me some chapati or some bread to eat (,) I mean it was

Example 6.31. ICE-IND: S1A-069.

d means to address the social needs economic needs of these persons because as (,) detective said he has gone in he has seen the need and I just like he is saying I am not advocating that because we are in need we are to jump on this bandwagon but when persons are really really desperate and when we ***go in*** to investigate we have seen the need we understand because we have cases where (,) parents have children who are very ill They need the medication (,)

Example 6.32. ICE-JAM: S1B-040

6.6.6 go into

There is a semantic preference of **something important about the place/time/state, or and important event**. This, in turn, implies a semantic force of **movement into something important**. The *jury room* in Example 6.33 is obviously important as is the decision that will emerge from the room; and, the colonisation of places in Example 6.34 is not something to be dismissed. However, where there is reluctance to go into something, often this is considered trivial and not worth mentioning, as in Example 6.35.

gain and I'll mention it one (,) last (,) time That every ingredient necessary to sustain the charge against each of the accused (,) must be proved by the Crown beyond a reasonable doubt (,) With respect to each ingredient every ingredient there's no obligation on the accused whatsoever Now when you ***go into*** that jury room (,) I want you to read them together (,) and use your good old-fashioned common sense and take whatever time you need (,) to reach your verdict (,) until you are all each and every one of you satisfied with it

Example 6.33. ICE-CAN: S2A-070

$A In the first place America doesn't have the attention span for colonialism ((laughter))
$B Have what
$A Doesn't have the attention span for colonialism {1 [1 (,) 1] I mean the British and the French are {2 [2 quite willing to ***go into*** a place 2] and {3 [3 stay 3] for two hundred years (,,) {4 [4 After ten after after 4]
$C [1 ((laughs)) 1] 1}
$C [2 ((laughs)) 2] 2}
$C [3 Stay there for 3] 3}

Example 6.34. ICE-JAM: S1A-010.

digress (,,) At that time Mayfair was a modest farm well removed from the cramped and crowded conditions of the walled City of London and for those of you know of the history of Mayfair you will know that they had a May fair there and all sorts of curious activities took place none of which I would ***go into*** in such a distinguished audience as this (,) However in the seventeen twenties as people became increasingly fed up with the overcrowded and insanitary conditions of the City they sought to live in rather more attractive and pleasant atmosphere of Mayfair (,)

Example 6.35. ICE-GB: S2A-045

6.6.7 come on

Of course, like all the other *come*-grams, *come on* does have a semantic force of **a movement to or from a specific location** as in Example 6.36 below. It is also employed as a discourse manager as described in Section 6.5.

$B No they have all the time and yes (,) see we don't have time (,)
$A That's right

$B You have ***come on*** a short holiday
$A That's right Kanna
$B You have met them in the Bank you can't be going to their home (,) The other day I met them in the Bank (,) { [okay (,) right yeah
$A [Okay] }

Example 6.36. ICE-IND:S1A-058

However, the two most frequent usages of *come on* are distinguishable by the subsequent colligation and in terms of spoken grammar they would be classed as inserts (see Section 2.2.2). The first is followed by a semantic preference, while the second does not.. The semantic preference is for **a correction to/reformulation of the previous utterance** the current speaker. The semantic force for this semantic preference is **to exhort the other person(s) to question their thinking – perhaps to re-think something, or to think like them, and often in the process they will require some demonstrable evidence that this has happened in that the receiver of the exhortation will do what the speaker wants**. In other words, it is a more polite way of saying "re-think this/think like me, and do it". When *come on* is used with no following semantic preference, as the whole speaker turn, it has a semantic force of **you don't say.** It is less emphatic than the first example, acting as a backchannel in the exchange. The former is illustrated in Example 6.37 and 6.38, and the latter in Example 6.39.

$A [Tried to explain it Matthew tried] } to explain it to you [you know]
$B [((laughs))] }
$B Well of course he wouldn't ***Come on*** he's only five ((laughs))
$A Yes But ***come on*** we expect our children to understand irony { [((laughs))]
$B [Linguistics] } { [((unclear-words))

Example 6.37. ICE-GB: S1A-018

hing the sun set in the evenings. " What are you thinking about?" Shannon asked me, turning off at the corner and heading down to the beach. I shook my head. " Nothing." drawing + I look over at him. " How are you feeling?" " Oh ***come on***, Shan, don't fret. I feel absolutely fine ! Loosen up," he cried. "

Example 6.38: ICE-IND:W2F-009

$A Mhhm
$B And she was telling me the evening nurses about the problems sh her son is having at school (,) and uhh she's she's fighting with her professor and she's thinking of quitting school and I think (,,) the writing's on the wall for that woman
$A ***Come on***
$B Her son who's who's had problems before (,) obviously and all of her children are younger than Madie I mean her eldest is younger than Madie Not like he's thirteen
$A Oh ya
$B { [That's the tough part]

Example 6.39. ICE-CAN: S1A-002.TXT

6.6.8 go on

Within the ICE corpora, *go on*, at its very basic, can be found to have a semantic force analogous with that of *go* – **a movement to a non-specific location**. If there is a structural preference of post noun phrase it appears that the *on* is shared between the *go* and the following noun phrase in the same way as *in* appears to do in *come in* and *go in* as previously discussed (see Section 6.2.5). The following Example 6.40 is a good illustration of this.

$A [((laugh))] }
$B Probably uh stay ((word)) (,) camping and go hiking Or my dad had a good idea {1 [1 (,,) we should 1] get into cross-country skiing for the {2 [2 winter 2]
$A [1 What was that 1] 1}
$A [2 Oh ya 2] 2} but I mean
$B We could ***go on*** cross-country ski trails (,,) cos there's real nice trails apparently { [like super] nice
$A [Oh really] }

Example 6.40. ICE-CAN: S1A-083.

Go on also has a semantic force of **continue**; this occurs when it is has structural preference of a following verb phrase or nothing to follow, and/or a colligation of a preposition or adverb. It is used to encourage someone to continue doing something (Example 6.41), to indicate that someone continued communicating in someway – such as saying (Example 6.42), as well as to just indicate a continuation of a process (Example 6.43).

$C Here we are (,) ((Peter)) what are you having
$D Some of that please (,,)
$C Anything on it (,)
$B Uhm (,) try the meringue I think Well ***go on*** then
$C Meringue on meringue (,) (( )) $A Are we to start
$C Yes (,,)
$B Well I wonder what's happened to ((unclear-words)) today (,,)
$A { [Lizzie] where did you put that Artemisia

Example 6.41. ICE-GB: S1A-(23)

What was stupefying however was to hear Pierre; ***go on*** to say that having your fundamental democratic right (,) your right to vote stolen from you was not an attack on democracy (,) but attending a unity rally was I mean that was mind boggling and there there's no logical way to defend that position other than the fact that it was dictated to him

Example 6.42. ICE-CAN: S1B-031.

The next stage in the cycle is are Advanced Level examinations after which the individual will usually ***go on*** to university. During these fifteen years of learning it can be observed that the amounts that can be spent on tuition and examination fees are quite substantial. A family that has one child attending school may find this affordable but when there are more children to support it can prove to b

Example 6.43. ICE-JA:W1A-005

Where *go* is followed by *on my own* (Example 6.44), I would be inclined to suggest that it is not *go on* that should be considered to be the pre-set collocate, but *on my own*.

$B On being called (,)
$A On being called by the police (,)
$D (,,)
$A Now can you give any explanation to the (,) to the (,,) can you give any explanation Mr Angale as to how why you did not ***go on*** your own to the police station right from twenty-eighth April (,) till sixth May (,) to report the matter (,) as you were the eye witness of the incident (,,) ?
$B Sir there were so many people (,)
$A Mr Angale I'm talking about you (,) Forget about the whole people
$B I didn't uh (,) I mean

Example 6.44. ICE-IND: S1B-062.

6.6.9 come out

When *come out* is used to describe, has a semantic preference of, physical movement, the semantic force indicates **the movement towards a specific place**. However, when the semantic preferences are for **non-physical movement and noteworthy outcome** the semantic force is **whatever does *come out* makes a noteworthy change to the situation**. In the first example which expresses physical movement, Example 6.45, dead animals are shaken out of a salt cellar – admittedly unpleasant but not a noteworthy change. The semantic preferences of phrases that are an indication of noteworthiness such as *a beautiful part of a loving relationship* and *tastier than the ordinary one* are shown in Examples 6.46 and 6.47. The equal relationship is not out of duty but out of love; the cake is an improvement on those cooked previously. What is also worth noting in this example is that both speakers seem to be contributing to the semantic force in that speaker B appears to prospect some sort of noteworthy difference in the baking of the cake which speaker A confirms.

No no not here At school
$D Oh right (,)
$B I didn't know they'd go for salt
$A Yeah You find little little uhm you shake the salt cellar and there're little sort of black blobs ***come out*** and if { [you look carefully you see the legs] start wandering about but they're dead definitely dead
$C [He said oh no it's the pepper] }
$D You're lucky the salt was flowing in my opinion { [Fancy complaining about ants

Example 6.45. ICE-GB: S1A-055

ild is evident because of her happiness with her husband and ability to mutually experience the joy of a family. Parvati can be very strongly contrasted with Mohan to exhibit great gender differences, Mala and her white husband have more of an equal relationship. The child for Mala does not seem to ***come out*** of any duty for for her husband, rather a beautiful part of a loving relationship. Children for Mohan and Parvati was a duty

rather than an act of love. It was this duty that Dr. Bridget was able to play upon to help Parvati escape her strong father in-law.

Example 6.46: ICE-CAN:W1A-020

$B No oven (,) where do you get the oven from ?
$A Uh granny has an old desi oven
$B Desi oven (,) on the stove
$A Yeah (,) You can put it in
$B In the cooker and you keep it on the stove
$A Yeah (,)
$B How does it ***come out*** like (,)
$A Very nice (,,) tastier than the ordinary one (,) And what about your new house (,) every thing is settled down na (,)

Example 6.47. ICE-IND: S1A-040

6.6.10 go out

The semantic preference associated with *go out* is **further significant action.** The semantic force of *go out* is to **leave a place in order to do something that is significant** – one does not just *go out* and do nothing. Pete Sampras *wants to wipe the smile of Greg Rusediki's face* in Example 6.48; and, in Example 6.49, the writer bemoans the fact that skills have been lost since the time when the *PWD* could do something significant.

against Pete Sampras Even when he was losing he managed to (,) maintain that smile I wonder sometimes if whether or not that's a psychological technique that that he's e- employing to to keep his spirit up uh when he plays tennis
$B Well Pete Sampras was was quoted as saying that he really wanted to ***go out*** during Wimbledon and wipe { [the smile off] ((laugh)) (,) Greg Rusedski's face so
$A [wipe the smile off ((laugh))] }

Example 6.48. ICE-CAN: S1B-039.

Whatever happened to the days when the PWD had sufficient skill in house to ***go out*** and build a road or a bridge or a building once the budget was approved? Now, it seems one cannot move without consultants of all kinds. No wonder the money cannot spread far.

Example 6.49: ICE-JA:W1B-028

6.6.11 go through

There is a semantic preference for a **neutral, problematic or unpleasant final outcome,** and the semantic force of *go though* is a **change from one state to another, where the final state has the potential to be problematic or unpleasant or at the very least neutral**. In Example 6.50, help is needed to cope with a *nightmare*; in Example 6.51, the advice is that angiography can have its problems; and, in Example 6.52, there are a number of stages to be negotiated before the *full potential* is realised.

saying within uh within an hour the uhh person involved in your incident will be freed she the person is getting out Uhh is everything alright They're working with the person (,) helping them to find solutions to the problem If the problem is security we might find a place for them And help them to ***go through*** this

nightmare
$A So those four officers are are contact people in both directions I mean if if a problem arises from the side of the woman or the citizen they can get in touch with them

Example 6.50. ICE-CAN: S1B-048.

$A Then he said another test we will take
$B Uhm uhm
$A And if necesary we will do angiography
$B Yeah (,) Angiography is the last thing I won't advice you to ***go through*** (,) lightly (,)
$A Lightly means
$B I know you will see the all process of that you are really a cardiac patient to ***go through*** the (,) uh angiography
$A So that will depend on the kind of uh (,) { [yeah (,) results that the ((one word)) tests show
$B [Yeah] }
$B I hope so

Example 6.51. ICE-IND: S1A-068

$B Hm
$A Okay How are you coping with your partner now How is that rose
$B Uh it's it's the rose is struggling basically still (,) yes uhm but like like uhm most things in life {1 [1 (,) 1] it has to to ***go through*** (,) different stages before it reaches that full potential {2 [2 (,) 2] uhm So basically the relationship has a lot of room for improvement {3 [3 and uhm 3] I can see how it could actually be a beautiful relationship
$A [1 Mhm 1] 1}
$A [2 Mhm 2] 2}
$A [3 Mhm 3] 3}

Example 6.52. ICE-JAM: S1A-090.

6.6.12 come to

When c*ome to* involves actually physical movement the semantic force is comparable to that of *come* – **movement towards a specific place related to the speaker** – see Example 6.53. However, when the following noun phrases – colligations – have a semantic preference of **end state** there is a semantic force of **movement from one state to an end state.** Examples of this usage include a progression towards a state in the future where the conclusion is reached (Example 6.54), a thing has stopped moving (Example 6.55), or *closure* is reached (Example 6.56).

$P Ha right may be you know the one that I met was two years ago
$T Bhattacharjee (,)
$P Bhattacharjee that is it (,) he came to
$T ((One or two words))
$P C I E F L there was a (,,) seminar and later in our own Kolhapur he had ***come to*** Guwahati
$T We've four centres (,) I was working in the E L T I English language teaching (,,)

Example 6.53. ICE-IND: S1A-019

It has to be made
$B Ya Ya it has { [to be made]

$A [For your] } sanity
$B For my sanity plus I mean I I've received money to do it { [and]
$A [Oh] } that's right
$B It's been an ongoing thing
$A Right right right So it does have to ***come to*** some kind of conclusion { [and then you're]
$B [Oh ya absolutely] } and and then I think I'll be in a in a better position to know how I feel about (,,) uh (,) my own sort of practice

Example 6.54. ICE-CAN: S1A-033

ng the majority (,) of the Crown case He knew what was not being said (,) and in my submission must have known of its importance Your Lordship knows from Mr Caine that the matters did not ***come to*** light because at some late stage (,) the breakdown in communication (,) was realised (,) that it did not ***come to*** light because it came belatedly from the customs officers and through their counsel (,) It came to light because word was given to defence counsel from defendants (,) and the matter was put forward to Crown counsel

Example 6.55. ICE-GB: S2A-068

Well I have nothing else to say hold up just want to know that I no longer feeling sorry for myself I got closure re my shitty ass relationship and I have ***come to*** the grand conclusion that I really dont hate men and love women all that much I mean they are nice but not that nice.

Example 6.56: ICE-JA:W1B-015

When the colligation that follows *come to* is a verb there is a semantic force of **progression towards a time when an action is going to occur in the future.** *Come to* appears to shift the beginning of the verb action to a specific point that is in the future in relation to the time framework in which it is used. It is in the *twenty-first century* that the changes in *Canada* will ascertain how it will be defined (Example 6.57); the money has been donated to educate the *Dominican amazons* to *value their natural heritage* (Example 6.58); freedom happened which was then followed, in the subsequent years, by the country occupying *a respectable position* with regard to industrialisation (Example 6.59). The semantic force could also be defined in terms of deixis. It is form of temporal deixis.

Canadian people (,,) In areas such as immigration and cult culture for instance (,) I think Canadians must be able to see how these changes (,) would in fact directly benefit them (,,) because all of these changes will have a profound impact on the future of our nation as a whole and how Canada will ***come to*** be defined in the twenty-first century I must restrate restate that the New Democratic Party believes that a strong Canada requires a strong (,) national government

Example 6.57. ICE-CAN: S2B-025

tion , clubbed together and raised $22,000 to fund research into the life-style and requirements of the two beautiful Dominican

amazons . The money was donated to the ICBP to administer . One priority is to start an educational programme to win the hearts and minds of local people so that they will ***come to*** value their natural heritage .

Example 6.58: ICE-GB:W2B-028

(,) And for number of years (,) following the Nehruvian (,,) outlook (,) this society has built itself (,) It is not as if we are a new country (,) like any African African country in dark continents (,) We are a country who have built in last (,) uh uhm so many years after freedon (,) to ***come to*** occupy a very respectable position so far as industrialisation goes (,) so far as our economy goes (,,) We are in a mess because of the debt that we took (,) Who took the debts (,) ?

Example 6.59. ICE-IND: S1B-054

6.6.13 go to and *go to the*

By far the majority of instances of *go to* have, as structural preference, a following noun phrase. This is underlined by the fact that the *go*-gram *go to the* also has a frequency of equal or above 40/million – the *the* indicating the beginning of a noun phrase. The noun phrase has a semantic preference of **a location that is non-specific to the participants in the exchange**. There are rare examples of a verb following *go to,* with *sleep* being by far the most frequent. In the instances where *work* follows *go to,* I have designated it as a noun rather than a verb as the structural preferences of *go to* would seem to indicate this. The semantic force for both *go*-grams is **physical movement to a location non-specific to the discourse participants.** It is always a physical movement not an abstract one. In Example 6.60 the movement is *to a shop.* In Example 6.61 the first *go to* prospects the verb *sleep,* while the second the noun *bed.* The Example 6.62 shows *work* following the *go*-gram. The final two examples in this section, Example 6.63 and 6.64 illustrate *go to the.* In the first the students are required to go to the *morgery* [sic] and the *dissection hall*, and in the second, an example of abstract movement, *Mister Patterson* is seeking *a mandate* from *the public.*

a combination question comes (,) otherwise (,) and moreover a saree and a blouse piece are not substitutes (,) ((laughs)) they are compliment (,) they are compliment and not substitutes and therefore you can't have law of substitution there Yeah (,) so actual life what happens actual life is you ***go to*** a shop (,) you ask the various prices and things like that You see the colour this that etcetera and so on (,) And you know how much you have the money and (,) and therefore (,) then you can order them

Example 6.60. ICE-IND: S2A-032

$A Quiz you don't watch (,)
$B No (,)
$A It's lovely yaar (,,)
$B I don't (,) I I I ***go to*** sleep very (,) means { [I ***go to*** bed very early]
$A [I think it's very sad] }
$A You see that this one uh (,) { [uhn (,) that ((two words))

Example 6.61. ICE-IND: S1A-052

$B Well ((four or five words)) cos at one point in time I used to be able to get up at ten o'clock in the morning (,) and drink a forty by myself and go to work (,,)
$A You'd get up and have a forty of (,) booze for breakfast and go to work
$B Yeah and then I'd ***go to*** work
$A { [Just kin]
$B [A guy] } can't handle himself in alcohol well then I (,) don't (,) you know ((rest of utterance))
$A So you're telling me you can drink forty ounces and nothing
$B At one point in time yeah

Example 6.62. ICE-CAN:S1B-063

nking about (,,) dissection hall itself they really get scared and that also in the midnight uh (,)
$A Haan
$B So they asked them to ***go to the*** (,) morgery (,,) whereas (,,) in the morgery even the (,,) fresh bodies are (,,) I mean kept (,,) So (,,) sometimes they (,,) I mean they ask the students to ***go to the*** (,,) dissection hall and identify (,,) so many things (,,) At the same time (,,) some of their (,) same collegues (,) they must be hiding (,)

Example 6.63. ICE-IND: S1A-090.

close for Mister Patterson {2 [2 (,) 2] to call an election right now
$A [1 Yes 1] 1}
$A [2 Yes 2] 2}
$A In terms of if if { [if]
$B [He] } wouldn't have had If he really wanted a mandate from the people he wouldn't have had a better opportunity to ***go to the*** public now for a mandate Uhm I mean the economy seem to be (,) the dollar seem to be stabling se settling down the economy settling down the dollar is fairly stable and the J L P is in a vast state of uh chaos I couldn't have he couldn't have a better I was listening to hear that kind of t

Example 6.64. ICE-JAM: S1A-092

6.6.14 come up

In ICE-India *come up* has semantic preferences for **class and gender**, resulting in semantic force of **movement that produces advancement of class or equality**. This is exemplified by Example 6.65 where woman *come and (,,) be one with (,,) men.* In ICE-Jamaica *there* is a collocation of *come up,* as in *come up there* (Example 6.66) which seems to work against the more common semantic preference of *come* as a location that is associated with either the speaker or the listener. However, the most common use of *come up* across the ICE corpora is collocated with *with,* and semantic preferences of **trying**, and **better plans, schemes or ideas**. This gives a semantic force of **attempted movement that, and this is by no means guaranteed, creates a change of state for the better**. In Example 6.67, taken from a news broadcast, the City will attempt *to compensate for the library's loss,* and the parents are going to try and find ways of keeping the school open; and in Example 6.68, the leaders hope – it

is not definite – that they can come up with a formula that will peacefully settle the crisis.

iven equal chances with men (,) and uh almost in all (,) uh position even women are (,,) have been working (,) I mean only thing we have to utilise the (,,) chances properly and (,,) get into proper positions and make (,) woman ***come up*** somewhere Educate them basically and then (,) make them also (,) ***come up*** and (,,) be one with (,,) men (,)

Example 6.65. ICE-IND:S1A-011

$A And by the third year trust me they don't want to come back here for a vacation And you are sending a ticket to them I mean you are paying a ticket for them to come home But I've seen those changes What I find as to planning a lot of Jamaicans ***come up*** there and they select subjects in college that they can't get a job in when they graduate A lot of them ***come up*** there to do agriculture Try to get a job afterwards in agriculture

Example 6.66. ICE-JAM: S1A-005

City says it's not backing down Marie Bourdain admits the public was not consulted before the decision but she insists people will have a say when the City tries to ***come up*** with ways to compensate for the library's loss Mary Hailey C B C News Montreal
$A About forty parents met last night to try to ***come up*** with ways of keeping the Statacona Elementary school in Quebec City's Limoilou district open next year Catholic School Commission officials say it's too expensive to keep open for eighty-five students

Example 6.67. ICE-CAN: S2B-006

$A The visiting dignitary called on the President Mr R Venkatraman (,) and discussed a wide range of subjects (,) including the Gulf crisis (,) Both the leaders hoped (,) that the forth coming NAM meeting in Belgrade (,) would ***come up*** (,) with a formula (,) for a peaceful settlement of the crisis (,,)

Example 6.68. ICE-IND: S2B-007

6.7 Ante pre-set collocates

6.7.1 they come and they go

There would appear to be no additional semantic forces, semantic preferences, colligations or collocates that I have been able to identify relating to *they come.* While *they go* is discussed further in Section 6.7 in relation to *go* and reported speech and/or thought, it, in the main, also has no additional co-selection components. The *they* in *they come,* acts as a collocate of the right pre-set collocate come-*grams,* and, similarly, the *they* in *they go* is normally just one of the collocations for the right pre-set collocate *go*-grams. For example in Example 6.69 the semantic preference is that of *come back* – **previous/usual/routine location** and **specific point in time**, giving a semantic force of **movement to return to a previous/usual/routine state or place at**

a specific time. In Example 6.70, the post pre-set collocation is *go out*, with the corresponding semantic preference of **further significant action** and a semantic force of **movement from a location to undertake something significant .**

ant thing of all is to make sure that all prisoners of war and everyone who's been trans-shipped into Iraq They talk of eight thousand Kuwaiti armed service men and some twenty-two thousand Kuwaiti citizens Uh uh I wouldn't want to say or do anything that upset the possibilities of making sure ***they come back*** as quickly as possible And that having been said uhm Mrs Beeton always said in her recipes didn't she first catch your hare We haven't got him in the jug uh Saddam Hussein uh No one knows where he is or what is going to happen to him There

Example 6.69. ICE-GB: S1B-036

say is history They start having sex Now at first this sex takes place in hotels because he gives her the excuse that he's recently been divorced and he has two children and he doesn't want to traumatise them with introducing a new woman Now she accepts this excuse it seemed reasonable to her ***They go out*** on a couple of dates and he's very controlling so he decides that he's going to get her an apartment puts her up in a very ritzy apartment He comes over there whenever he wants sex Now she (,) this whole time she don't know that he's married She's very happy I mean who wouldn't be the man is

Example 6.70. ICE-JAM: S2A-065

6.7.2 to come and to go

The most prominent collocate common to both *to come* and *to go* is *want.* Additionally, the main semantic preference is words or groups of words indicating **(not) want** or **(not) desire**. There is a semantic force of (**no) desire for movement**. In Example 6.71, the speaker is asked if they would like to return in *fifteen or twenty minutes*; in Example 6.72 they want to produce a better *course,* and in Example 6.73he writer wants Dr Babar present. It should also be noted that the following pre-set collocates in all the examples work in tandem with these ante pre-set collocates. For example, in 6.72 the post pre-set collocate is *come up*, and with this there is the collocate *with*, the semantic preferences of **trying** and **better plan/scheme/idea**, giving a semantic force of **attempted movement that, and this is by no means guaranteed, creates a change of state for the better.**

wasn't ready So ((unclear-words)) a lot of anxiety and crossness and feeling let down was coming out So (,) I said oh no That's that's so annoying And so he then said well it will only be fifteen or twenty minutes **if you 'd like** ***to come*** back And then I showed him what was wrong with the other one So he then said to me (,) I 'll drop it round to you (,) And it was very strange because my reaction to that was oh no No That's all right

Example 6.71. ICE-GB:S1A-064

a ((a few words)) (,,) Ah Another thing behind another concept behind the new courses uh was (,,) and there's been a

great discussion as to what the translation of compétence is it is Is it competency or proficiency or (,) you know pick a word out of a hat sort of thing Uh **the aim was to** (,) ***to come*** up with an outline for a course where students would actually be able to do something when they finish the course uh as in be able to speak to someone be able to write something be able to read something as opposed to saying (,) the students will will be able to manipulate (,) six and a half ver

Example 6.72. ICE-CAN:S2A-027

But accommodation and local hospitality is assured. Do you think that Dr. Babar will **agree** ***to come*** on anything less than airfare. I **want** him ***to come***, but I can not think of airfares at this stage. Is there any other source that can support or subsidize his travel? Can you raise this question and get his views for me?

Example 6.73. ICE-IND:W1B-012

If there is a semantic preference of **measurement time** or **distance** preceding *to come* or *to go*, combined with a semantic preference of **attainment.** There is a semantic force of **time** or **distance left before something is attained.** In Example 6.74, it is the action of men *in the years to come* that will enable *scientific and technological development,* and in Example 6.75 the *entry-level machine* will provide *service and enjoyment.*

Because we are going to provide them (,) the men who are going to later (,) uh act and who who **in the years** ***to come*** are going to lead our scientific and technological development (,,) Now this aspect (,) was (,,) well thought out by our national leaders even before we became independent (,) The (,) science and technology policy gre owes a great deal (,) to the foresight of Pandit Jawaharlal Nehru (,,) Even b

Example 6.74. ICE-IND:S2A-023

buying for a younger rider, remember to think ahead. Consider what you might want in a snowmobile in the future, and think about whether you'll need a down-size or full-size chassis. With proper planning, the entry-level machine you buy now will provide years of service and enjoyment **in winters** ***to come***. SINGULARLY SATISFYING The best in single-seater luxury-touring machines.

Example 6.75. ICE-CAN: W2D-019

6.7.3 have to go

There is a semantic preference within the concordances lines of *have to go* of **necessity, obligation**, or of **problem** or **difficulty**: suggesting a semantic force of **movement that goes beyond the normal run of things that could be**

potentially difficult. Standard grammars often suggest that the use of *have* has necessity and obligation but not necessarily the idea of potential difficulty (eg: Leech 189: 177). In Example 6.76, there is a *whole legal nightmare* to be negotiated; in Example 6.77, *extra-ordinary lengths* have to be taken; and in Example 6.78, the *Toronto Maple Leafs* will have to *work hard* to be able to win their match.

It's just not biologically possible Uhm adoption is a very (,,) different thing in the north For one y if if it's between northern families they don't go through the whole legal (,) nightmare that everyone seems to ***have to go*** through down here They often will adopt uhm (,) like i if if a girl gets pregnant

Example 6.76. ICE-CAN: S2A-039

Uhm (,) they're commonly found throughout this part of Indonesia this time slice (,) uh in Indonesia Uhm they haven't been used before (,) and they're they're rather nice to look at as you'll see later (,) I hope (,,) OK (,,) We ***have to go*** to extraordinary lengths to actually find these fossils Most of the the work that I do is in the jungles of Indonesia (,) And we have to collect our rocks uhm along rivers because that's where (,) uhm rocks are exposed

Example 6.77. ICE-GB: S2A-046

This is nothing new for the Toronto Maple Leafs. They're into their third NHL playoff series, and they've lost Game 2 in all three, including a 3-2 setback Wednesday night to the Los Angeles Kings. " We're confident," says head coach Pat Burns. " We know we're going to ***have to go*** down there and work hard" " We have given them some life. We'll make some changes and find some ways and see what we can do."

Example 6.78: ICE-CAN:W2C-007

6.7.4 you come and you go

There would appear to be two notional semantic preferences that I can identify with *you come* and *you go.* There is a specific semantic preference of *you* of **the other participant or participants of the exchange**, and a semantic preference of *you* of **a random person or persons (non-specific) that could exist somewhere.** The first example, Example 6.79, is non-specific, and the next one, Examples 6.80, is specific. Example 6.81 is an example of a direct question to the participant.

And of course most of us uhm myself included use our hands to prepare these hamburger patties Uh if the bacteria is there and if you are not (,) really really uh conscious of washing your hands well with warm soapy water (,) after you have (,) handled the ground beef before ***you come*** into contact with your family members (,) or before you go on to prepare the salad as the next part of your meal (,) you really are risking spreading that bacteria within your kitchen
$A So this is uh a significant finding in the study that that you can get it also through personal contact

Example 6.79. ICE-CAN: S1B-050

they are very hot (,) I don't suppose you have so hot summers in Goa because
$A Yeah Goa is not that very hot You know summer I think that is the right time you can enjoy in Goa (,) We have lot of social activities (,) uh cultural activities that time you { [know] That is the right time I think if ***you come*** to Goa you will enjoy (,) { [uh uh (,) and not rainy season I don't will not advise any one to come ((laughs)) in rainy season

Example 6.80. ICE-IND: S1A-065

$C Was it the second time I've (,) ever (,) ever been there No It was the third time I've been to Lindos because we (,) we went to Lindos { [originally]
$A [I went] } (,,) Did ***you come*** with me
$C I went to Lindos once and slept on a beach there { [((unclear-words))]

Example 6.81. ICE-GB: S1A-063

In the first example, Example 6.99, the usage is non-specific and *go* is a replacement for a speech verb. The final example 6.101 is a non-specific question – in this case a rhetorical question.

ing
talkshows ((words)) ((laugh)) I don't know { [I can't ((word)) so much]
$B [Although they can] } be addicting
$A { [They can jeez]
$B [As much as ((word))] } When you watch a little bit of Donahue or ((laugh)) something and ***you go*** holy cow this is interesting I wanna hear about these people's lives slobber drool drool slobber { [((laugh))]

Example 6.82. ICE-CAN: S1A-021

$J The big problem is how do you get from the wide band to the narrow band Well that's relatively easy when once inflation converges but the bigger issue is where do ***you go*** from there in in phase two A and certainly our view is that the hard ecu proposals of the Chancellor are extremely helpful in that direction
$A But you see the Prime Minister said and presumably this is one of the areas of confusion which led to Sir Geoffrey's resignation the Prime Minister

Example 6.83. ICE-GB: S2B-007

6.8 Discourse managers

There are *come*- and *go*-grams that have a semantic force related to how speakers intend to organise the message, in other words discourse deixis. These superficially divide – they would appear to be an overlap of semantic forces – into those that are

right pre-set collocates, *come back*, *come in*, *come on*, *come to*, *come up* and *go back*, and those that are left pre-set collocates, *I go* and *we go*. I think it worth restating that this overview only examines those pre-set collocates of *come* and *go* of a frequency equal to or above 40/million that occur in the spoken colonies of all the ICE corpora. In the examples that follow *I come* and *we come* are also in evidence, but they do not meet the frequency requirements and thus are not examined.

Both the *come*- and *go*-grams have a discourse preference of **spoken language,** but the *come*- and *go*-grams differ in their semantic preferences. The *come*-grams have a semantic preference of **a specific location in the discourse and time (often in the near future)**, while the semantic preference of *go*-grams are much less specific in comparison. The *go*-grams are temporally less specific and more abstract in what they are pointing to in the discourse. The *come*-grams have a semantic force of **an indication that the discourse will move at a specific time to a specific subject/topic**, and the *go*-grams, **an indication that the discourse will move at some point in time to an area that includes a subject/topic**. The difference is subtle but, I think, underlines the idea that *come* has a more specifically directed semantic force than does *go*. There are, however, instances where *I go* and *go back* are more specific in what they are pointing at, temporally and topically, and these would appear to occur where there is an additional structural preference of **hesitation**. Here, perhaps the use of a *go*-gram rather than a *come*-gram is an instance of purposefully making the language seem more vague or less specific in order to soften the expression so that it does 'not appear too direct or unduly authoritative and assertive' (Carter 2004b: 32)?

The first set of examples are come-*grams* and the second, *go*-grams. In Example 6.84 the radio announcers will *come back* after the break and initially *talk a little more about that* and then they will *come back* to Michael's point. The next example, Example 6.85, insists that *John Mortimer* should be allowed to *come in there* – there is nothing vague and face saving about this language. The speaker states that they want to *come on to when you went off to Germany* in a short time – Example 6.86, while the speaker in Example 6.87, is insistent that they want to discuss the *main thrust* of the plan *now*.

Howard is gonna be joining us
$A You know you know Jimmy is is is Danny's { [father
$C [Big name} big name in j now you ((word)) big my big name selecting and him will anyone recall Danny's name
$A We'll we will talk a little bit more about that when we ***come back*** we will take a break right back Smile Jamaica continues it's morning time
$C ((chuckle)) Hands this thing Michael wants to make a point but Geoffrey has had to go to the phones so when we ***come back*** we'll take that So hello good morning

Example 6.84. ICE-JAM (59)

the arts world is is booming and bubbling but there is real problem with the London Boroughs ((unclear-words)) ((several-speak-at-once))
$A No let John John Mortimer { [***come in*** there]
$C [Uh I just want to] } to take ((unclear-words)) because I want to thank Mr Renton very much and and I certainly wouldn't cast him as the Prince of Darkness (,) And I would like to say that i it's the Labour council

Example 6.85. ICE-GB: S1B-022

when we came back we would be the owners of the property at Unit Eight Mill Road (,,)

$A I'll ***come on*** to when you went off to Germany (,) shortly (,) In your statement on page three the last but one paragraph you say ((unclear-words)) throughout the seventh of January Mr Sainsbury discussed the situation and particularly mentioning the urgency to arrange the funds as the completion date wa

Example 6.86. ICE-GB: S1B-061

and perhaps (,) later on we can discuss (,) in what ways that may be achieved (,,)
$A Thank you very much Now (,) I would like to (,) ***come to*** (,) the area of the main thrust of the (,) eighth five year plan (,)

Example 6.87. ICE-IND: S1B-026

In Example 6.88, the speaker is hesitant in wanting to *go back a bit* and in Example 6.89 the speaker asks for an unspecific amount of time to allow him to *go over things*; and in Example 6.90, it is not clear how long the speaker is going to take in explanation before he continues with the main message. The speaker in Example 6.92 indicates that they are continuing with the discourse with the statement *I go on* for an unspecified amount of time. However, in the Example 6.93 the speaker is specific in what they want to return to but quite hesitant in the approach.

rally agreed that by the nine the mid nineteen seventies that this approach (,) through the use of Systran and many other systems was inherently uhm (,) incapable of development of great development (,) Therefore by this time it's generally agreed (,) that Oh no sorry Let me ***go back*** a bit The foc The syst The The (,) approach which was adopted immediately after the direct system was known as the interlingua approach where you had one type of representation interlin uh an interlingua which was intermediary between (,) a source language and a target language and you had t

Example 6.88. ICE-GB: S2A-032

And ultimately the judge will give you some instructions (,,) What I'd ask at this time (,,) is that you (,,) just give me maybe ten or fifteen minutes (,,) of your attention while ***I go*** over things that I feel are important (,) on behalf of Herbert Wasylesko (,,) I want to reinforce just a couple of points that Mister Phelan had made when he was standing chatting with you The first one is that (,) Herbert Wasylesko is an innocent man (,) unless and until you say otherwise Poi

Example 6.89. ICE-CAN: S2A-063

ause that slide wasn't prepared specifically fo for for this talk but for another one (,) Uh and it it's certainly uh I think a fair assumption that uh when talking about innovation most people assume that innovation is something that's carried out within the industrial context But let me say before ***I go*** any further that that isn't a fair assumption and that what uh I will try and illustrate what I mean by innovation and how it 's different from and subsumes research But innovation to a large extent can be considered to be an attitude of mind and in some things it is equally applicable in uh a

Example 6.90. ICE-GB: S2A-037

red every night for the next five nights She went to work the next morning as usual (,) carried on her work responsibilities as usual ((clears-throat)) because there was nothing unusual happening other than she had ((word)) lost her boyfriend (,,) So Ladies and Gentlemen (,) we Ladies (,) ***I go*** on she was so distraught after she supposedly was raped that not only did she accept the roses and the gift card Mr Jack sent her the following day but she took them home and had them in the apartment Why not throw them away (,)

Example 6.91. ICE-JAM: S2A-066

$A [8 Yes he's 8] 8}
$A [9 Yes yes yes 9] 9}
$A [10 Yes 10] 10}
$A [11 Yes he is the ((word)) 11] 11}
$A [12 Ex exactly exactly exactly 12] 12}
$A Uhm the I think you mentioned something that I'd like to ***go back*** to the background {1 [1 (,) 1] of the individuals the how they were brought up {2 [2 (,) 2] Uhm and it is so that if children grow up seeing their parents being beaten or especially seeing their fathers beating their mothers {3 [3 (,) 3] many of these men themselves (,) {4 [4 turn 4] out to

Example 6.92. ICE-JAM: S1B-032

6.9 Sports reporting

On examination of the instances of the *come*- and *go*-grams in sports reporting it would appear that the deictic centre – located by the semantic force – has, in relation to the commentator, shifted and it is the winning position related to the particular player/horse/athlete being currently described. The deictic centre has no relation to where the commentator is situated whether it is back at the studio or adjacent to the pitch, or somewhere entirely different. So, basically, the discourse preference is **live radio sports reporting,** the semantic preferences are **sporting terms**, **and the winning position.** The semantic force for *come*-grams is **movement towards the winning position**, and that of the *go*-grams, **movement away from the winning position.**

In the following I have included three football commentaries; two from India (Examples 6.93 and 6.94) and one from Great Britain (Example 6.95): two cricket, India (Example 6.96) and Great Britain (Example 6.97): and two from the same British horse race, (Example 6.98).

In the three examples of football commentary the deictic centre seems to be the goal scoring position in relation to the particular player described at that point. In Example 6.93, in coming up to the ball the player places himself in the position to kick it forward towards the goal. Example 6.94 describes a substitution: the substitute comes on to the pitch, thus making himself available to assist in winning the game. In the final example, Example 6.95, the player comes up to the ball in order to kick the ball at goal for the penalty.

he big centre (,,) Alok Das (,,) Amitabh Chandra (,,) Back to Das (,,) Long ball played (,,) And (,) Akeel Ansari in no mood to oblige Vijayan is on his own way (,,) Mohun Bagan forward (,,) constantly on the run (,,) Athanu Bhattacharya sensing a bit of danger in Vijayan (,,) did the right thing to ***come up*** to the ball (,,) Foul was committed by Vijayan (,,) Here comes the

(,) a free kick (,,) Tushar Rakshit (,,) dispossessed (,,) Biswas backing up (,,) It's uh stretches (,,)

Example 6.93. ICE-IND:S2A-008

And the substitution (,,) being made (,,) by Bidhan Nagar (,,) Vimal (,,) no it is Raja Chatterjee (,) who's coming on (,) Raja Chatterjee (,,) the number thirteen who'll ***come on*** (,,) And coming out is the number twelve (,,) that is Santosh Gupta (,,) So Raja Chatterjee will ***come in*** (,) in the next back position (,) Bidhan Nagar (,) in attack (,,) Once again (,,) to Sanjoy Das (,) the skipper (,,) This is in the second half Bidhan Nagar leading their opponent Sports

Example 6.94. ICE-IND:S2A-003.txt

And a free kick to Manchester United which 'll be taken by their captain Bryan Robson A left-footed one into the middle Webb is up there It goes over his head (,) takes a knock off Kenny Sanson goes behind for a corner on the near side (,) And Denis Irwin the full back has ***come up*** to take it (,) Places it (,) as the (,) home fans (,,) cheer on faithfully looking for that opening goal as Irwin floats the corner in Pallister goes up just misses Nice ball for Sharp there (,) Wanted little too much time and his shot eventually came off Barker

Example 6.95. ICE-GB:S2A-003

In the following two examples of cricket commentary, the deictic centre would appear to be related to the stumps. The two batsmen in Example 6.96, are safe when they have returned – come back – to their respective crease; and, in Example 6.97, Jose comes on to the pitch towards the stumps.

ts it down through square-leg They've taken one They're *coming* back for the second He's t haring in The ba throw is just wide of the stumps from Suli Mallik And so they ***come back*** safely for the second run

Example 6.96. ICE-GB:S2A-013

Jose goes back for (,) total of (,) individual total of seventeen (,) this (,,) this session always (,) necessary (,) to (,) know what type of bowling is there (,) and she has ***come on*** last evening say (,) she is a (,) off spinner (,) She is very success successful bowler in the nineteen ninety-three World Cup (,,) where (,,) she had a (,,) very good throw with wonderful fielding (,) She is a very good fielder also (,) And she is coming up from the (,) Darya Ganj (,) end

Example 6.97. ICE-IND:S2A-016

And, with the racing commentary below, regardless of what direction the horses are running in relation to the reporter, the deictic centre would appear to be always the winning post.

mber ten is Golden Frieze still with the whip hand from Mr Frisk Over The

Road Up on the outside General Chandos Garrison Savannah just in behind the leaders And so they jump that one and there 's no casualties The tail-end are still coming through But they ***come up*** towards this open ditch Golden Frieze still the leader from General Chandos Over The Road Then comes Mr Frisk Racing with him on the near-side is Rinus And so that's the leading quartet Closely followed then by Garrison Savannah

Example 6.98. ICE-GB:S2A-005

It is also worth noting that it is the *come*-grams that predominately point at the deictic centre, as the *go*-grams indicate a movement to where the centre is not. However, when the *come*-grams and *go*-grams are examined in contrast to each other there is some evidence that *go back* and *go through* are also used in this way. In Example 6.99, Thompson has to *go back* to play the shot which suggests that he has to move away from the opposition's goal (in which he wishes to score) in order to do it. In Example 6.100, *Amir* lets the ball *go through* rather than taking a possible run making shot; and, in Example 6.101, the ball is again let through to the wicket keeper. There also would appear to be a particular use of *go into* in reporting of horse racing. In the first two out of the following three examples (Examples 6.102, 6.103 and 6.104) that utilise *go into*, it takes place in the first part of the race, but it is not possible to ascertain to which bit of the race the commentator is referring to in the final example. Perhaps there is an indication that the horses are moving away from the winning post at this point as in the first two examples? There would appear to be some evidence for this in Example 6.104, as when they enter the *home straight* the reporter describes them as *coming back.*

Calgary as the Dinosaurs fire it out front Back to the point it comes for Pegg Flips it towards the net That (,,) redirects just wide Thompson'll ***go back*** to play it (,) Chips it back deeper in his own zone and Krywko clears (,) it out (,) Schoneck back to pick it up for Calgary (,) Nine and a half minutes to go in the second here (,) of game two of this best-of-three series Last night Calgary won by a score of five three (,,) Bears have it i

Example 6.99. ICE-CAN: S2A-004

five (,,) I beg your pardon Jahur Ilahi (,,) has three to his credit (,,) A bit of lift on that ball (,,) and Amir (,,) seeing it well in time let it ***go through*** (,,) And Kapil Dev too sort of angling through his first over (,,) Two slip (,,) another young Hyderabadi (,,) doing some duty for the fielding side Noel David (,,) is now brought in (,,) at third place (,,)

Example 6.100. ICE-IND: S2A-001

be focus as much attention to much closer together now uh rather drop it at first and vivious ((((word)) at second because the ((two words)) that went through between them There's Srinath He's bowling now to Hooper who's letting it ***go through*** outside the off-stump just gentle little loosing up Only a bit tough though for Sourav Ganguly to to resist throwing the ball at Javagal Srinath every time

Example 6.101. ICE-JAM: S2A-006

And the tapes will be sent up by the referee in the box just above the start They're roaring up at the start now and they ***go into*** the first bend Doncaster got a bad start there and it 's Leram who moves into the lead although he's tussling all the way with Mullet And Mullet and Leram are really having a neck and neck battle And it's Leram who takes it now in the bend coming back into the home straight

Example 6.102. ICE-GB: S2A-012

$A They're off (,,) Jam Rock comes out leaping like a bucking bronco and left at the back early (,) Mr Butcher takes off quickly (,) So do Summer Princess Rule By Secrecy moving fluently against the rail as they ***go into*** that (,) first turn and Rule By Secrecy has stepped off it but holds the lead Mr Butcher joining Summer Princess right there on the premises Sound Of Speed kept just off those fractions joined up by Helta Skelta Lady Suhaagraat is next and the bucking bronco

Example 6.103: ICE-JAM: S2A-019.

in fourth just ahead of Port Royalty and Special K some six and a half lengths off the lead Heart Of Steel now making progress on the fence With Honour at the back of the field (,) The leaders about to arrive at the half mile They ***go into*** that turn They've still got to get back to Rush To Victory Here's Sir Kunjabihari with his search on the outside taking him to within a half a length as they lead the seven sixteenth and he now joins Rush To Victory Prime Minister right on their heels as they go passing the three A

Example 6.104. ICE-JAM: S2A-014

6.10 Replacement speech/thought verbs

In the ICE-Canada, when *go* is colligated with a pronoun – *I, you, we, they,* and a discourse marker such as *oh* or *hey,* or collocated with *and* and again colligated with a discourse marker and then followed by reported speech or thought (structural preference), the semantic force of the *go*-gram is **this is what was said or thought.** I should mention that there are a few of examples of this usage in ICE-GB, one in ICE-India, and none in ICE-Jamaica. This phenomenon predominately has a discourse preference of spoken language and ICE-Canada and the following, Examples 6.105-108, are illustrations of this. Example 6.107 is taken from ICE-GB. I would suggest that the *moan* in Example 6.106 is a speech activity as it communicates meaning.

$B Oh ya
$A Oh she didn't say anything The worst effect of her that her phone calls have is that I sit there and say oh Barry she's such a jerk
$B Ya ((laugh)) And of course I call ***and go*** **oh god** I called you again I was calling my dad And Barry goes oh well (,) I'm on the other line with my mom { [And I went (,,) what for]
$B [((laugh))] }

Example 6.105. ICE-CAN: S1A-069.

You know if they don't have a pencil then that means they're pushing it You know maybe You know maybe you should bring pen I dunno It depends on your situation you know But I didn't have to bring pencils at least but I did have to bring paper (,,) you know So ***they go*** ((imitates student moan)) Go **Ah ha you know**
$A Take out the paper and what do you do for the first class (,,) What would you do as a first (,) class (,) activity

Example 6.106. ICE-CAN: S2A-025

Uhm (,) yeah they're really good Really nice (,,) I was really waiting for the moment for people to have faxes at home (,,) because you can write really good messages (,) and fax them whereas when you call somebody up you have to go hi how are you and **they go** I'm fine how are you what have you been doing you know whereas on a fax you can just send a really good message and a picture pictures of other people and pictures of you (,,)
$A You send pictures of you
$B No I don't send pictures of me Other people send pictures

Example 6.107. ICE-GB: S1A-015

Actually I find (,,) do you watch Rosanne
$B Ah used to { [not (,) like anymore]
$A [Uh Dan her husband] } I find him ((laugh)) really funny (,) { [((word))] Rosanne
$B [Ya] }
$A I don't know c he's a typical male and **you go** right on man ((laugh)) you know yes
$B Pass me a beer
$A ((laugh)) Well you know (,) it's just kinda funny the these taken to extremes these situations When uh one of the things when I uh can't remember what it is the daughter with the dark hair who's (,) tough (,) mean (,) Charlene or {1 [1 s

Example 6.108. ICE-CAN: S1A-02

In summary

I summarise the main points and conclusions reached in this chapter below. The co-selection components of the lexical items are recapitulated in table form in Appendix X, as are the co-selection components identified in the following chapter.

- As a result of my research two additional co-selection components are included. Structural preference and discourse preference. The former relates to traditional grammatical structures and less traditional structures. The latter relates to specific discourses in which a lexical item is most likely to be found. By introducing structural preference, colligation is restricted to the identification of word class.
- Again, as a result of my research, semantic prosody has been replaced by the term semantic force.
- The *come*- and *go*-grams are divided into those that have pre-set collocations before and those that have pre-set collocations after *come* and *go*. The former are called ante and the latter post.
- It was noted that post pre-set collocates and ante pre-set collocates semantic forces would appear to act in tandem – the sum is equal to the parts.

- Ante co-selection components are primarily associated with ante pre-set collocates, and post co-selection components are primarily associated with post pre-set collocates.
- Any co-selection component identified in the written colonies will also be present in the spoken colonies.
- Familiar idioms are identified in the corpora, but none at high frequency so they are not explored further.
- *Come*- and *go*-grams are utilised to manage the discourse.
- There is evidence of deictic shift in live radio sports reporting. It would appear that the deictic centre is allied to the winning position.
- *Go* is utilised as a speech or thought verb predominately in the spoken colonies of ICE-Canada.

In the chapter that follows I continue my examination of the semantic forces associated with ante pre-set collocates and post pre-set collocates in order to investigate the tandem effect of the semantic forces identified in this chapter. I show that the identification of the co-selection components depends on the choice of the node. I also go on to suggest that there would appear to be layers of semantic forces that function on the same stretch of text simultaneously.

Chapter 7 Narrowing the focus

Meaning is always provisional
Teubert 2010: 21

I have shown that the *come*- and *go*-grams can have individual co-selection components, and that they can also have the deictic co-selection components that are commonly associated with *come* and *go.* I have also shown that it is possible that the semantic forces of particular *come*- and *go*-grams can be added together to work in tandem rather than producing a new semantic force. In other words, the ante and post pre-set collocations operate as collocates of each other This, I suggested, shows that the chosen node of the concordance would appear to have some impact on the resulting co-selection components of the proposed lexical item.

In the first section that follows I explore in detail the *go*-gram *to go to*, examining the concordance lines from all the private colonies of the ICE corpora. I illustrate how the semantic forces of *to go* and *go to* work in tandem with each other to produce a combined semantic force that should not be considered a new semantic force. I have chosen these private colonies as they give sufficient concordance lines, and, as I have observed before, it is within the spoken colonies of the ICE corpora that there is the most variety of co-selection components (see Section 6.1). The concordance lines under investigation are reproduced in Appendix VI. The concordance is subdivided into the different ICE corpora. In the appendix the concordance lines are 140 characters (including spaces), but where I reproduce them below I have shortened them to fit the available space. The numbers at the beginning of the lines correspond to the numbers in the appendix.

In Sections 7.2 and 7.3, I explore in detail *come and*, *and go* and *come and go* in order to demonstrate that the selection of the node can have considerable impact on the resulting co-selection components. In Section 7.2 I examine all the concordance lines from all the ICE corpora generated with the nodes *come and* and *and go.* The concordance lines from which I have taken the examples are reproduced in Appendices VII and VIII. The concordance lines reproduced in these appendices consist only of the colonies from which I have taken the examples. Again, the concordance lines are 140 characters including spaces, and where reproduced below these have been shortened with the numbers corresponding to those in the appendices. And, in the final section, Section 7.3, I examine all the concordance lines of *come and go* in all the ICE corpora. These are reproduced in the text.

7.1 to go to

In Chapter 6, Section 6.4.1, I demonstrated that *to go* is collocated with **want** or has a semantic preference of **(not) want or desire,** and a semantic force of **(no) desire for movement**, and that *have to go* has semantic preferences of **necessity/obligation** and **problem/difficulty**, with a semantic force of **obligated movement that has the potential to be difficult.** I also demonstrated in Section 6.3.13 that *go to* has a strong structural preference of **post noun phrase** and has a semantic preference of **a location non-specific to the discourse participants**.

Table 7.1 provides the frequencies of the three most frequent ante co-selection components associated with *to go to* in the private colonies of all the ICE corpora, and

Table 7.2, the post co-selection components, again in the private colonies. *Going* and *have/had/has* and *want/wanted/wanting* are all ante collocations of *to go to*, and *the* is a post collocate (as part of a noun phrase), although *going* might be considered to be colligation as it has a function of indicating that an event will be happening in the future. The post **verb** is a colligation, and the post noun phrase is structural preference. The post noun phrases are non-specific to the participants of the exchange. The final column in the first table shows the semantic preference **(not) want.** In the second table I have included the frequency of the remaining post qualifications in order to emphasise the high instances of a post noun phrase.

	to go to	going	have/had /has	want /wanted /wanting	**(not) want**
ICE-Canada	42	5	15	5	3
ICE-GB	36	6	5	9	5
ICE-India	47	0	20	4	9
ICE-Jamaica	69	0	18	13	6

Table 7.1: Frequency of the ante co-selection components associated with *to go to* in Private ICE-Canada, -GB, -India and –Jamaica.

	to go to	*the*	verb	noun phrase	other
ICE-Canada	42	11	1	28	1
ICE-GB	36	3	0	28	5
ICE-India	47	10	2	32	3
ICE-Jamaica	69	7	1	40	5

Table 7.2. Frequency of the post co-selection components associated with *to go to* in Private ICE-Canada, -GB, -India and –Jamaica.

In the following concordance lines I illustrate the different ante co-selection components as discussed above. Concordance 7.1 shows *going to go to*, taken from private ICE-GB, Concordance 7.2 both *have/had to go to* and *want/wanted to go to* from private ICE-Jamaica, and Concordance 7.3, semantic preference of **(not) want/desire**. I include *ask* as the person who is asking, I would presume, wants the event to happen, and *forced* is included as the person who is forced to do something patently does not want to do it. Likewise *couldn't get a chance* implies an unfulfilled desire to do the action.

GB

1 basically but I'm not (,) $B Are you going to go to all uhm the day on of the phonology
13 rt Hall { [one place and uh he was going to go to I don't think he's at Edinburgh

Concordance 7-1: *Going to go to* in Private ICE-GB

Jamaica

13 o the same rules (,) like (,) uhm you had to go to classes you had to (,,) after
15 hat's cool 2] 2} eh That's cool I I I want to go to { [Sweden I want to travel the
18 r the bio and the and the English you have to go to n a nursing school $B Yeah then
19 $A Uhm even though it's my goal uhm I want to go to Emerson uhm I realise that the
20 joy myself {3 [3 (,) 3] Right And I wanted to go to the movies too probably next week
21 r just like straight technicians they have to go to field at some point in time Before
23 No I think I staying for the summer I want to go to summer school $A Okay That's why
25 en after finishing I decided I didn't want to go to that area and now I'm doing my
26 tudents get in free {2 [2 (,) 2] So I want to go to that but I'm not sure $A [1 Okay
28 e $B [Mhm] } $B Hold on You mean you had to go to your bed at nine o'clock $A Right

Concordance 7-2: *Have/had to go to* and *want to go to* in Private ICE-Jamaica

Canada

23 just (,) maybe you know I I would prefer to go to the museum on Sunday Like so
25 niture and everything I have (,) and try to go to RADA or something in { [England]

India

1 omething I'm missing and I we're planning to go to Tawan (,,) when it becomes less (,)
14 ight uh (,) $A Haan $B So they asked them to go to the (,) morgery (,,) whereas (,,)
15 aren't you (,,) So I would better ask you to go to the (,) Mahalaxmi temple (,) have
16 one college is that one (,) guy was asked to go to the ladies hostel (,) $A Haan (,)
25 es they (,,) I mean they ask the students to go to the (,,) dissection hall and
32 don't know I can't say whe that I like to go to Bombay (,) but (,) I wanted to be
41 perienced it that I couldn't get a chance to go to temple (,,) and uh I just pray in
43 m his home town Lucknow and he was forced to go to Calcutta And there (,) amidst lots

Concordance 7-3: Semantic preference of (not) want/desire for *to go to* taken from Private ICE-Canada and Private ICE-India

And in those that follow I illustrate the post co-selection components apart from *the* as a sufficient number of these are included in Concordance 7.3 for illustration purposes. Concordance 7.4 shows all the instances of post verb in all the private colonies of all the ICE corpora and Concordance 7.5, a selection of post noun phrases. I would suggest that *like* (lines 31 and 33) is acting as a determiner in these instances. The final collection of concordance lines are examples of those that are none-of-the-above, taken from private ICE-GB.

Canada - Private

18 my parents were bit better off you know Go to go to eat in the restaurant there at the Sheraton Hotel there's a the

GB - Private

India - Private

5 l now there is no intimation $B Do you like to go to see (,) ((two words)) in Bombay
38 ay rickshaw has not come no so (,,) I have to go to (,) collect $B I also go to

Jamaica - Private

30 to tell her back {3 [3 stories for 3] her to go to sleep $B [1 Mhm 1] 1} $B [2 Mhm

Concordance 7-4: All instances of post verb in Private ICE-Canada, -GB (0), -India and –Jamaica.

Canada - Private

28 me to meet you at the hospital cos I have to go to work right now but I can meet you
29 or something $B Ya maybe I definitely want to go to Alison 's $A I'm having that
30 's party { [((laugh))] $A [You wanted to go to] } { [Janette 's] $B [We're]
31 e said no let's do it And we really wanted to go to like San Fransisco and stuff $B
33 ws] } I might have to go and I might have to go to like some teachers' college out in
35 lly different mindspace And we had planned to go to California the year before and I
36 t the two corroborate cos then she'll have to go to a third { [if the second guy ((a

40 guess and (,)] } he said he wasn't going to go to the wedding or was thinking not

Concordance 7-5: Post noun phrases in Private ICE-Canada.

GB - Private

12 ut I mean (,,) And I've got so many events to go to I mean ((laughs)) I know that
13 ert Hall { [one place and uh he was going to go to I don't think he's at Edinburgh
14 if you start giving her a series of events to go to it'll make her shy away $B I see
21 Oliver Oliver Pemb Pemberton $B [1 I used to go to s 1] 1} $B [2 I went to school
35 ly got four days there $C Yeah I'm tempted to go to ((unclear-words)) $A Yeah wish

Concordance 7-6: 'Other' in ICE-GB

When one then examines *to go to*, taking into account the co-selection components identified for *to go* and *go to* it is possible to see that they appear to be working in tandem. The reason that there are two sets is that those associated with *to go* are present even if the post collocation *to* is absent (see Section 6.4.1 and summary at the beginning of this section) and those associated with *go to* are also present even if the ante collocation *to* is absent (see Section 6.3.13 and summary at the beginning of this section). The examples below confirm this.

The first two examples have *have* as an ante collocation and a post structural preference of a noun phrase, the second also includes a post collocation of *the.* The semantic force for both is **obligation with some problem/difficulty to move towards a location that is non-specific to the participants –** non-specific in the sense that it is a location that is not specific to the location of the participants at the time of utterance. This semantic force corresponds to the combined semantic force of *have to go* and *go to (the).* In Example 7.1 the *host-mother* is unable meet the girl at the hospital – the problem – because she is obligated to go to work. She is not at the time of the conversation at work. In Example 7.2, in order to *see the exam* the student must visit *the professor* – it is not as simple as just handing *back the exams sort of thing so you can see what you did wrong.* The location of *the professor* is non-specific to the participants in the exchange.

Mhh hmm
e kids came here (,) if they had the same problem
$A I'm sure they did Well I know they did One of uh one of my participants had a had a health problem This is the one with the single mother host family And the single mom said okay well (,) you know want me to meet you at the hospital cos I have ***to go to*** work right now but I can meet you at the hospital And the girl said no no no And so the (,) host-mother took it at face value The girl (,) said she could take care of herself Fine (,,) What this girl was expecting was her host-mother to insist upon coming {1 [1 and and 1] spending twenty-four

Example 7.1. ICE-CAN:S1A-057

$B Huh I wonder if the they'll be available before that (,,)
$A I don't know You mean { [to see these]
$B [I wonder if you] } like if they're going to hand back the exams sort of thing so you can see what you did wrong
$A I think if you want to see the exam you have ***to go to*** the professor
$B Oh
$A But you don't get them back { [at all]
$B [Oh (,)] } ya That's usually the way with final exams anyway

$A Ya
$B Ya Huh Oh well
$A Mhh hmm (,,)
$B Oh am I ever tired

Example 7.2. ICE-CAN:S1A-040

The following three examples have a semantic preference of **want/desire.** In the first of the three, Example 7.3, the semantic preference is in the form of the collocation *want*, in Example 7.4 the semantic preference is represented by *everybody dream*, and in the final Example 7.5, the semantic preference is again a collocation – *wanted.* In all three examples the following noun phrase is a location that is non-specific to the participants of the exchange – *Japan, UWI* or *the movies*. The semantic force of the three examples is **want/desire to move to a location that is non-specific to the participants of the utterance.** This semantic force is equivalent to the combined semantic forces identified of *to go* and *go to.*

$B [2 Uhm 2] 2}
$B Well no It was a vague { [sort of]
$A [Oh] } right Well it's now definite (,) {1 [1 uhm 1] and I I definitely want ***to go to*** Japan uhm until I {2 [2 uh unclear-words ((two or three words)) 2]
$B [1 Oh yeah 1] 1}
$B [2 But no no 2] 2} The desire is definite or you've actually got a job
$A Oh the just the desire {1 [1 is 1] definite

Example 7.3. ICE-GB:S1A-097

pared for this new exam We knew that it wa these A levels determined if we could go on to u go to university and stuff like that I think that was very challenging I know for me studying for my A levels and getting prepared and getting the right grade and you know cos everybody dream is ***to go to*** UWI or some university abroad so I mean getting prepared was a really difficult time a really challenging time I remember our principal had this motto challenging but not impossible and I mean that motto brought me throughout sixth form in {6 [6 everything 6]
$A [1 Mhm 1] 1}
$A [2 Mhm 2] 2}
$A

Example 7.4. ICE-JA:S1A-087#51

ah for the whole weekend But of course I'm not wearing pyjamas ((laughter))
$A Why ((laughs))
$B Uh I don't really go according to the theme stuff {1 [1 (,) 1] I will just stay normal {2 [2 (,) 2] enjoy myself {3 [3 (,) 3] Right And I wanted ***to go to*** the movies too probably next week
$A [1 Okay 1] 1}
$A [2 Mhm 2] 2}
$A [3 Okay 3] 3}
$A Oh what movie you're going to watch
$B I think it's Shrek

Example 7.5. ICE-JA:S1A-067

It is evident that semantic forces combine not to create new forces but to work in tandem with each other. This would suggest that the choice of node has a considerable impact on the identification of the co-selection components of the lexical item. In the sections that follow I examine the *come*- and *go*-grams *come and*, *and go* and *come and go* in order to examine to what extent the choice of node might influence the identification of the lexical item.

7.2 come and and *and go*

Looking initially at *come and*: the first row in Table 7.3 gives the frequencies of *come and* in all the ICE corpora. The subsequent rows give the frequencies of the relevant ante and post co-selection components of *come and* in all the ICE corpora and all the colonies of the ICE corpora. The ante co-selection components chosen are those that have been identified in order to show that, again (see Section 6.7.2), the semantic force of the left and right *come*-grams add to each other. These are the collocation *to*, and the semantic preference of **want/desire** associated with *to* as discussed in Section 6.4.

		ICE-Can	ICE-GB	ICE-Ind	ICE-Jam
	come and	***43***	***43***	***65***	***94***
ante	*to*	13	16	12	27
	semantic preference **want**	9	9	6	10
post	colligation **verb**	26	30	46	46
	go	6	4	0	2
	semantic preference **social interaction**	12	19	31	29
	structural preference **clause**	8	6	8	16

Table 7.3. Frequencies of ante and post co-selection components of *come and*.

The post co-selection components include the colligation verbs normally directly after *and*, but I have also included those that have the same pronoun before *come* and before the verb as shown in Collocation 7.7 as, where there is this repetition of the pronoun the co-selection components are the same as where there is no repetition. The other post co-selection components are the collocation *go*, the semantic preference of the collocation verbs of **social interaction** (meeting, communicating, etc), and the structural preference of a **clause**. As the table shows, a following verb is much more common than a separate clause. When the *and* is used to join two clauses that are dissimilar, there seems to be no particular relationship between the two words – they just happen to occur because *come* ends one clause that is attached to the next by means of the conjunction *and*. However, where the *and* is followed by a verb there would appear to be a distinct semantic force as discussed below. The frequencies of *go* are included in order to compare those occurring in the concordances of *come and* with those of *come* that occur in the *and go* concordances in the section that follows.

Jamaica Private

19 er Anyway me no know a few months later she come and she say How come you didn't tell
20 Kedisha is the quietest one in my room She come and she o knock on the door and she
23 in the kitchen just right beside me so she come and she realized what was happening

Jamaica Unscripted

57 on if they'll R B C books this is where you come and you borrow the books to take out

Concordance 7-7. The same pronouns occurring before *come* and before post verb.

The concordances that follow give examples of the various ante and post co-selection components. Concordance 7.8 shows all the instances of *to* in ICE-Canada with the ante semantic preference of **want** associated with *to* in bold type. (Where the semantic preference is not included in the concordance line I have added it at the end in bold italic type and in brackets – e.g. line no. 19 ***(invited)***: for the full concordance line see Appendix VII.) Concordance 7.9 shows the post colligate verbs, and where there is a semantic preference of **social interation** it is shown in bold type. The final Concordance 7.10 gives examples where the *and* is used to join two separate clauses.

Canada Private

7 the stu like one of the student teachers to come and do the dialogue {3 [3 with me for
8 [2 Sure 2] 2} sure $B See if they **wanted** to come and (,) have supper with us and Or
9 $B Ya $A I think I'll just **ask** uh Morris to come and get it (,,) remove remove it { [

Canada Public

18 was uhm (,,) a child who was **longing to** to come and play in the playground $B { [Aw]
19 ited (,) uhm everyone on the street (,) to come (,) and uh meet with us And it was a
(invited)
20 ers with the community and **invited** them to come and visit and we had an open house and
21 I guess the the Quebec elections police to come and see us Uh from there they
22 (,) uh the parents that were there (,) to come (,) and participate (,) in {3 [3 uh
(invitation)
23 h of June All E P P parents are **invited** to come and the reason for that was school-

Canada Unscripted

32 en (,) Browning (,) Eldredge (,) Bowman to come and right in the middle of it is Elvis
33 hard to find people who are are **willing** to come and work in the school and do that (,)

Canada Scripted

[none]

Canada Non-Printed

41 ons of words. You are certainly **welcome** to come and use the resources of the Language

Canada Printed

43 ll go; phoned my husband in Cape Breton to come and take my child, and drove me to the

Concordance 7-8. Ante collocation *to* and semantic preference want in ICE-Canada

India Unscripted

1 have migrated from our (,,) districts and come and settled down here Now we've become
2 the ground then easily (,) the confidence come and they can practise shirsana (,) And
3 le (,,) or (,,) the society (,,) which has come and settle down here (,) become a part
4 his coach (,,) as the British teacher have come (,) and (,,) she feels thankful to the
5 ,) So he all (,) the customer has to do is come and pick up the application (,) fill
6) some people (,,) so (,) some people just come and **ask** me to paint things (,,) on any
7 bles have to prove (,) the watchman should come and **make a statement** before thepolice 8
y the school has given them the day off to come and **watch the match** (,,) There the
9) Our Gudi was there (,,) I forced Gudi to come and join this institution He was not
10 are there (,) or you are there (,) now to come and **tell** them (,,) look I don't mind
11 ffs of course (,) this is the time to (,,) come and **meet** (,,) some of their favourite
12 ment (,,) or if he is not (,) then he will come and **argue** himself (,,) and at themost
13) but we have a soft corner for you if you come and sacrifice (,,) and ultimately I
14 view her father in fact (,,) Why don't you come and **share your views** with us (,,) ?

Concordance 7-9. Post verbs and semantic preference social interaction in ICE-India unscripted

GB Public

3 lage hall he simply said uh uh let Kingdom come and my name perish {2 [2 ((laughs))
4 let's say Baghdad where the main supplies come and to the south where the Republican
5 e grown Londoners and of the tourists that come and they are being denied access to

Concordance 7-10. Post structural preference clause in ICE-GB public

The concordance lines show that, as with *to go to*, there is an ante semantic force of **want/desire** with *to come* which, I would suggest, is then added to the post semantic force of *come and*. There is also a frequent post semantic preference of **social interaction** when *come and* prospects a verb. However, *come and* like *go and* can often be considered to be surplus to the utterance (see Section 6.3.1.). In other words, the utterance can be easily understood without the inclusion of *come and* so there is also a structural preference of **surplus.** I would suggest that there seems to be some sort of temporal and physical shift from the moment of utterance to the moment action. Thus when there is a verb following *come and* there is a semantic force of **movement to undertake an action that will happen after the utterance with emphasis on the participant(s) being physically present in a place specific to the participants and the action often involves social interaction.**

In Example 7.6, the invitation occurs before the date that the *community* make a social visit to the speaker's *open house,* and the visit, in Example 7.7, is for the *general scrape* and is subsequent to the hygienist appointment. In both cases the *come and* could be deleted from the transcript with, ostensibly, little change in overall meaning. In Example 7.8, the students are punished and then the parents intervene, in Example 7.9, the killing occurs a year later. Again, in both these examples the *come and* could be considered additional to requirements and it serves to emphasise the physicality of the subsequent action.

$A Ah in in reality I mean we've really blended into {1 [1 the into the community 1] quite nicely and uhm (,,) and we've we've had some interesting uh (,) uh get-togethers with the community and invited them to ***come and*** visit and we had an open house and they could come in and (,) tour the {2 [2 facilities and what not 2]
$B [1 I would think so 1] 1}
$B [2 Mm hmm 2] 2}

Example 7.6: ICE-CAN: S1B-072

11 illing off the germs of the moment If we leave the uh precipitating condition (,) {2 [2 it 2] will recur [3 (,) {3 3] so we want to clean them up thoroughly {4 [4 and 4] uh so what I'm going to do is get you an appointment with our hygienist as soon as you can [5 (,) {5 5] to get them clean and then ***come and*** see me for a general scrape round before you go to Japan

Example 7.7, ICE-GB: S1A-087

aid that okay we are giving in writing that we'll not trouble any teacher we'll be serious about our studies (,) But the same thing we find (,) {2 [2 still they are troubling 2]
$A [1 Or they would be given by 1] 1}
$A [2 Even I face 2] 2} the same problem (,) When I punish the student their parents ***come and*** ask me why you punished the

student (,) They don't want punishment to be given to their children (,) and they don't

Example 7.8. ICE-IND: S1A-085

nsed that him him always around cos them cannot find him cos he's one of the best marine corps ((word)) and everything (,) Mona was secret agent ((words)) known to assassin So he can be anywhere and them cannot find him And then uh also around maybe one year later him ***come and*** kill off all of the all of the family too You know she said she a female one No it's it's after that she escaped and she run away and everything And then she trained under a one different name and become a police
S1A-066.TXT

Example 7.9. ICE-JAM: S1A-066

Turning now to *and go*; the first row in Table 7.4 gives the frequencies of *and go* in all the ICE corpora. The subsequent rows give the frequencies of the ante and post co-selection components of *and go* in all the ICE corpora and all the colonies of the ICE corpora. The ante co-selection components include the collocation *leave* and the semantic preference **leave** as well as the collocation *come*. These two co-selection components are the most frequent across the ICE corpora as a whole. It should be noted that there would appear to be more instances of *come* with *and go*, than *go* with *and come*. I will comment on this in the following Section 7.4. The post co-selection components show the frequencies of the post pre-set collocates discussed in Section 6.3 – such as *back* in *go back*, and the frequencies of structural preference of **reported speech/thought**, also discussed earlier in Section 6.6.

		ICE-Can	**ICE-GB**	**ICE-Ind**	**ICE-Jam**
	and go	***56***	***27***	***33***	***44***
ante	*leave*	2	2	6	7
	semantic preference **leave**	4	5	1	17
	come	6	6	1	4
post	pre-set collocates	19	17	10	27
	structural preference **reported speech/thought**	10	0	1	0

Table 7.4. Frequencies of ante and post co-selection components of *and go*.

The concordances that follow give examples of the co-selection components featured in Table 7.4. Concordance 7.11 shows a selection of instances of the collocation *leave*¸ and the semantic preference **leave** from the colony ICE-Jamaica private. Concordance 7.12 shows all the instances of the post pre-set collocates in the ICE-GB colonies – these are marked in bold type. Additionally it also shows instances of the ante collocation *leave* and the ante semantic preference for **leave** – also in bold type. Concordance 7.13 shows a selection of the structural preference of

reported speech/thought from ICE-Canada private and the one instance from ICE-India printed.

Jamaica Private

7 6] 6} $A [7 But then to 7] 7} **leave** Spanish and go to I R it's not easy though $B I
8 you (,) but usually you have to **leave** here and go to another institution although and
9 that's an option maybe I should just **leave** and go study (,) and then later on maybe
10 re the providers So guys will **leave** school and go straight into a job start working
11 t was one thing Let's **change the topic** now and go on to something else ((laughs))
14 e had a class in the morning and then **left** and go into the production After that I
17 m problem and him say I **could not stay** (,) and go home at the time $A [Mhm] } $A 20
into the night Five o'clock I **leave** campus and go down to run a race at five o'clock 23
don't get it you're going have to **pack up** and go back home You have to plan ahead
27 [1 yeah 1] {2 [2 And you didn't 2] **get up** and go {3 [3 Mhm 3] $B [1 Right 1] 1} $B [2

Concordance 7-11. A selection of the ante collocation *leave* and semantic preference leave from ICE-Jamaica private.

GB Private

4 've already taken down out down the West End and go **and** see (( )) something fantastic

GB Public

5 ou know it's seen it over the centuries come and go **and** alas it still sees it today $A
6 e pregnancy and uh get it over and done with and go **and** start again (,) on the next
7 er is to def define it for one's own purpose and go **and** use that $B What frightens me
8 **iac note on which the curtain came down** and go **back** if we may to the beginning
10 elf for every minute of it to go on writing and go **on** working (,) And on other days it

GB Unscripted

11 e than is absolutely necessary should **leave** and go **into** employment without training
14 re's such a close uhm similarity Let me try and go **back** on the (,) right to convince
16 e dies the nutrients in it rot away uhm (,) and (,) go **back** into the soil and provide
17 it's clear look behind with a lifesaver (,) and go **And** once you're past the parked
18 ght oh I'm going to **get out of here** I'll go and (,,) go **and** find a hotel (,) And I 19
t that more and more healthy babies survive and go **through** childhood (,) But this

GB Scripted

21 ou and Cranley Onslow and th a third person and go **in** and say Margaret we've sen sed
22 s were instructed to put on their gas masks and go **into** sealed rooms as a protection

GB Non-Printed

23 t this I will buy them an answering machine and go **to** the beach . Anyhow I'm now

GB Printed

25 , reluctant to end one awkward conversation and go **back** to another . Even after
26 t Pension , and then **cancel your retirement** and go **back** to work . You cannot get
27 ell out in front of your body . Be positive and go **to** meet the ball . Remember :

Concordance 7-12. Post pre-set collocates of *and go* in ICE-GB.

Canada Private

6 ugh))] } $A { [((words)) gonna stand up and go **hey** that's not true I feel guilty
7 gh)) $A Well $B { [Just wander up] to you and go **hey** what are you doing $A [((
9 jerk $B Ya ((laugh)) And of course I call and go **oh god** I called you again I was
10 mostly (,) a piece of art that you look at and go **oh** isn't that beautiful { [and]

India Printed

4 et Gray said ` Eat drink, be merry yourself and go **the deuce** - - if there be a deuce"

Concordance 7-13. Structural preference of reported speech/thought from ICE-Canada private and ICE-India printed.

The semantic force arising from the node *and go* and the semantic preference **leave** (including the collocation *leave* on the basis that collocation is a sub-division of semantic preference) is **depart from/leave behind in order to move into something new and unrelated – an emphasis of end before a new beginning.** The concordance lines show that, as with *to go to*, that there could possibly be semantic forces that work in tandem. An ante semantic force associated with post semantic forces related to the following pre-set collocate – for example, *back*, *into* or *to*. The following concordances show that this is indeed so.

The previous chapter identified a semantic force for *go out* of **movement from a location to undertake a mentionable action**; for *go back,* **a return to a non-specific location and/or time**; for *go to*, **movement to location non-specific to participant**; and, for *go on*, **continue.** In Example 7.10 the writer describes how they would leave their bed and *start walking around our neighbourhood* in the middle of the night - and end of sleep before undertaking something that is not altogether normal. The speaker, in Example 7.11, seeks to change the subject suggesting that they return to the beginning. While there is a specific point to which the speaker would like to return to, they are hesitant that this might be possible by the hedge *if we may*. It is less specific than if *come back* had been utilised, so the hedge could be considered to be included in the use of *go back* as well. I would suggest that this might be an instance where the semantic force is utilised to accentuate the tentative nature of the request. The narrator in Example 7.12 describes how his friend eventually abandons *the country* for *the USA* – a new beginning in a location that is non-specific to either the narrator or his friend. And, in Example 7.13, discussion is ended on one topic before continuing (*go on*) with another topic. Also, there is the deictic semantic force of going to something non-specific to the participants.

nking of all that effort, for what? For what? What's wrong with us? As we pulled away into the dark I started to cry, and then caught my self, crying over a spider and a rusted Malibu. Somehow it didn't seem absurd. I cried a lot that year. Sometimes I couldn't sleep at night and I would get up ***and go*** out and start walking through our neighbourhood. Nor mally winter is torture for me, but all that winter I was warm and would walk around with my coat unbuttoned and without any mittens. When the snow came, it came in huge soft flakes, laying down bright white blankets over each shabby house, gen

Example 7.10. ICE-CAN:W2F-009

is Joe that when uh (,) the old problems recur I reach for the old solutions and I don't have anything new to offer I thought that was a very significant and perceptive remark from a man who claimed he had nothing new to offer
$A Well let's leave that elegiac note on which the curtain came down ***and go*** back if we may to the beginning Roy Jenkins could I ask you in nineteen sixty-three was there a serious alternative to Harold Wilson being Gaitskell's successor or was his choice absolutely inevitable (,)
$D No I don't think it was absolutely inev inevitable

Example 7.11. ICE-GB:S1B-040

at he found solace in my company. I kept on visiting him almost everyday and he used unfold his agony. Another year in 1964-65 when I was Research fellow at the C I E. ! How often we met. Then the storm had not blown over. He was still unhappy and finally the plunge he took to leave the country ***and go*** to the USA. What a loss. H.B. has several vols of poetry in English to his credit. I wonder if we could do something to perpetuate his memory at W.C. Is it possible ? Least ways would you be able to write an article on him ? If I get some information I will pass it on to you. But it

Example 7.12. ICE-IND:W1B-006

Alright You wish it in a ((((word(s))) Trust me that administrating this university is such a task (,) Anyway we we're finishing up the financial situation thing So that was one thing Let's change the topic now ***and go*** on to something else ((laughs)) Yes so hm the point I was making I think was that people are now listening to somebody's radical suggestions like thinking outside of the box and those sorts of things For example here at Carimac we've never time-tabled a class like a theory related class in the a

Example 7.13. ICE-JA:S1A-018

This section has shown that the post and ante semantic forces can again be seen to function in tandem with each other with both *come and* and *and go*, and again the choice of node is influencing the resultant co-selection components. In the following section, I show that not only do semantic forces function in tandem with each other, but they can function in layers.

7.3 come and go

The following co-selection components have been identified that are relevant to this section:

Node: *come*
Semantic preference: locations specific and non-specific to the participants
Semantic force: movement associated with the participants or towards their location

Node: *go*
Semantic preference: locations specific and non-specific to the participants
Semantic force: movement from a specific location towards a non-specific location

Node: *come and*
Colligation: Post verb
Structural preference: surplus to requirements, optional repetition of pronoun
Semantic preference: post action/social action
Semantic force: movement to undertake an action/social action that will happen after the utterance with emphasis on the participant(s) being physically present in a place specific to the participants.

Node: *and go*
Colligation: *leave*
Semantic preference: leave
Semantic force: depart from/leave behind in order to move into something new and unrelated – an emphasis of end before a new beginning.

Node: *and go*
Colligation: post discourse marker
Semantic preference: post reported speech or thought
Discourse preference: ICE-Canada
Semantic force: this is what was said or thought

The frequencies of all the instances of *come and go*, (*come* *) *and go*, and *come and* (* *go*) taken from ICE-Canada, -GB, -India and –Jamaica are shown in Table 7.5 and the instances are re-produced in Concordance 7-14. The concordance lines are sorted into the different ICE corpora and colonies in which they occur. While *come and go* do not appear in all of the ICE corpora, there are examples of them in each of the colonies investigated. Each concordance line, except line 8, is 140 characters long in order that there is normally a sufficient quantity of text to allow for an adequate examination. I have included additional text in line 8 as I analyse it further at the end of the section. Where I consider that the text is not sufficient, I have included a further explanation in italicised brackets after the concordance line. (Of course, this is a subjective view point and there might be some lines that I consider to have sufficient text for understanding and others will not.)

Canada Private

1. ike less than two weeks
   $A Ya
   $B Anyway the deadline's gonna *come and go* and then you're gonna be kind of disappointed later
   $C It doesn't h
2. not done it for a few years But this would be like a constant And you *come and* you *go* and and you use it to touch base and (,) it (,) might *(discussion about regular visits to a counsellor over time)*

Canada Public

3. words))
   $A Sorry I didn't mean to cut you off You You could *come and go* as you please though
   $B Yeah
   $A Nobody'd ever tell you what to do $ *(visiting a property)*

Canada Unscripted

4. te seventies the very late seventies And so I've seen trends *come and go* and different attitudes and whatnot It i (,) In spite of my ability

5. re very torn between this And then you have the union who *come* in *and go* (,) you know you're a member of the union now and (,) and things yo

Canada Non-Printed

6. use sometimes I'm away too. Good thing about volunteer jobs; *come and go* without asking. I always *go* on my arranged-for day when I'm here. I

Canada Printed

7. ge, by focusing on elements old and new. Man-made structures *come and go* in relatively rapid succession. A Halifax city block is slated for *(discussion of how photography can capture time change)*

GB Private

8. the family of Arne had a bay grey (,) that size with((unclear-words))and they used to have a small rural farm with big Wyandotte ducks [2 (,) {2 2] you see Well (,) at uh you know where the toll house is over at Huxley (,){3 [3 Then you 3]you know that that th there used to be a stream of water in the gutter running down there perpetual it never stopped {4 [4 and 4] it used to come down from the moor (,) down by s by the the I've got to use my hand to show it down by the uhm Royal Oak (,) underneath the Royal Oak and the big chute affair and then come out and go down in the gutter [5 (,) {5 5] And the ducks used to come out there every day (,) And that is how it originally was named as Duck Street
   $B [1 How was that 1] 1}
   $B [2 Uh uhm 2] 2}
   $B [3 Yeah 3] 3}

GB Public

9. { [conflict] And you know it's seen it over the centuries *come and go* and alas it still sees it today
   $A [Uhm] }
   $A Indeed indeed But u

10. usiness a good number of years Haven't you seen these crises *come and go* before Isn't this an age-old story
    $C The arts are always in crisis

GB Unscripted

11. es one though even that has seen some large corporations (,) *come and go* since since since its founding Uh (,) ((clears-throat)) and one s *(longevity of institutions)*

12. us about the cycle of vibration how the vocal folds *come* together *and go* apart (,) And I'll show you some examples from four different voice

GB Scripted

13. its advantages in an intensely volatile region where regimes *come and go* even if in the end time cannot halt the flow of advanced knowledge

India Private

14. us will be crowd again (,) and we'll *come* at ((one word)) again *and go* (,)
    $B Yeah who ever has *come* no let's get back (,)
    $A They said
    $B *(meaning not clear – but probably relates to catching a bus)*

Jamaica Private

15. sit with them in meetings et cetera So you can't just *come* here *and go* to your classes and not do anything You have to be involved especia

16. sh it's the longest one of {1 [1 the W B shows (,) 1] Others *come and go* {2 [2 and it's still there 2]
    $B [1 W B show I have to say cos 1] 1*(length of time a particular show had been broadcasting)*

Jamaica Printed

17. She was extremely happy that she had a house where Bim could *come and go* as often as he wished. She went to her friend Jackie and said to he

18. business, yuh mek mi feel miserable at times. *Come* out a mi life *and go* bout you business." Bim left, feeling to himself that Janet was jok

Concordance 7-14: All examples of *come* * *and go*, *come and go*, and *come and* * *go* taken from ICE-Canada, -GB, -Jamaica, and –India.

Table 7.5 demonstrates that depending on the choice of node, different combinations of *come* and *go* are identified. Both *come and* and *and go* will reveal *come and go* but they will not show each other. This would suggest that the way in which *come and go* are identified in the concordance will shape how the co-selection components are expressed.

ICE	Canada	GB	India	Jamaica
come and go	5	4	0	2
come and (go)*	1	0	0	0
*(come *) and go*	1	2	1	2

Table 7.5. Frequencies of *come and go, come and,* and *go and* across all the ICE corpora.

I would suggest that whether *come and go* is used in a physical sense – lines 3, 6 and 17 – , or in a more abstract sense – lines 1, 4, 7, 9, 10, 11, 13 and 16, there are a semantic preferences of **time and/or cycles,** for example *late seventies*, *relatively rapid succession*, *a good number of years*, *longest* and where there is physical

movement, a semantic preference of **unhindered** is present – *as you please, without asking,* and *as often as.* I would suggest that this semantic preference seems to be implicitly suggested in the more abstract forms. The events could not have happened on a cyclical basis if they were hindered in some way suggesting a semantic force of **unhindered repetitious movement from a non-specific location/state to a specific and back to a non-specific location/state over a time period**.

There is only one instance of *come and * go* and this is concordance line 2. There is also a semantic force of **unhindered repetitious movement from a non-specific location/state to a specific and back to a non-specific location/state over a time period**. This can also be considered to be an instance where there is a structural preference of an optional repetition of a pronoun as identified in the concordances of *come and.* This structural preference is not exclusive to *come and go* but occurs with other verbs associated with *come and* so, I would suggest, should be considered as such.

When one examines *come * and go,* there seems to be some variation of co-selection components. Concordance line 5 is of the type that has a discourse marker as a colligates (*you know*), post reported speech as a structural preference and thus a semantic force of **this is what was said or thought.** The *go* is a substitute speech verb.

Lines 8 and 18 have ante semantic preference of **leave –** in both cases this is indicated by *come out*: line 8, describes water flowing, coming out of, or leaving, a place and going to another place; and, in line 18 the speaker demands the hearer leaves the speaker's life *and go bout you business.* However, I would suggest that while the co-selection components are commensurate with that of *and go* and the semantic preference **leave**, there is evidence that in line 8 there is also co-selection components associated with *come and go* as well as well as deictic *come* and *go.* It is as if four different co-selection components are active in this exchange. The co-selection components are layered.

The water *never stops* and it goes on *every day*, showing a semantic preference of **time and/or cycles** and of the action being unhindered. This suggests and underlying semantic force of **unhindered repetitious movement from a non-specific location/state to a specific and back to a non-specific location/state over a time period.** These are the co-selection components of *come and go.* Additionally, the location in which this occurs is important as that is where the ducks visit regularly, and that is why the road was *named Duck Street.* The water flows towards (*come*) this specific location, and then it flows away to somewhere else (*go*). This has traces of the semantic force of deictic *come* and *go* in that it is explicitly pointing at a specific place that is relevant to the narrative.

The text in line 14 is not clear (even after examination of additional co-text), but it would appear to relate to taking a bus journey, and, as such, describes some sort of sequence of movement. Only two of the *go*-grams can be linked with the same semantic preferences and forces detected with *come and go:* line 13 describes the movement of *vocal folds* during speech, and as such there is no obstacles and, in being cyclical, there also a time element; and, in line 15, there is an indication that it is not possible to just come and go unhindered – another instance that would appear to be, although negative, analogous to the semantic force of *come and go.*

Come and go has a semantic force of **unhindered repetitious movement from a non-specific location/state to a specific and back to a non-specific location/state over a time period**.. There is also evidence to suggest that instances of both *come and * go* (line 2), and *come * and go* (lines 12 and 18) also have the same semantic force. I would now suggest that there is evidence that *come and go, come and * go,* and *come * and go* can all have co-selection components that are linked with *come and go.* However, it must also be noted that it is somewhat dependant on the choice of the node what co-selection components are highlighted. For example while no further claims can be made about *come and * go* as there is only one

example in the four corpora, *come * and go* exhibits other semantic forces that have also been identified in relation to *and go*, although in at least one instance when one examines more of the surrounding text, it would seem that there is a combination of semantic forces that relate to four different sets of co-selection components.

Furthermore, I would suggest that there are always strong connections between the semantic forces identified for *come and go,* deictic *come* and *go.* Deictic *come* has a semantic force of **movement associated with the participants or towards their location,** and deictic *go* **movement from a specific location towards a non-specific location**: *come and go* indicates movement or change of state from the non-specific to the specific and back to the non-specific. If one had examined *come and go* through the spotlight of either *come* or *go* the resultant semantic force would have probably have not included specific or non-specific location/state as it is only the examination of both verbs that has highlighted this difference. It is the co-selection components of different nodes that are giving a slightly different slant to the resultant semantic forces. With *come and go* the resultant semantic force is not just one adding to another but combinations of co-selection components have produced layers of semantic forces, and the allied co-selection components, that are only identified by examining the concordances generated by different nodes.

In summary

The investigation of *to go to* and *come and go* in this chapter has revealed that

- The ante and post *come*- and *go*-grams or pre-set collocations work in tandem with each other.
- Each word or combination of words in the node has the potential to have co-selection components that, when examined in concordance, can create identifiable layers of semantic forces.
- This leads to the hypothesis that the selection of the node always impacts on the co-selection components identified.

Chapter 8 The lexical item revisited

All words have the "taste" of a profession, a genre, a tendency, a party, a particular work, a particular person, a generation, an age group, the day and hour. Each word tastes of the context and contexts in which it has lived its socially charged life; all words and forms are populated by intentions
.
Bahktin 1981: 293

In the previous chapters I have examined the *come*- and *go*-grams that occur above or equal to a frequency of 40/million in the spoken ICE corpora. In Chapter 5, I included all the *come*- and *go*-grams, but in Chapter 6 I narrowed the selection to only those that occurred at this frequency in all the ICE corpora. In Chapter 7 I narrowed the focus to only four *come*- and *go*-grams in order to show that the choice of node influences the resultant co-selection components identified. All the results of this investigation are shown in Appendix X. The first column gives the node, which is then followed by collocation, colligation, structural preference, semantic preference, discourse preference and, finally, semantic force.

In this chapter I discuss, in light of my research, firstly the co-selection components of the lexical item, and then I consider what this might mean in terms of the whole.

8.1 The Parts

8.1.1 Collocation

Collocation was sub-divided into those adjacent collocates of *come* and *go* that occurred in the ICE corpora above or equal to a frequency of 40/million and used in the *come*- and *go*-grams as pre-set collocates, and those that were co-selection components of these *come*- and *go*-grams. The nodes were further divided into those that had collocations preceding the *come* or *go* and those that followed it. I called the former, ante pre-set collocates, and the latter post pre-set collocates. From these nodes co-selection components were identified and the semantic force of the resultant combination of co-selection components posited.

The co-selection collocations included those that were an indication of a potential semantic preference such as *want* for *to come* or *to go*, or *leave* with *and go*, and those that with their addition to the node produced a shift in the semantic force, for example, *with* with *come up* and *there* with *go in.* I am inclined to suggest that where the collocation is obviously an indication of a potential semantic preference it should be taken as part of the semantic preference, but where it is not consideration should be given to re-classifying it as part of a pre-set collocation, in other words, part of the node for generation of additional concordances. For example, *leave* is a collication for *and go* but it is also a semantic preference, in that there are other word groups that occur that also have the equivalent meaning, but the collocate *with* that occurs with *come up* has no corresponding word groups with the same meaning. One ends up with those collocations that are to be included in the node, and those collocations that are a sub-division of semantic preference.

With regard to the ante and post pre-set collocations, I think it worth noting here that there seems to be quite a distinction between them in terms of the resulting co-selection components. The ante pre-set collocates have a tendency to generate a greater amount of ante co-selection components, and the post pre-set collocates, a greater amount of post co-selection components. And, whereas the ante pre-set collocates tend to have similar co-selection components for both *come* and *go*, the post have distinct differences. This would suggest that Gries' proposal for a new measure for collocation, ΔP, that gives information with regard to direction as well as attraction and/or repulsion strength will be of great value in collocation studies (Gries 2013: 152). The calculation of the relative strengths of the ante and post pre-set collocates might shed further light on this phenomenon.

However, it is my conjecture that this could be a good indication of prospection in language and/or a by-product of the original selection of nodes. In terms of prospection, for example, the semantic preference of **want** with the pre-set collocation *to* prospects a verb of movement. Then, once the verb of movement, *come* or *go*, is selected this then can prospect a post collocate with the related co-selection components. With regard to the original choice of nodes: concordance lines were generated using either *come* of *go*, and the subsequent pre-set collocates were generated from these concordance lines. Thus, the focus of the pre-set collocates could be said to be hinged on *come* and *go*, leading, perhaps, to the conclusion that the choice of node can influence the resultant co-selection components identified.

8.1.2 Colligation

As I stated in Section 6.1.1, I have restricted colligation to word classes and introduced a new co-selection component discussed in the following section, structural preference. I would suggest that this helps avoid the problem that I identified in Section 2.2.2: if the identification of word class is dependent on the grammatical structures in which a word is found, then if one is identifying colligation as co-occurrence of word class or grammatical structure the whole definition process is circular. In other words, if one wishes to identify the class to which a word belongs – its colligation, one examines the grammatical structure in which it is found – which is again, its colligation. One is using the colligation of a word to identify the colligation of the word. The preference for particular grammatical structures is, I suggest, a structural preference not colligation.

The co-selection colligations identified are verbs, nouns and discourse markers. The pre-set collocates could be grouped into two sets of colligates. The main ante pre-set collocate is pronoun – *I, they*, *we* and *you.* The main post pre-set collocates are multi-word verb particles – such as *in, on, through*, and *back*. This would suggest that one could consider using pre-set colligates (word classes) in the node (this would not possible for this research as the ICE corpora used are not POS tagged) rather than pre-set collocates. Additionally, it appears that, at least for *come* and *go,* functional colligates are more likely to be found in the node and lexical colligates are more likely to be classified as co-selection components. That does presume, however, that discourse markers are more lexical than functional. This is something that is under consideration; Carter and McCarthy (2006) suggest that discourse markers should be classified as an additional word class to nouns, verbs, adverbs, etc, but Biber *et al* (1999) are inclined to classify them as an additional major word class that lies between function and lexical.

8.1.3 Structural Preference

I believe that this new co-selection component is required as, not only does structural preference include traditional grammatical structures such as noun phrases (*go to*) that follow *go to,* but other structures that might not be considered so grammatical: ellipsis (*come in*) and surplus (*go and*), list (*come back*), repeated pronouns (*you come and you go*), hesitation (*go back*), reported thought/speech (*and go*), and qualification and backchannelling (*come on*). These structural preferences are identifiable in the text so, I would suggest, that they should be taken into account when delineating the co-selection components. Also, in introducing this category, it does mitigate the problem of word class identification by grammatical patterning.

8.1.4 Semantic Preference

As I stated earlier I have extended the scope of semantic preference to include both the co-occurrence of a word or words of a particular meaning regardless of word class with the node, and the co-occurrence of words belonging to particular semantic sets: for example, sports terms (*come* and *go*), and types of objects (*come in*).

I am increasingly of the opinion that semantic preference has been much misjudged in the literature. Hunston suggests that it should be confined to evaluative meaning (Hunston 2007: 266), and Bednarek advocates sub-divisions of positive/negative collocation and semantic collocation. While Xiao and McEnery, as I have already indicated, suggest that semantic preference has a distinct collocational meaning that is 'a feature of the collocates' (Xiao and McEnery 2006: 107). However, Hoey does suggest that semantic preference should be seen in terms of semantic sets (Hoey 2005: 23).

My research indicates that there can be evaluative meaning: there is vagueness (*go back*), there is importance (*go into*), there is better plans, schemes and ideas (*come up*), there is obligation and necessity (*have to go*), and there is desire (*to come*). There is also positive/negative meaning: the positive outcome signalled by *come out*, and the negative state often arrived with *go through.* There is also, if I interpret Xiao and McEnery's definition correctly, some semantic preferences that have a connection with a particular collocation: both the collocates *leave* and *want* are also a variety of the semantic preferences **leave** and **want.** However, there are other semantic preferences that do not have a corresponding collocation: class/gender (*come up*), specific or non-specific location (*come* and *go*), and measurement of time/distance (*to come* and *to go*). Are these collocational features in that there is a re-occurrence of meaning, or are they something different as they are not a feature of the collocates?

It seems to me that semantic preference is all these things and more. It is the predilection of a node to associate with a word or words that are semantically equivalent and/or with a word or words that belong to semantic sets. There is ample evidence that word(s) do hang around in semantically similar environments and I would suggest that as broad a definition as possible is of the greatest of use. I suppose one could sub-divide the semantic preferences according to the various categories identified above, but I can see no reason why it should be necessary.

I would also suggest that it is often the semantic preference of the node word that is at the root of the subsequent semantic force that is identified. It is the semantic preferences of the node, the semantic environment about which the node hangs, that regularly informs the resultant semantic force except where the node would appear to be organising the discourse. Where the node is utilised to indicate temporal shift (*come to*), to backchannel (*come on*), to amplify (*go and*), or to introduce reported speech/thought (*and go*), there are structural preferences but there is no apparent semantic preferences. I would be inclined to suggest that where *come* and *go* are

used functionally – to organise rather than to mean, it is indicated by the absence of semantic preferences. However, it should be noted that when the *come*- and *go*-grams are used to manage – explicitly inform the receivers what has or is going to happen in the discourse – semantic preferences are present.

8.1.5 Discourse Preference

Discourse preference is the predilection for the node and, but not necessarily, a particular combination of co-selection components to a particular discourse type. The statistical investigation in Chapter 5 established that there was a significant difference in frequencies of occurrences within the colonies of the ICE corpora, but there was no such significant differences between the ICE corpora themselves, although additional pairwise investigations did show, for example, that the frequency of *come to know* was significantly different between ICE-India and the rest of the ICE corpora and that the (spoken) scripted colonies were much more homogenous with the written colonies than the spoken colonies. However, what is important is that no account of differences in co-selection components was revealed, only differences in the frequencies of occurrence of the *come*- and *go*-grams.

Chapter 6 established that, although there are significant differences in frequencies across the colonies the co-selection components were, and this must be stressed, mainly consistent across all the ICE corpora and all the colonies. Some of the differences between the ICE corpora were cultural – in particular, the approach in ICE-India to class, gender and roots (*come from* and *come up*), but not all of them. The use of *go* to indicate following reported speech or thought appeared to be almost exclusively associated with ICE-Canada which I find surprising. I cannot comment on Indian or Jamaican Englishes, but I would certainly think, from a "native" speaker perspective, that it is used in Great Britain. Perhaps it is just a much more common usage in Canada and thus it is occurring in the ICE-Canada but not in the other ICE corpora.

Come to know is not only an example of a *come*-gram that is significantly more frequent in ICE-India, but it also only occurs in the spoken colonies of that ICE corpora. I would suggest that this is a form of *come*- and *go*-gram that is emergent in terms of frequency. This is the only *come*- and *go*-gram that I have identified as new, but it is quite possible that there are other emergent *come*- and *go*-grams that are occurring at a lower frequency than 40/million. It is interesting that this emergent *come*-gram occurs exclusively in the spoken component of the ICE corpora and not the written. This might suggest that language change has a tendency to begin in spoken language not written language. However, as I have already noted, the ICE corpora do not contain computer-mediated discourse (CMD) and additional research into innovation here would be useful to ascertain whether there is also a tendency for language change to occur here. If this is so, it should be important that research of language should at least include spoken language, and probably, in an ideal world, more spoken language than written language and CMD. Additionally, with regard to further research, these particular ICE corpora were collected nearly 20 years ago, so it would be interesting to see if this particular *come*-gram is now used more generally across all World Englishes, or still confined to Indian English at such high frequencies.

Other *come*- and *go*-grams, like *come to know,* also appear to have discourse preferences within the ICE corpora, for example, *come on* has a preference for dialogue and the discourse managers, predictably I would suggest, occur predominately in spoken language. However, I would suggest that the most striking discourse preference is the live radio sports reporting that occurs across the ICE corpora.

This type of language use must be considered a new phenomenon. The live reporting of sports events on radio began in the early Twentieth Century and the co-selection components of the *come*- and *go*-grams that indicate the deictic shift that would appear to be occurring in this type of reporting are equivalent across all the ICE corpora, and all the sports that are reported. This, I believe, not only shows that a discourse preference is a useful category of co-selection component of the lexical item, but also shows that language, when necessary, can very quickly adapt to the new. This, I would suggest, is an example of language as a local practice (Pennycook 2010) – the language has adapted to meet the needs of live reporting of sport.

8.1.6 Semantic Force

Semantic force is the reason why the particular set of co-selection components have been utilised. I have previously suggested that semantic force could be considered to have three elements based on those identified in relation to Linear Unit Grammar (LUG) – interactive organisational elements (OI), text-oriented elements (OT) and message-oriented elements (M) (Sinclair and Mauranen 2006). The speaker/writer wants, with the particular set of co-selection components chosen, to relay a particular message; or they want to organise the interaction or they want to organise the text. I would suggest that the *come*- and *go*-grams investigated demonstrate all these aspects of semantic force. However, what has also become apparent is that the semantic force is, to some extent, tied to the selection of the node. I have shown that the semantic force of a particular node can be the same as the semantic force of two separate nodes working in tandem, and I have also shown that the choice of the node can influence how one sees the subsequent semantic force – aspects that might have been picked up with one node might not have been with a different one.

The majority of the *come*- and *go*-grams have a semantic force that is integral to the message of the texts. They are, at their very basic, an indication of movement to or from a location; for *come* this is a movement that is linked to the location of the participants and for *go* this is a movement that is not linked to the location of the participants. The former tends towards the usual, the finite, ends, the routine, the related, sureness – specific to the current actions/locations/movements of the participants, while the latter tends toward the unusual, the infinite, beginnings, the new, the un-related, vagueness – non-specific to the current actions/locations/movements of the participants. I *come back* to a specific point in the discourse or I *go out* to do something significant that I am not currently doing: I *come to* some sort of conclusion or I *go to* a new place.

Some *come*- and *go*-grams organise the interaction, and these have a preference for spoken language or for dialogue. There are the discourse managers, the *come*- and *go*-grams that have a semantic force that indicates to the receiver what is going to, or has happened in the text. There is the discourse marker, *come on* that has a semantic force of either an exhortation to re-think, or exhortation to continue as a back channel. And, there are the *go*-grams that indicate a following reported speech or thought.

There are also the *come*- and *go*-grams that have a semantic force related to text organisation. All *come*- and *go*-grams have this potential in that they can have a semantic force that relates to the deictic centre of the text. They point to the deictic centre. However, in addition, there are those *come*- and *go*-grams that appear to have a semantic force that indicates temporal shift or emphasis of the following text. The semantic force of *come and* indicates an action that is going to happen, and emphasises that the participants are physically present. The *come*-gram is, I suggest, surplus to what is needed for understanding hence the fact that I suggest that it must have some text organisation usage. The same can be said for *go and* in that it has a semantic force of amplification of the action that follows it, and is surplus to the

coherence of the message. *Come to* also has a semantic force, when followed by a verb, relating to text organisation – **progression towards a time when something will occur.** As with *come and* there is the indication of a temporal shift from the time of the utterance to a time that is later than the utterance time.

Turning now to the connection the semantic force has with the node: the evidence would suggest that the choice of the node influences the resultant co-selection components. I touched on this in Chapter 6, and in Chapter 7 I confirmed that it would appear that this is very much the case. The examination of *to go to* demonstrated that the co-selection components constituted those of *to go* working in tandem with those of *go to.* The co-selection components of *to go* are associated with not just the post *to* but with other post word(s), and the co-selection components of *go to* are not just associated with the ante *to*, but with other ante word(s). Having established that semantic forces can work in tandem with each other, I then examined *come and, go and* and *come and go.* Again, it can be seen that the semantic forces of *come, go,* and *come and* are contributory to the semantic force of *come and go*. The examination of the different *come*- and *go*-grams reveal a layering of meaning that interacts, combining or adding to, with each other. This can also been seen, I would suggest, with *and go.*

I expressed the opinion that the co-selection components of *and go* could be contested. Should *come* and *leave* be considered to be examples of a semantic preference of **movement** or should they be considered separately – a collocation of *come,* and a semantic preference of **leave**. The semantic force of *come and go* will depend on which of these view points is adopted. If it is the former then *come and go* might be up for consideration as a separate lexical item with *come* as a pre-set collocate of *and go,* but if it is the latter then *come and go* should be considered as one example of a lexical item that includes other ante verbs of movement and *and go.*

8.1.7 Core

In Chapter 2, Section 2.2.5, I concluded that with the generation of concordances of a particular node, a potential lexical item can be identified. This item is only a potential item as the node is not necessarily a fixed feature of the lexical item. The semantic force of this potential item can then be used to identify variations in the item, and thus identify the core. I suggested that the node and the core have separate identities. While I do not dispute that the lexical item, as described by Sinclair, would fit these parameters, the evidence of this research suggests that the choice of node impacts on the resultant co-selection components. I would suggest that it would be better to re-locate the lexical item and, in so doing, the node rather than the more nebulous core should be seen to be integral to the location of the lexical item.

8.2 The Whole

In the previous section I have discussed each of the co-selection components individually. In this section I discuss them as parts of the lexical item.

When a concordance is generated using a particular node, the node is a fixed feature of the paradigmatic axis and each particular line is a feature of the syntagmatic axis. It is from the individual concordance lines that the collocation, colligation, structural preference, semantic preference, discourse preference and semantic force are found. However, if one considers the node to be the independent variable, and the other co-selection components the dependant variables it is possible to explain why

the choice of node influences the resultant co-selection criteria. The independent variable is fixed on the paradigmatic axis and this then produces a snapshot of the co-selection components on the syntagmatic axis. As the independent variable is changed – another node is fixed on the paradigmatic axis, so there is a different focus on the syntagmatic axis. The dependent variables, collocation, colligation, structural preference, semantic preference, discourse preference and semantic force, shift. Each shift in node focuses on a different layer, a different sedimentation, of language.

A lexical item is then an item in which the node is integral. A lexical item is the item that is identified when a node is generated by a concordance. It consists of an obligatory node and semantic force, and optional collocation, colligation, structural preference, semantic preference and discourse preference. This new definition of the lexical item might appear to be similar to that given by Sinclair. He suggests the lexical item consists of an obligatory core and semantic prosody, with optional collocation, colligation and semantic preference. The addition of two optional co-selection categories and the substitution of semantic force for semantic prosody would, in themselves, have only been modifications to his model. However, in dispensing with the obligatory co-selection of the core and the use of the node instead signifies a radical departure from Sinclair's model. This differs from Sinclair's definition of the lexical item in that he assumes the lexical item is a unique object contained in text, whereas I am suggesting that the lexical item is an object that is observed when one looks at a text through a unique focus – a particular node. If one shifts the focus, one shifts the lexical item. The former, Sinclair's model, is a feature of text, the latter, the model proposed by this research, is a feature of how one looks at the text. The paradigmatic choice is constraining the syntagmatic results.

In order to illustrate this in practical terms, I return to the Concordance 7.14 line 8, shown again here, and I then re-assess the possible links with the Carter (2004a), Pennycook (2010) and Wray (2008).

```
the family of Arne had a bay grey (,) that size with(( unclear-words
))and they used to have a small rural farm with big Wyandotte ducks [2
(,) {2 2] you see Well (,) at uh you know where the toll house is over
at Huxley (,){3 [3 Then you 3]you know that that th there used to be a
stream of water in the gutter running down there perpetual it never
stopped {4 [4 and 4] it used to come down from the moor (,) down by s
by the the I've got to use my hand to show it down by the uhm Royal Oak
(,) underneath the Royal Oak and the big chute affair and then come out
and go down in the gutter [5 (,) {5 5] And the ducks used to come out
there every day (,) And that is how it originally was named as Duck
Street
$B [1 How was that 1] 1}
$B [2 Uh uhm 2] 2}
$B [3 Yeah 3] 3}
```

I established that it depended on the node as to what co-selection components might be identified. In this example I determined that there were four lexical items that, depending on which node was chosen, were functioning over this stretch of the text containing *come out and go down*. I suggested that the semantic forces appeared to be layered. There was the semantic force associated with the node *and go* and a semantic preference of **leave**, the semantic force associated with the node *come and go*, and the semantic forces associated with nodes deictic *come* and *go*. Each node selection gives different co-selection components, all of which have relevance to the meaning in this stretch of text. Sinclair, however, suggests that a language consists of lexical items that have considerable internal variation that vanishes when an appropriate level of abstraction is invoked (Sinclair 2004a: 35). He also suggests that texts consist of strings of lexical items that are statistically autonomous (Sinclair 2004a: 39). If this is the case then I am not sure how one accounts for the presence of four lexical items over the same part of the text each contributing individually their own semantic force. The evidence would suggest that Sinclair's account of the lexical item does not necessarily account for what would appear to be happening in the text.

In Chapter 2 I suggested that Sinclair's lexical grammar could have theoretical links with Carter (2004a), Pennycook (2012) and Wray (2008). Carter argues that while language is creative, or perhaps re-creative, there is a stream of flexible formulaic expression that runs parallel to the creativity, stabilising and routinising language (Carter 2004a: 48, 133). Similarly, Pennycook advocates considering that creativity and repetition in language are the norm, and that grammar is a product of a repeated sedimentation of form. I would maintain that the links I suggested existed with Sinclair's view of the lexical item are still valid, perhaps more so, with mine. In suggesting that the co-selection components can work together in different ways would suggest a greater integration of the creativity and the repetitious or formulaic aspects of language.

Like Sinclair, Wray (2008) proposes that language consists of variable units which she calls Morpheme Equivalent Units (MEUs), but she has reservations with regard to Sinclair's idea that language exists on a continuum from the terminology to the phraseological tendency. She does not think this model is adequate for the processing of language, suggesting that language users store MEUs as parts as well as wholes, and process language on a Needs Only Analysis (NOA) basis. While I would suggest that it is questionable whether MEUs exist, the conclusions reached in the research would fit better with Wray's NOA. If one makes the hypothetical jump by contextualising the theoretical and the psychological, from the idea that each node has the potential to have distinct co-selection components to, and I adapt this from Hoey (2005), each node is primed for the individual user to have distinct co-selection components, language processing on NOA basis is quite feasible. The individual language user could be chunking language so that when a particular *n*-gram (node) is recognised, with the appropriate co-selection components, understanding is reached.

In summary

In this chapter I have discussed the co-selection components individually and I have then discussed the lexical item as a whole. The conclusions I have come to are presented in the following chapter. I have also suggested that links can still be made to the other theoretical paradigms I introduced in Chapter 2: some links are stronger and some links are weaker, but they are still evident. In the final chapter, I will be discussing possible links within the corpus linguistic paradigm.

Chapter 9 Conclusion

> *An Introduction is to introduce people, but Christopher Robin and his friends, who have already been introduced to you, are now going to say Good-bye. So this is the opposite. When we asked Pooh what the opposite of an Introduction was, he said "The what of what?" which didn't help us as much as we had hoped, but luckily Owl kept his head and told us that the Opposite of an Introduction, my dear Pooh, was a Contradiction; and, as he is very good at long words, I am sure that that's what it is.*
>
> *Milne 2001: 124*

When I began this research my initial view was that I would hope to modify Sinclair's model of the lexical item by examining high frequency verbs between and within different World Englishes. I find instead that I am challenging his model and offering something that is fundamentally different. His model considers the lexical item to be an object in the discourse. I am suggesting that the lexical item is determined by the paradigmatic choices made – it is generated from the syntagmatic axis by a specific word or words selected from the paradigmatic axis. By changing the word or words one changes that which is realised. The lexical item is determined by the paradigmatic focus on the syntagmatic axis, and, as such, could be considered to be a theoretical construct.

In this chapter I first summarise the research conclusions and I answer the research questions posed, I then consider the empirical problems associated with this research. I continue by suggesting further research that could be undertaken within this paradigm and additional research that could be undertaken as a result of my conclusions. I end by suggesting what this research might signify in terms of the investigative and theoretical corpus linguistic strands, bar lexical priming, introduced in the first chapter. As the aim of lexical priming is to contextualise 'theoretically and psychologically Sinclair's insights about the lexicon' (Hoey 2005: 158), and as this research has determined that there are problems with these insights, links between this research and lexical priming is difficult. In the last chapter I did suggest that one might conceptualise each node as being primed for the individual user to have distinct co-selection components, but in so doing I emphasised that this was a hypothetical jump and not necessarily supported by this research which remains neutral on the cognitive language processes.

9.1 Research conclusions

I started this research by examining *come* and *go* quantitatively before examining them qualitatively. The quantitative investigation examined frequency differences between and within the ICE corpora and the colonies of the ICE corpora. The qualitative investigation examined meaning in terms of co-selection components of the lexical item. While the co-selection components and methodology originally proposed by Sinclair (2004a) were utilised as a starting point I have introduced additional methods and adapted the co-selection components. Whereas Sinclair investigated the second-order data of concordance lines, this research investigates third-order data, specifically *come*- and *go*-grams and their associated statistical information. I have dispensed with the core and I have introduced two new co-selection components – structural preference and discourse preference. I have also substituted the term semantic force for semantic prosody as I believe semantic prosody now has so much baggage attached to it that its use is problematic.

The most important conclusion reached from the quantitative investigation is that the difference in frequency is much greater within the ICE corpora than between the ICE corpora. This would suggest that World Englishes are much more homogeneous at a core level than might be supposed by the amount of research that is undertaken into the differences (e.g. Schneider 2007), and that much more research should be undertaken into the differences between discourses particularly the differences between spoken and written. The quantitative investigation also showed that the frequencies of the *come*- and *go*-grams of the scripted colonies were much more homogenous with written colonies than the spoken colonies. This again would suggest that care should be taken if spoken scripted language is used to examine spoken language as it would appear to be much more like written language.

As a result of the qualitative investigation, I argued that the collocates that have been identified could be an indication of new nodes for investigation, could be an indication of semantic preference, could be an indication of prospection, and could be a preliminary indication of the node influencing the co-selection components detected. I demonstrated that confining colligation to word classes has a benefit. I have also shown that, for *come* and *go* at least, functional words would appear to be more likely to be part of the node, and lexical words further colligates. I have also demonstrated that there is benefit in having structural preference as an additional co-selection component as it both allows less traditional structures to be included and separates word classes from structures. I have advocated that semantic preference should be seen as a broad category, encompassing evaluation, positive and negative, collocation, and semantic sets. I also have suggested that it is an undervalued component of the lexical item as it appears to be an important pre-cursor to the identification of the semantic force. However, I have also pointed out that where it was absent the semantic force was more likely to be related to text organisation, not the message or interaction.

I have included discourse preference as a co-selection component as it became increasingly apparent that some lexical items were predominately, if not exclusively, limited to specific types of discourse such as spoken language. However, while there were differences associated with discourse types, the predominating factor was homogeneity. While the use of *come*- and *go*-grams in live radio sports reporting was confined to this discourse it had spread through the Englishes and had spread to all types of sports. It was a local innovation; it is now sedimented across English as a whole.

I have established that semantic force can be sub-divided: the semantic force of message; the semantic force of interaction; and the semantic force of text organisation. This, I would suggest, could allow for the integration of Linear Unit Grammar (LUG) and lexical grammar. I will discuss this further below. I have demonstrated that semantic force appears layered. The choice of node impacts on the semantic force; as the node changes, the semantic forces of a word or words contained in the node can add together or they can layer. In changing the node, the other co-selection components also change.

I have proposed a new definition of the lexical item that has its roots in the syntagmatic manifestation of the paradigmatic choice. The lexical item consists of seven co-selection components, two that are obligatory and five that are optional. The obligatory co-selection components are the node and the semantic force, and the optional co-selection components are collocation, colligation, structural preference, semantic preference and discourse preference.

9.2 Research question(s)

This research asked where lexical items are located in World Englishes and discourses. This was further divided into three sub-questions asking where do lexical items begin and end, are they restricted to specific discourses and are they restricted to specific World Englishes. In reverse order this research has shown that lexical items can be restricted to specific World Englishes and they can also be restricted to specific discourses. For example, *go* as a replacement speech verb would appear to be a feature of the spoken colonies of ICE-Canada in this data, and *come* and *go* in live radio sports reporting have distinctive co-selection components. I have introduced the additional co-selection component of discourse preference to address this point.

The answer to the last sub-question is, I would suggest, the key to the conclusion of this research. Lexical items do not have a beginning or an end as they cannot be considered to be objects in the discourse. They are located in discourses as syntagmatic realisations of paradigmatic choices. I would go even further and say that I am becoming more and more convinced that linguists are making a mistake in trying to locate syntagmatic objects in the discourse in that whatever paradigmatic chunk, or node, one chooses will always have the potential to have distinctive syntagmatic, or co-selection components, realisations.

9.3 Empirical problem

This research has its roots firmly embedded in the work of Sinclair, starting with his methodologies and structures and then adapting them to accommodate the data under investigation. I would suggest that its main weakness is that which has also been directed at Sinclair's research: the corpus linguistic paradigm is an empirical paradigm and this research is predominately qualitative (see McEnery and Hardie 2012). I became increasingly aware that the more abstract the co-selection component, the more it was a feature of order rather than of sequence, the more it became my interpretation of the data. However, I do think that there are two points that can be made in mitigation. The first I have made before and that is that if one is going to use co-selection components to compare *n*-grams across discourses and World Englishes then how one interprets the data is secondary, all one requires is for the same or different co-selection components to be in evidence.

My second point is that having established that lexical items are products of paradigmatic choices on the syntagmatic axes, researchers can begin to look at the co-selection components in terms of algorithms and other recoverable procedures. This has not been necessarily possible before as the evidence of the existence of the lexical item on the syntagmatic axes has entailed a certain amount of variability. As a starting point, I would envisage the use of POS and semantic tagging for the identification of colligation and semantic preference. While semantic force cannot necessarily be quantified, it could conceivably be possible to quantify collocation, colligation, structural preference, discourse preference and semantic preference. This in turn could create the opportunity to examine electronically meaning variation across discourses rather than just frequency variation.

9.4 Further research

While it might be possible to identify co-selection components electronically, what is certainly possible is further research with different n-*grams* and different ICE corpora or other corpora. The aim of this type of research would be to corroborate this research and, possibly, further fine-tune the co-selection components. This research

has only investigated the *come*- and *go*-grams in four of the ICE corpora, as more ICE corpora come on line it will be possible to examine the co-selection components in them as well. I would suggest that this type of research need not necessarily examine all the *come*- and *go*-grams tackled in this research, but a selection of them. It would also be of interest to see if these new *come*- and *go*-grams exhibit the same type of profile of differences within the ICE corpora as shown in the box plots in Chapter 5.

It would also be of interest to see if this type of profile occurs with other *n*-grams associated with different high frequency words, both verbs and other word classes. The same type of research could be undertaken with medium to low frequency words, but this would require the compilation of larger parallel corpora. Additionally, research into the co-selection components of these different *n*-grams in firstly the four ICE corpora used in this research and then other ICE corpora and corpora would also be of benefit. Again, I would suggest that only a selection of *n*-grams associated with a particular high frequency word need be investigated. As the polysemous nature of high frequency words make them more difficult to study, I would suggest that this research paradigm goes some way to mitigating this problem. I also think that it would be useful to undertake research into *and*. I am convinced that it does far more than just link like to like – word to word, phrase to phrase, clause to clause –, but as it is so ubiquitous we see no reason to pay it the attention I think it should deserve. For example, *come and* could almost be considered to be a form of phrasal verb as it has distinct co-selection components such as a semantic force of **movement to undertake an action that will happen after the utterance with emphasis on the participant(s) being physically present in a place specific to the participants** when there is a colligation of **post verb**, and a structural preference of **surplus.**

This research has identified a deictic shift in live radio sports reporting. I would suggest that this is an area that would benefit from further investigation. Research could show if this type of deictic shift is confined to just sports and radio or is it something that can also be seen in other live broadcasts where events are being described to an audience. I would also be inclined to investigate how deictic centres are managed in documentaries and news broadcasts. Television and radio are new-comers as language mediums and, as such, are there co-selection components that can be associated with only them, and are they present in different World Englishes?

In addition this research has shown that lexical items can occur in tandem, and that they can be layered. This would suggest, from a pedagogic perspective, that phrasal verbs should be learnt in groups associated with the particular verb from which they originate, rather than just as a list of phrasal verbs. It would be useful if further research could be undertaken to see if this is the better approach for the language learner.

9.5 Links

I believe that it is important that one considers the possible links between the various investigative and theoretical strands of corpus linguistics as an integrated theoretical approach to meaning would be a positive step in the search to understand meaning creation in language. They are lexical bundles, concgrams, pattern grammar, and LUG. I would suggest that both lexical bundles and concgrams can be re-defined as a form of pre-set collocations with the associated co-selection components. Pattern grammar can be linked to structural preference. Finally, I discuss further my proposal that the elements of the discourse identified with LUG can be equated with the different types of semantic force (Section 6.1).

Lexical bundles are *n*-grams that occur at or above a specific frequency in a specific minimum of texts. While it has been suggested that analysis of any lexical

bundle can be undertaken, the majority of the analysis has been undertaken where $n \geq 4$ (Biber and Barbieri 2007, Biber 2009, Cortes 2004, Breeze 2013). The analysis has shown that they regularly straddle structural boundaries, and they have a basic communicative function (Biber 2009). The straddling of structural boundaries could be considered to be a structural preference, while the basic communicative function is the semantic force. I am concerned that the research on lexical bundles has been confined to those where $n \geq 4$ as this research would suggest that each value of n has the potential to have co-selection components.

A concgram is a set of words that occur together in any combination of constituent or position. They are identified by frequency criteria, the canonical (most frequent) form is determined and from this a meaning shift unit (MSU) is posited (Cheng *et al* 2009). In the sense the research paradigm associated with concgrams seek to detect objects, or MSUs, in the discourse, there are no links with this research. However, the concgrams are nothing more than non-contiguous and non-positional pre-set collocates and as this research proposes that it is the paradigmatic choices that dictate the syntagmatic co-selection components this must also apply to concgrams.

Patterns, in pattern grammar, are structures that occur with a restricted set of words. These words can also occur with a restricted set of patterns (Hunston and Francis 2000: 3). The words that are associated with a particular pattern habitually have some aspect of meaning in common, and different meanings of polysemous words are usually associated with different patterns (*ibid*). I would suggest that the patterns can be also classified as structural preferences. The word or words associated with the particular structural preference would be the collocations, and these could then be, as is done with pattern grammar, sub-divided into groups with specific semantic preferences. The colligations would be the word class to which these groups belong. The semantic force, the meanings associated with the patterns and the other co-selection components and, finally, the discourse preferences, the preference a particular pattern has for a particular discourse. Interestingly, I would suggest that the node – the word or words used to generate the concordances to examine the associated co-selection components – would be the structural preference. Not only can the node consist of collocations but it can also comprise structural preference, which would suggest that it would be possible to utilise other co-selection components as nodes. In this respect it would be easy to generate concordances utilising colligations as nodes, but semantic preference, discourse preference and semantic force would be much more difficult with current software capabilities. I would suggest that this might be something to aim for in the future, although as we are still awaiting automatic spoken language transcription this might take a little time.

The combination of LUG and lexical items is where I believe further research could have the potential to disclose how language means. LUG is both syntagmatic in its orientation and it can be used to model any type of language. It is particularly suitable for modelling spoken language as it delineates the interactive from the message. When the text is initially chunked the chunking is at the discretion of the "chunker" and it pre-suppose that to any user of language a text falls into smallish chunks and that each user has different perceptions as to where a chunk might begin or end (Sinclair and Mauranen 2006: xx, 6). This research has determined that it is the paradigmatic choice that influences the syntagmatic selections. The initial chunking of the language could be considered to be a succession of paradigmatic choices, each choice having the potential to have co-selection components that are identifiable along the syntagmatic axis. It is the investigation into the potential for the chunks to be lexical items that I believe could be revealing.

Once the text has been chunked it is then re-classified into three different types of elements. These are interactive organisational (OI), text-oriented (OT) and message-oriented (M). This research has suggested that the elements identified with LUG are also the same categories associated with semantic force, the relaying of a message, the organisation of an interaction and the organisation of a text (Section

8.1.6). It seems to me that the links are there and now we need to find some way to model language, especially spoken language, along the syntagmatic axis. Too long have we over-estimated the paradigmatic and under-estimated the syntagmatic, and too long have we ignored spoken language. The combination of lexical items as envisaged by this research and LUG might just be the answer.

References

Adolphs, S. (2008) *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse.* Amsterdam and Philadelphia: John Benjamins.

Austin, J. (1975) *How to do Things with Words* (2nd Edn.). Oxford: Oxford University Press.

Bakhtin, M. (1981) *The Dialogic Imagination: Four Essays.* Emerson, C. and Holquist, M. (Trans.). Austin: University of Texas Press.

Barlow, M. (2011) 'Corpus linguistics and theoretical linguistics.' *International Journal of Corpus Linguistics* 16 (1): 3-44.

Barnbrook, G., Mason, O. and Krishnamurthy, R. (2013) *Collocation: Applications and Implications.* UK and USA: Palgrave Macmillan.

Behre, F. (1973) *Get, Come and Go: Some Aspects of Situational Grammar.* Stockholm: Almqvist and Wiksell.

Bednarek, M. (2008) 'Semantic preference and semantic prosody re-examined.' *Corpus Linguistics and Linguistic Theory* 4 (2): 119-139.

Biber, D. (2006) *University Language: A Corpus-based Study of Spoken and Written Registers.* Amsterdam and Philadelphia: John Benjamins.

Biber, D. (2009) 'A corpus-driven approach to formulaic language in English.' *International Journal of Corpus Linguistics* 14(3): 275-311.

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

Biber, D. and Conrad, S. (2004) 'Corpus-based comparisons of registers.' In Coffin, C., Hewings, A. and O'Halloran, K., *Applying English Grammar: Functional and Corpus Approaches*, Great Britain: Hodder and Arnold, 40-56.

Biber, D, Conrad, S., and Cortes, V. (2004) 'If you look at Lexical Bundles in University Teaching and Textbooks.' *Applied Linguistics* 25 (3): 371-405.

Biber, D. and Barbieri, F. (2007) 'Lexical bundles in university spoken and written registers.' *English for Specific Purposes* 26: 263-286.

Bolinger, D. (1971) *The Phrasal Verb in English.* USA: Harvard University Press.

Bolton, K. (2009) 'World Englishes today.' In Kachru, B., Kachru, Y. and Nelson, C. *The Handbook of World Englishes*, UK: Wiley-Blackwell, 240-269.

Brazil, D. (1995) *A Grammar of Speech.* Oxford: Oxford University Press.

Breeze, R. (2013) 'Lexical bundles across four legal genres.' *International Journal of Corpus Linguistics* 18 (2): 229-253.

Carroll, L. (2003) *Alice's Adventures in Wonderland and Through the Looking Glass and What Alice Found There.* New York and London: Bloomsbury.

Carter, R. (2004a) *Language and Creativity: the Art of Common Talk*.. London and New York: Routledge.

Carter, R. (2004b) 'Grammar of spoken English.' In Coffin, C., Hewings, A. and O'Halloran, K., *Applying English Grammar: Functional and Corpus Approaches*, Great Britain: Hodder and Arnold, 25-39.

Carter, R. and McCarthy, M. (2006) *Cambridge Grammar of English*. Cambridge: Cambridge University Press.

Cermak, F. (2001) 'Substance of idioms: perennial problems, lack of data or theory?' *International Journal of Lexicography* 14 (1): 1-20.

Cermak, F. (2009) 'Spoken corpora design.' *International Journal of Corpus Linguistics* 14 (1): 113-123.

Chafe, W. (2006) 'Reading Aloud'. In Hughes, R. (ed.), *Spoken English, TESOL and Applied Linguistics: Challenges for Theory and Practice*, UK: Palgrave MacMillan, 53-70.

Cheng, W., Greaves C. and Warren M. (2006) 'From n-gram to skipgram to concgram.' *International Journal of Corpus Linguistics* 11 (4): 411-433.

Cheng, W., Greaves, C., Sinclair, J. and Warren, M. (2009) 'Uncovering the extent of the phraseological tendency: towards a systematic analysis of Concgrams.' *Applied Linguistics* 30 (2): 236-252.

Cortes, V. (2004) 'Lexical bundles in published and student disciplinary writing: Examples from history and biology.' *English for Specific Purposes* 23: 397-423.

Crystal, D. (2003) *English as a Global Language* (2nd Edn.). Cambridge: Cambridge University Press.

Darwin, C. and Gray, L. (1999) 'Going after the phrasal verb: an alternative approach to classification.' *TESOL Quarterly* 33(1): 65-83.

Duchan, J., Bruder, G. and Hewitt, L. (Eds.) (1995/2009) *Deixis in Narrative: A Cognitive Science Perspective.* New York and Oxford: Routledge.

Ellis, N., Frey, E. and Jalkanen, I. (2009) 'The psycholinguistic reality of collocation and semantic prosody (1): lexical access.' In Römer, U. and Schulze, R. (Eds.) *Exploring the Lexis-Grammar Interface*, Amsterdam and Philadelphia: John Benjamins, 89-114.

Fillmore, C. (1966) 'Categories in the semantics of 'come'.' *Foundations of Language* 2 (3): 219-227.

Fillmore, C. (1971/5) *Santa Cruz Lectures on Deixis 1971*. Indiana: Indiana University Linguistics Club.

Firth, J. (1957) *Papers in Linguistics; 1934-1951*. London: Oxford University Press.

Firth, J. (1968) *Selected Paper of J.R. Firth*. Palmer, F.R. (Ed.) United Kingdom: Longmans, Green and Co Ltd.

Freadman, A. (2012) 'The trap and trappings of genre theory.' *Applied Linguistics* 33 (5): 544-563.

Gardner, D. and Davies, M. (2007) 'Pointing out frequent phrasal verbs: a corpus-based analysis.' *TESOL Quarterly* 41 (2): 339-359.

Gilquin, G. & Gries, S. (2009) 'Corpora and experimental methods: A state-of-the-art review.' *Corpus Linguistics and Linguistic Theory* 5 (1): 1-26.

Gray, B. and Biber, D. (2013) 'Lexical frames in academic prose and conversation.' *International Journal of Corpus Linguistics* 18 (1): 109-135.

Green, K. (Ed.) (1995) *New Essays in Deixis: Discourse, Narrative, Literature.* Amsterdam and Atlanta: Rodopi.

Greenbaum, S. (1991) 'The development of the International Corpus of English'. In Aijmer, K. and Altenberg, B. (Eds.), *English Corpus Linguistics,* London and New York: Longman, 83-91.

Gries, S. (2013) '50-something years of work on collocations: what is or should be next…' *International Journal of Corpus Linguistics* 18 (1): 137-165.

Halliday, M. (1985) *Spoken and Written Language.* Australia: Deakin University Press.

Hoey, M. (2001) *Textual Interaction: an introduction to written discourse analysis.* UK: Routledge.

Hoey, M. (2005) *Lexical Priming*. UK: Routledge.

Howitt, D. and Cramer, D. (2005) *Introduction to Statistics in Psychology* (3rd Edn.). England: Pearson Education Limited.

Hughes, R. (1996) *English in Speech and Writing: Investigating Language and Literature*. London and New York: Routledge.

Hughes, R. (ed.) (2006) *Spoken English, TESOL and Applied Linguistics: Challenges for Theory and Practice.* UK: Palgrave MacMillan.

Hunston, S. (2007) 'Semantic prosody revisited.' *International Journal of Corpus Linguistics* 12(2), 249-268.

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

Hunston, S. and Thompson, G. (Eds.) (2000) *Evaluation in Text.* Oxford: Oxford University Press.

Jespersen, O. (1924) *The Philosophy of Grammar.* Chicago and London: The University of Chicago Press.

Kachru, B. (1992) 'Teaching World Englishes'. In B. Kachru (Ed.), *The Other Tongue: English across Cultures,*(2nd Edn.), Urbana and Chicago: University of Illinois Press, 255-265.

Kachru, Y. (2008) 'Language variation and corpus linguistics.' *World Englishes* 27 (1): 1-8.

Kilgarriff, A. (2002) 'Comparing corpora.' *International Journal of Corpus Linguistics.* 6 (1): 97-133.

Leech, G. (1989) *An A-Z of English Grammar.* England: Longman.

Leech, G. (2000) 'Grammars of Spoken English: New Outcomes of Corpus-Oriented Research.' *Language and Learning* 50 (4): 675-724.

Leech, G., Rayson, P., and Wilson, A. (2001) *Word frequencies in Written and Spoken English: Based on the British National Corpus.* UK: Longman.

Louw, B. (1993) 'Irony in the text or insincerity in the writer? – the diagnostic potential of semantic prosodies. In Sampson, G. and McCarthy, D. (Eds.) (2005) *Corpus Linguistics: Reading in a Widening Discipline*, London and New York: Continuum, 229-241.

Louw, B. (2000) 'Contextual prosodic theory: bringing semantic prosodies to life'. In Heffer, C.,Saunstson, H. and Fox, G. (Eds.) *Words in Context, A tribute to John Sinclair on his Retirement,* Birmingham: University of Birmingham. Downloaded from http://www.revue-texto.net/docannexe/file/124/louw_prosodie.pdf, last accessed 15 August 2011.

McCarthy, M. (2001) *Issues in Applied Linguistics.* Cambridge: Cambridge University Press.

McCarthy, M. and Carter, R.(2006) '*This, that* and *the other*: multi-word clusters in spoken English as visible patterns of interaction.' In McCarthy, M., *Explorations in Corpus Linguistics*, Cambridge: Cambridge University Press, 7-26.

MacIntyre, D. (2007) 'Deixis, cognition and the construction of viewpoint.' In Lambrou, M., and Stockwell, P. (Eds.) *Contemporary Stylistics. Contemporary Studies in Linguistics,* Continuum, 118-130.

Mahlberg, M. (2005) *English General Nouns: A Corpus Theoretical Approach.* Amsterdam: John Benjamins.

Malinowski, B. (1923) 'The problem of meaning in primitive language'. In Ogden, C. and Richards, I. (Eds.), *The Meaning of Meaning*, New York: Harcourt, Brace and World, 296-355.

McEnery, T. and Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice.* [Kindle Version]. Cambridge: Cambridge University Press.

Milne, A.A. (2001) *Winnie-The-Pooh: the Complete Collection of Stories and Poems.* UK: Egmont.

Morley, J. and Partington, A. (2009) 'A few *Frequently Asked Questions* about semantic – or *evaluative* – prosody.' *International Journal of Corpus Linguistics* 14 (2): 139-158.

Mukherjee, J. and Gries, S. (2010) 'Lexical gravity across varieties of English: an ICE-based study of n-grams in Asian Englishes.' *International Journal of Corpus Linguistics* 15 (4): 520-548.

Nesi, H. & Basturkmen, H. (2006) 'Lexical bundles and discourse signalling in academic lectures.' *International Journal of Corpus Linguistics* 11 (3): 283-304.

Nelson, G. (2004) 'Introduction' *World Englishes* 23 (2): 225-226.

Nelson, G. (2009) 'World Englishes and corpora studies.' In Kachru, B., Kachru, Y. and Nelson, C. *The Handbook of World Englishes,* UK: Wiley-Blackwell, 733-750.

O'Keeffe, A., McCarthy, M. and Carter, R. (2007) *From Corpus to Classroom: Language Use and Language Teaching.* Cambridge: Cambridge University Press.

Partington, A., Duguid, A. and Taylor, C. (2013) *Patterns and Meanings in Discourse: Theory and Practice in Corpus Assisted Discourse Studies (CADS).* Amsterdam and Philadelphia: John Benjamins.

Pennycook, A. (2010) *Language as a Local Practice.* UK: Routledge.

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

Römer, U. (2009) 'The inseparability of lexis and grammar: corpus linguistic perspectives.' *Annual Review of Cognitive Linguistics* 7: 141-163.

Rühlemann, C. (2006) 'Coming to terms with conversational grammar: 'dislocation' and 'dysfluency'. *International Journal of Corpus Linguistics* 11 (4): 385-409.

Saussure, F. De (2013) *Course in General Linguistics.* Harris, R. (Ed. and Trans.). London: Bloomsbury.

Scott, M. (2015) *WordSmith Tools ver.6.* Stroud: Lexical Analysis Software.

Scott, M. and Tribble, C. (2006) *Textual Patterns: Key Words and Corpus Analysis in Language Education.* Amsterdam and Philadelphia: John Benjamins.

Searle, J. (2010) *Making of the Social World.* Oxford: Oxford University Press.

Segal, E. (1995/2009) 'Narrative comprehension and the role of Deictic Shift Theory.' In Duchan, J., Bruder, G. and Hewitt, L. (Eds.), *Deixis in Narrative: A Cognitive Science Perspective,* New York and Oxford: Routledge, 3-17.

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

Sinclair, J.. (1999) 'A way with common words'. In Hasselgard, H. & Okesefjell, S. (Eds.), *Out of Corpora,* Amsterdam: Rodopi, 157-179.

Sinclair, J. (2003) *Reading Concordances: An Introduction.* Great Britain: Pearson Education Limited.

Sinclair, J. (2004a) *Trust the Text*. London and New York: Routledge.

Sinclair, J. (2004b) 'Meaning in the framework of corpus linguistics.' *Lexicographica* 20: 20-32.

Sinclair, J., Jones, S. and Daley, R. (1970/2004) *English Collocation Studies: The OSTI Report.* Krishnamurthy, R. (Ed.). London and New York: Continuum.

Sinclair, J. and Mauranen, A. (2006) *Linear Unit Grammar: Integrating speech and writing.* Amsterdam and Philadelphia: John Benjamins.

Stockwell, P. (2002) *Cognitive Poetics: An Introduction.* London and New York: Routledge.

Stubbs, M. (2002) *Words and Phrases: Corpus Studies of Lexical Semantics.* UK: Blackwell Publishing Ltd.

Stubbs, M. (2006) 'Corpus analysis: the state of the art and three types of unanswered questions.' In Thompson, G. and Huston, S. (Eds.) *System and Corpus: Exploring Connections,* London and Oakville: Equinox, 15-36.

Stubbs, M. (2007a) 'Collocations and semantic profiles: on the cause of the trouble with quantitative studies.' In Teubert, W. and Krishnamurthy, R. (Eds.), *Corpus Linguistics: Critical Concepts in Linguistics,* London & New York: Routledge, 166-193.

Stubbs, M. (2007b) 'On texts, corpora and models of language'. In Hoey, M., Mahlberg, M., Stubbs, M. and Teubert, W. *Text, Discourse and Corpora: Theory and Analysis,* New York and London: Continuum, 127-161

Stubbs, M. (2009a) 'Memorial Article: John Sinclair (1933-2007). The search for unit of meaning: Sinclair on Empirical Semantics.' *Applied Linguistics* 30 (1): 115-137.

Stubbs, M. (2009b) 'Technology and phraseology: with notes on the history of Corpus Linguistics.' In Römer, U. and Schulze, R. (Eds.), *Exploring the Lexis-Grammar Interface,* Amsterdam and Philadelphia: John Benjamins, 15-32.

Stubbs, M. (2013) 'Sequence and order: the neo-Firthian tradition of corpus semantics.' In Hasselgård, H., Ebeling, J. and Oksefjell Ebeling, S. (Eds.), *Corpus Perspectives on Patterns of Lexis,* Amsterdam and Philadelphia: John Benjamins, 13–34.

Sweet, H. (1925) *A New English Grammar.* Oxford: Clarendon Press.

Teubert, W. (2005) 'My version of corpus linguistics.' *International Journal of Corpus Linguistics* 10 (1): 1-13.

Teubert, W. (2010) *Meaning, Discourse and Society.* Cambridge: Cambridge University Press.

Teubert, W. and Čermáková, A. (2004) 'Directions in Corpus Linguistics.' In Halliday, M., Teubert, W. Yallop, C. and Čermáková A. (Eds.), *Lexicology and Corpus Linguistics: an Introduction,* 112-165.

Thompson, G. (2004) *Introducing Functional Grammar* (2nd edn.). Great Britian: Arnold.

Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work.* Amsterdam and Philadelphia: John Benjamins.

Tsur, R. (2008) 'Deixis in literature: what isn't cognitive poetics?' *Pragmatics and Cognition* 16 (1): 119-150.

Whitsitt, S. (2007) 'A critique of the concept of semantic prosody.' *International Journal of Corpus Linguistics* 10 (3): 283-305.

Widdowson, H. (2003) *Defining Issues in English Language Teaching.* Oxford University Press: Oxford.

Widdowson, H. (2004) *Text, Context, Pretext: Critical Issues in Discourse Analysis.* UK: Blackwell Publishing.

Wilks, Y. (2005) "What would a Wittgensteinian Computational Linguistics be like?" *Proceedings 10th International Congress Pragmatics, 212-227.*

Willis, D. (1990) *The Lexical Syllabus.* Great Britain: William Collins Sons & Co. Ltd.

Wittgenstein, L. (2009) *Philosophical Investigations* (4th Edn.). Hacker, P and Schulte, J. (Eds.). Anscombe, G., Hacker, P. and Schulte, J. (Trans.). UK: Wiley-Blackwell.

Worlock Pope, C. (Ed.) (2010) 'The Bootcamp Discourse and Beyond.' Special issue of the *International Journal of Corpus Linguistics* 15 (3).

Wray, A. (2008) *Formulaic Language: Pushing the Boundaries.* Oxford: Oxford University Press.

Xiao, R. and McEnery, T. (2006) 'Collocation, semantic prosody, and near synonymy: a cross-linguistic perspective.' *Applied Linguistics* 27 (1): 103-129.

Zubin, D. and Hewitt, L. (1995/2009) 'The Deictic Center: a theory of deixis in narrative.' In Duchan, J., Bruder, G. and Hewitt, L. (Eds.) *Deixis in Narrative: A Cognitive Science Perspective,* New York and Oxford: Routledge, 129-155.

Appendix I: ***Frequency/million words of come and come-grams.***

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**Extra**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
		COME			1	Canada		4	1000	524	1054	902	587	1330	823	427
		COME			1	GB		4	1087	528	1359	1140	755	909	676	480
		COME			1	India		4	1582	614	2543	1363	574	1422	946	510
		COME			1	Jamaica		4	1726	617	2121	1936	650	1696	934	509
	THEY	COME			2	Canada		4	47	12	15	48	10	118	20	10
	THEY	COME			2	GB		4	56	0	36	42	28	119	0	0
	THEY	COME			2	India		4	51	5	89	36	17	44	21	0
	THEY	COME			2	Jamaica		4	110	7	121	119	0	165	0	10
	TO	COME			2	Canada		4	257	119	270	242	186	306	238	80
	TO	COME			2	GB		4	226	162	347	198	168	139	284	123
	TO	COME			2	India		4	270	122	400	282	113	227	350	51
	TO	COME			2	Jamaica		4	328	131	349	363	191	357	292	76
	YOU	COME			2	Canada		4	47	7	60	48	0	63	10	6
	YOU	COME			2	GB		4	50	14	82	60	19	20	49	3
	YOU	COME			2	India		4	115	24	235	48	17	107	51	16
	YOU	COME			2	Jamaica		4	96	15	97	131	48	89	19	13
		COME	AND		2	Canada		4	62	12	70	54	10	97	30	6
		COME	AND		2	GB		4	52	29	72	54	9	53	88	9
		COME	AND		2	India		4	79	17	127	66	9	88	21	16
		COME	AND		2	Jamaica		4	131	30	189	113	48	130	58	20
		COME	BACK		2	Canada		4	100	15	115	54	59	160	30	10
		COME	BACK		2	GB		4	118	31	158	156	28	86	20	35
		COME	BACK		2	India		4	115	37	212	72	0	126	82	22
		COME	BACK		2	Jamaica		4	171	20	228	191	19	178	29	17
		COME	FROM		2	Canada		4	54	29	30	85	59	49	50	22
		COME	FROM		2	GB		4	76	36	61	84	112	60	20	41

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**Extra**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
		COME	FROM		2	India		4	72	27	113	78	9	63	10	32
		COME	FROM		2	Jamaica		4	96	69	155	89	38	62	97	60
		COME	IN		2	Canada		4	98	24	70	73	29	216	30	22
		COME	IN		2	GB		4	64	24	72	72	0	93	29	22
		COME	IN		2	India		4	72	17	75	66	78	76	21	16
		COME	IN		2	Jamaica		4	110	25	160	113	10	110	29	23
		COME	ON		2	Canada		4	59	34	125	30	20	28	10	42
		COME	ON		2	GB		4	61	10	128	42	9	33	10	9
		COME	ON		2	India		4	87	22	179	66	9	50	31	19
		COME	ON		2	Jamaica		4	59	10	82	72	10	48	10	10
		COME	OUT		2	Canada		4	56	29	65	48	39	63	30	29
		COME	OUT		2	GB		4	87	14	102	138	9	66	0	19
		COME	OUT		2	India		4	73	34	127	66	26	50	41	32
		COME	OUT		2	Jamaica		4	109	17	136	143	29	89	0	23
		COME	TO		2	Canada		4	98	124	50	115	88	153	248	83
		COME	TO		2	GB		4	145	108	118	150	196	139	196	79
		COME	TO		2	India		4	343	149	452	300	165	409	329	93
		COME	TO		2	Jamaica		4	229	99	213	280	201	213	165	76
		COME	UP		2	Canada		4	95	19	90	91	78	118	40	13
		COME	UP		2	GB		4	61	14	66	78	47	46	20	13
		COME	UP		2	India		4	61	24	61	90	52	44	10	29
		COME	UP		2	Jamaica		4	104	30	131	125	38	89	39	26
	WILL	COME			2	Canada	extra	3	11	39	5	6	29	14	99	19
	WILL	COME			2	GB		3	40	29	26	54	37	46	39	25
	WILL	COME			2	India		3	99	29	151	132	26	57	72	16
	WILL	COME			2	Jamaica		3	50	25	53	30	38	76	49	17
		COME	DOWN		2	Canada		3	46	19	60	24	10	77	10	22

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**Extra**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
		COME	DOWN		2	GB		3	44	12	92	12	9	40	0	16
		COME	DOWN		2	India		3	42	10	52	48	35	31	10	10
		COME	DOWN		2	Jamaica	extra	3	16	5	10	30	19	7	10	3
	HAS	COME			2	Canada	extra	2	10	27	0	6	29	14	69	13
	HAS	COME			2	GB	extra	2	27	24	5	6	56	60	29	22
	HAS	COME			2	India		2	121	27	104	174	96	120	31	26
	HAS	COME			2	Jamaica		2	43	2	10	36	48	96	0	3
	HAVE	COME			2	Canada	extra	2	23	27	25	24	29	14	40	22
	HAVE	COME			2	GB	extra	2	24	12	36	12	19	27	0	16
	HAVE	COME			2	India		2	94	46	146	48	61	107	31	51
	HAVE	COME			2	Jamaica		2	38	32	24	54	57	27	68	20
	I	COME			2	Canada	extra	2	8	2	10	18	0	0	0	3
	I	COME			2	GB	extra	2	24	7	66	6	0	7	10	6
	I	COME			2	India		2	42	17	89	24	9	25	51	6
	I	COME			2	Jamaica		2	62	17	126	36	0	48	39	10
		COME	HERE		2	Canada	extra	2	28	10	55	18	29	0	10	10
		COME	HERE		2	GB	extra	2	13	2	20	12	9	7	0	3
		COME	HERE		2	India		2	42	22	61	18	9	69	31	19
		COME	HERE		2	Jamaica		2	59	10	87	48	10	69	10	10
		COME	INTO		2	Canada	extra	2	23	10	10	24	0	56	10	10
		COME	INTO		2	GB	extra	2	35	24	26	66	28	20	0	32
		COME	INTO		2	India		2	43	12	19	48	70	57	10	13
		COME	INTO		2	Jamaica		2	45	10	34	66	29	48	10	10
	AND	COME			2	Canada	extra	1	11	2	10	12	10	14	10	0
	AND	COME			2	GB	extra	1	19	7	15	36	9	13	0	9
	AND	COME			2	India	extra	1	28	15	61	12	0	25	51	3
	AND	COME			2	Jamaica		1	46	10	63	60	10	34	29	3
									26	**9**	**37**	**30**	**7**	**22**	**23**	**4**

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**Extra**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
	HAD	COME			2	Canada	extra	1	10	27	20	0	10	7	40	22
	HAD	COME			2	GB	extra	1	18	31	10	24	19	20	10	38
	HAD	COME			2	India		1	46	32	80	42	17	31	10	38
	HAD	COME			2	Jamaica	extra	1	8	44	5	12	0	14	0	60
	NOT	COME			2	Canada	extra	1	5	7	5	6	0	7	10	6
	NOT	COME			2	GB	extra	1	16	22	10	18	19	20	20	22
	NOT	COME			2	India		1	34	22	47	42	17	25	21	22
	NOT	COME			2	Jamaica	extra	1	21	2	10	24	10	41	10	0
	WE	COME			2	Canada	extra	1	16	7	5	12	10	42	30	0
	WE	COME			2	GB	extra	1	10	0	5	30	0	0	0	0
	WE	COME			2	India		1	46	2	42	42	17	82	10	0
	WE	COME			2	Jamaica	extra	1	32	0	19	60	29	21	0	0
		COME	UP	WITH	3	Canada		2	57	12	25	61	68	91	20	10
		COME	UP	WITH	3	GB	extra	2	19	12	10	48	19	0	10	13
		COME	UP	WITH	3	India	extra	2	7	5	0	12	26	0	10	3
		COME	UP	WITH	3	Jamaica		2	42	15	48	48	38	27	0	20
	TO	COME	AND		3	Canada	extra	1	21	5	20	36	0	21	10	3
	TO	COME	AND		3	GB	extra	1	16	17	26	18	0	13	49	6
	TO	COME	AND		3	India	extra	1	16	5	24	12	0	25	10	3
	TO	COME	AND		3	Jamaica		1	35	12	24	30	38	55	39	3
	TO	COME	TO		3	Canada	extra	1	21	22	10	18	29	35	69	6
	TO	COME	TO		3	GB	extra	1	27	26	46	18	28	13	69	13
	TO	COME	TO		3	India		1	36	44	52	42	9	31	144	13
	TO	COME	TO		3	Jamaica	extra	1	30	7	44	42	0	21	29	0
	TO	COME	UP		3	Canada		1	38	2	15	24	59	70	0	3
	TO	COME	UP		3	GB	extra	1	11	2	20	6	9	7	0	3
	TO	COME	UP		3	India	extra	1	28	10	33	30	26	25	10	10
	TO	COME	UP		3	Jamaica	extra	1	30	0	29	42	19	27	0	0

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**Extra**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
	YOU	COME	TO		3	Canada	extra	1	8	0	0	18	0	14	0	0
	YOU	COME	TO		3	GB	extra	1	15	5	10	12	19	20	20	0
	YOU	COME	TO		3	India		1	40	2	57	48	0	44	10	0
	YOU	COME	TO		3	Jamaica	extra	1	24	0	10	48	29	14	0	0
		COME	BACK	TO	3	Canada	extra	1	18	5	10	12	10	42	10	3
		COME	BACK	TO	3	GB	extra	1	35	10	31	48	19	40	20	6
		COME	BACK	TO	3	India	extra	1	25	10	33	24	0	38	10	10
		COME	BACK	TO	3	Jamaica		1	38	10	58	30	10	41	10	10
		COME	OUT	OF	3	Canada	extra	1	11	7	0	18	10	21	10	6
		COME	OUT	OF	3	GB	extra	1	21	2	10	54	0	13	0	3
		COME	OUT	OF	3	India	extra	1	15	10	24	18	9	6	10	10
		COME	OUT	OF	3	Jamaica		1	40	7	44	36	19	55	0	10
		COME	TO	KNOW	3	Canada	extra	1	2	0	0	6	0	0	0	0
		COME	TO	KNOW	3	GB	extra	1	0	0	0	0	0	0	0	0
		COME	TO	KNOW	3	India		1	40	0	52	54	17	31	0	0
		COME	TO	KNOW	3	Jamaica	extra	1	2	0	0	6	0	0	0	0
		COME	TO	THE	3	Canada	extra	1	16	15	15	0	20	35	30	10
		COME	TO	THE	3	GB	extra	1	21	19	26	18	9	27	49	9
		COME	TO	THE	3	India		1	57	12	28	84	17	101	10	13
		COME	TO	THE	3	Jamaica	extra	1	32	10	10	54	67	14	39	0

Appendix II: ***Frequency/million words of go and go-grams***

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**extr a**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
		GO			1	Canada		4	2356	665	3298	1919	656	2758	972	565
		GO			1	GB		4	1606	538	2626	1314	699	1247	774	461
		GO			1	India		4	1619	609	3192	1122	426	1070	936	507
		GO			1	Jamaica		4	2329	595	3612	2163	583	1957	827	516
	AND	GO			2	Canada		4	75	27	100	67	20	91	60	16
	AND	GO			2	GB		4	39	10	31	36	28	60	10	9
	AND	GO			2	India		4	36	22	94	12	9	6	31	19
	AND	GO			2	Jamaica		4	66	30	140	30	0	48	29	30
	I	GO			2	Canada		4	70	22	95	42	10	111	50	13
	I	GO			2	GB		4	50	17	97	36	19	27	69	0
	I	GO			2	India		4	72	10	179	24	26	19	10	10
	I	GO			2	Jamaica		4	110	17	232	42	38	69	49	7
	THEY	GO			2	Canada		4	74	10	105	18	10	139	0	13
	THEY	GO			2	GB		4	42	10	61	12	9	73	0	13
	THEY	GO			2	India		4	52	5	89	36	17	50	10	3
	THEY	GO			2	Jamaica		4	78	5	82	48	10	158	10	3
	TO	GO			2	Canada		4	661	211	904	557	264	724	387	154
	TO	GO			2	GB		4	529	201	848	420	205	464	324	161
	TO	GO			2	India		4	479	200	876	396	191	296	350	154
	TO	GO			2	Jamaica		4	666	200	1085	477	153	659	273	175
	WE	GO			2	Canada		4	65	10	115	67	29	21	30	3
	WE	GO			2	GB		4	47	7	82	24	0	60	10	6
	WE	GO			2	India		4	87	7	122	72	43	94	0	10
	WE	GO			2	Jamaica		4	74	15	58	72	57	110	0	20
	YOU	GO			2	Canada		4	249	15	315	182	39	383	10	16
	YOU	GO			2	GB		4	142	22	281	150	9	46	20	22

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**extr a**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
	YOU	GO			2	India		4	175	20	396	84	9	113	41	13
	YOU	GO			2	Jamaica		4	234	7	412	179	57	172	19	3
		GO	BACK		2	Canada		4	85	22	120	67	20	104	30	19
		GO	BACK		2	GB		4	60	22	56	72	9	86	39	16
		GO	BACK		2	India		4	88	32	132	66	43	94	10	38
		GO	BACK		2	Jamaica		4	133	37	213	137	10	103	78	23
		GO	IN		2	Canada		4	118	15	100	103	20	230	20	13
		GO	IN		2	GB		4	42	14	66	24	19	46	20	13
		GO	IN		2	India		4	46	20	85	48	9	25	41	13
		GO	IN		2	Jamaica		4	59	12	82	72	19	41	0	17
		GO	INTO		2	Canada		4	106	19	70	115	59	181	30	16
		GO	INTO		2	GB		4	229	29	572	84	37	80	0	38
		GO	INTO		2	India		4	45	17	19	60	35	76	10	19
		GO	INTO		2	Jamaica		4	107	17	140	77	29	151	10	20
		GO	ON		2	Canada		4	77	34	70	67	20	139	40	32
		GO	ON		2	GB		4	139	33	194	144	93	93	59	25
		GO	ON		2	India		4	99	20	160	84	78	57	41	13
		GO	ON		2	Jamaica		4	136	22	218	89	76	117	58	10
		GO	OUT		2	Canada		4	118	34	170	115	49	97	30	35
		GO	OUT		2	GB		4	77	26	169	42	47	20	39	22
		GO	OUT		2	India		4	42	27	104	24	9	6	31	26
		GO	OUT		2	Jamaica		4	86	10	131	77	19	82	39	0
		GO	THROUGH		2	Canada		4	95	22	100	79	10	167	40	16
		GO	THROUGH		2	GB		4	58	12	41	84	19	80	20	9
		GO	THROUGH		2	India		4	39	20	57	48	26	19	62	6
		GO	THROUGH		2	Jamaica		4	69	12	63	89	29	82	19	10
		GO	TO		2	Canada		4	380	133	625	315	157	272	248	96

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**extr a**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
		GO	TO		2	GB		4	237	112	450	162	84	153	206	82
		GO	TO		2	India		4	346	144	749	252	17	183	339	83
		GO	TO		2	Jamaica		4	450	128	906	310	76	233	136	126
HAVE	TO	GO			3	Canada		4	103	15	170	61	10	125	10	16
HAVE	TO	GO			3	GB		4	42	7	66	36	37	20	10	6
HAVE	TO	GO			3	India		4	79	12	179	66	9	19	0	16
HAVE	TO	GO			3	Jamaica		4	88	25	155	77	29	48	58	13
	TO	GO	TO		3	Canada		4	111	53	210	79	59	49	119	32
	TO	GO	TO		3	GB		4	94	50	184	54	19	73	118	28
	TO	GO	TO		3	India		4	111	49	226	96	17	50	134	22
	TO	GO	TO		3	Jamaica		4	152	57	334	107	19	41	29	66
		GO	TO	THE	3	Canada		4	93	34	170	73	20	63	60	26
		GO	TO	THE	3	GB		4	47	26	77	60	19	13	29	25
		GO	TO	THE	3	India		4	73	22	104	96	35	44	41	16
		GO	TO	THE	3	Jamaica		4	106	27	160	119	19	76	19	30
	CAN	GO			2	Canada		3	75	15	125	42	29	77	20	13
	CAN	GO			2	GB	extra	3	32	5	20	48	56	13	0	6
	CAN	GO			2	India		3	72	7	179	30	17	19	0	10
	CAN	GO			2	Jamaica		3	66	12	97	48	0	89	10	13
	JUST	GO			2	Canada		3	57	2	65	54	10	84	0	3
	JUST	GO			2	GB		3	39	0	66	30	0	40	0	0
	JUST	GO			2	India	extra	3	33	0	99	6	0	0	0	0
	JUST	GO			2	Jamaica		3	48	2	116	18	0	21	0	3
		GO	AND		2	Canada		3	141	24	220	61	20	209	20	26
		GO	AND		2	GB	extra	3	147	7	276	96	0	139	29	0
		GO	AND		2	India		3	118	29	245	90	0	76	31	29
		GO	AND		2	Jamaica		3	102	20	145	101	10	110	39	13

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**extr a**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
		GO	FOR		2	Canada		3	65	22	90	67	10	70	40	16
		GO	FOR		2	GB		3	35	10	61	48	9	7	0	13
		GO	FOR		2	India		3	93	20	207	42	0	69	21	19
		GO	FOR		2	Jamaica	extra	3	38	7	77	24	0	27	0	10
	TO	GO	AND		3	Canada	extra	3	23	7	25	18	10	35	10	6
	TO	GO	AND		3	GB		3	44	2	97	30	0	20	10	0
	TO	GO	AND		3	India		3	43	10	80	30	17	31	10	10
	TO	GO	AND		3	Jamaica		3	46	2	68	36	0	62	0	3
	YOU	GO	TO		3	Canada		3	44	5	70	18	0	70	0	6
	YOU	GO	TO		3	GB	extra	3	21	2	36	24	0	13	10	0
	YOU	GO	TO		3	India		3	51	5	118	18	0	38	10	3
	YOU	GO	TO		3	Jamaica		3	78	0	174	42	10	34	0	0
		GO	BACK	TO	3	Canada		3	39	22	70	30	20	21	30	19
		GO	BACK	TO	3	GB	extra	3	24	19	15	30	9	40	39	13
		GO	BACK	TO	3	India		3	42	22	66	42	35	19	10	26
		GO	BACK	TO	3	Jamaica		3	58	15	92	42	10	62	29	10
	WILL	GO			2	Canada	extra	2	29	24	25	24	20	49	50	16
	WILL	GO			2	GB	extra	2	31	19	10	36	47	40	10	22
	WILL	GO			2	India		2	115	27	245	96	26	38	21	29
	WILL	GO			2	Jamaica		2	54	12	44	60	48	69	39	3
		GO	DOWN		2	Canada		2	56	15	135	24	0	21	0	19
		GO	DOWN		2	GB	extra	2	23	5	41	30	9	0	20	0
		GO	DOWN		2	India	extra	2	10	7	5	12	0	25	0	10
		GO	DOWN		2	Jamaica		2	40	0	58	24	0	62	0	0
		GO	THERE		2	Canada		2	47	10	100	30	20	14	40	0
		GO	THERE		2	GB	extra	2	15	2	26	12	9	7	0	3
		GO	THERE		2	India		2	40	5	104	24	0	6	10	3
		GO	THERE		2	Jamaica	extra	2	32	0	77	12	0	14	0	0

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**extr a**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
GOING	TO	GO			3	Canada		2	74	5	95	48	39	97	0	6
GOING	TO	GO			3	GB		2	64	5	138	48	0	33	20	0
GOING	TO	GO			3	India	extra	2	7	0	5	12	0	13	0	0
GOING	TO	GO			3	Jamaica	extra	2	19	0	10	42	0	21	0	0
WANT	TO	GO			3	Canada	extra	2	31	15	60	36	0	7	40	6
WANT	TO	GO			3	GB		2	53	7	143	24	0	7	20	3
WANT	TO	GO			3	India	extra	2	28	5	66	12	0	19	21	0
WANT	TO	GO			3	Jamaica		2	54	10	116	18	10	41	10	10
	CAN’T	GO			2	Canada	extra	1	20	2	35	18	0	14	10	0
	CAN’T	GO			2	GB	extra	1	16	2	20	18	0	20	0	3
	CAN’T	GO			2	India	extra	1	10	0	28	0	0	6	0	0
	CAN’T	GO			2	Jamaica		1	35	7	53	48	10	14	0	10
	GONNA	GO			2	Canada		1	64	0	90	36	10	97	0	0
	GONNA	GO			2	GB	extra	1	2	0	5	0	0	0	0	0
	GONNA	GO			2	India	extra	1	0	0	0	0	0	0	0	0
	GONNA	GO			2	Jamaica	extra	1	21	0	29	24	0	21	0	0
	SHOULD	GO			2	Canada	extra	1	18	5	30	12	10	14	20	0
	SHOULD	GO			2	GB	extra	1	15	19	26	24	0	0	0	25
	SHOULD	GO			2	India		1	39	10	85	24	0	25	21	6
	SHOULD	GO			2	Jamaica	extra	1	16	7	19	30	10	0	19	3
		GO	AHEAD		2	Canada	extra	1	34	2	35	54	20	21	0	3
		GO	AHEAD		2	GB	extra	1	23	10	10	36	47	7	10	9
		GO	AHEAD		2	India	extra	1	16	20	14	36	17	0	10	22
		GO	AHEAD		2	Jamaica		1	50	7	73	89	0	7	19	3
		GO	HOME		2	Canada	extra	1	20	5	40	0	10	21	0	6
		GO	HOME		2	GB	extra	1	11	2	15	12	19	0	10	0
		GO	HOME		2	India	extra	1	6	5	5	18	0	0	10	3

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**extr a**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
		GO	HOME		2	Jamaica		1	64	17	140	24	10	41	19	17
		GO	UP		2	Canada	extra	1	36	2	70	36	0	14	0	3
		GO	UP		2	GB	extra	1	32	7	51	42	0	20	0	9
		GO	UP		2	India	extra	1	16	2	9	36	0	19	0	3
		GO	UP		2	Jamaica		1	45	7	29	42	0	103	29	0
		GO	WITH		2	Canada		1	46	15	90	30	10	28	30	10
		GO	WITH		2	GB	extra	1	31	2	56	18	9	27	10	0
		GO	WITH		2	India	extra	1	12	2	28	6	0	6	10	0
		GO	WITH		2	Jamaica	extra	1	16	5	24	18	19	0	0	7
	I	GO	TO		3	Canada	extra	1	36	7	95	6	10	7	10	6
	I	GO	TO		3	GB	extra	1	11	2	36	0	0	0	10	0
	I	GO	TO		3	Jamaica		1	40	10	87	18	19	14	29	3
	I	GO	TO		3	India	extra	1	25	0	61	24	0	0	0	0
HAD	TO	GO			3	Canada	extra	1	33	0	60	12	10	35	0	0
HAD	TO	GO			3	GB	extra	1	10	12	15	6	19	0	20	9
HAD	TO	GO			3	India	extra	1	24	10	57	12	0	13	0	13
HAD	TO	GO			3	Jamaica		1	35	10	82	18	0	14	10	10
	TO	GO	BACK		3	Canada	extra	1	34	12	50	18	0	56	10	13
	TO	GO	BACK		3	GB	extra	1	15	7	10	18	0	27	20	3
	TO	GO	BACK		3	India	extra	1	33	20	57	30	35	6	0	26
	TO	GO	BACK		3	Jamaica		1	54	20	77	66	0	48	39	13
	TO	GO	IN		3	Canada		1	41	2	20	30	10	104	0	3
	TO	GO	IN		3	GB	extra	1	19	5	36	12	0	20	10	3
	TO	GO	IN		3	India	extra	1	16	7	28	24	0	6	10	6
	TO	GO	IN		3	Jamaica	extra	1	21	2	39	12	10	14	0	3
	TO	GO	INTO		3	Canada	extra	1	31	2	10	30	29	63	0	3
	TO	GO	INTO		3	GB	extra	1	34	0	26	54	9	40	0	0

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
L2	**L1**	**Node**	**R1**	**R2**	**n-gram**	**ICE**	**extr a**	**no. ICE**	**Norm Spoken**	**Norm Written**	**Norm. Private**	**Norm. Public**	**Norm. Scripted**	**Norm. Un-scripted**	**Norm. Non-Printed**	**Norm. Printed**
	TO	GO	INTO		3	India	extra	1	16	12	5	30	17	19	10	13
	TO	GO	INTO		3	Jamaica		1	37	5	58	24	0	48	0	7
WHEN	YOU	GO			3	Canada	extra	1	28	0	35	24	10	35	0	0
WHEN	YOU	GO			3	GB	extra	1	8	5	10	18	0	0	10	3
WHEN	YOU	GO			3	India		1	22	0	33	6	0	44	0	0
WHEN	YOU	GO			3	Jamaica	extra	1	21	0	39	30	0	0	0	0
		GO	INTO	THE	3	Canada	extra	1	31	0	30	30	10	49	0	0
		GO	INTO	THE	3	GB	extra	1	21	10	20	18	28	20	0	13
		GO	INTO	THE	3	India	extra	1	16	10	5	24	9	31	0	13
		GO	INTO	THE	3	Jamaica		1	42	5	48	30	29	55	0	7

Appendix III: ***Between and Within the ICE p-values***

Where $p \leq 0.05$ the number is shown in red (the null hypothesis is rejected).

1	*2*	*3*
N-Gram	***Colonies***	***ICE***
go	0.001	0.816
and go	0.044	0.352
I go	0.007	0.597
they go	0.002	0.928
to go	0.001	0.909
we go	0.008	0.751
you go	0.002	0.918
go back	0.005	0.622
go in	0.004	0.687
go into	0.006	0.539
go on	0.008	0.478
go out	0.035	0.244
go through	0.007	0.800
go to	0.002	0.745
have to go	0.007	0.613
to go to	0.005	0.944
go to the	0.007	0.584
can go	0.017	0.611
just go	0.004	0.548
go and	0.001	0.998
go for	0.007	0.332
to go and	0.002	0.864
you go to	0.002	0.933

1	*2*	*3*
N-Gram	***Colonies***	***ICE***
go back to	0.079	0.821
will go	0.242	0.395
go down	0.027	0.894
go there	0.015	0.375
going to go	0.021	0.169
want to go	0.003	0.900
can't go	0.009	0.374
gonna go	0.161	0.039
should go	0.052	0.832
go ahead	0.050	0.999
go home	0.650	0.055
go up	0.005	0.790
go with	0.087	0.078
I go to	0.034	0.131
had to go	0.088	0.974
to go back	0.130	0.401
to go in	0.006	0.963
to go into	0.012	0.994
when you go	0.033	0.805
go into the	0.007	0.667

1	*2*	*3*
N-Gram	***Colonies***	***ICE***
come	0.002	0.554
they come	0.005	0.973
to come	0.007	0.631
you come	0.007	0.458
come and	0.004	0.436
come back	0.004	0.969
come from	0.187	0.345
come in	0.008	0.875
come on	0.003	0.832
come out	0.003	0.957
come to	0.124	0.055
come up	0.004	0.562
will come	0.224	0.154
come down	0.093	0.276
has come	0.282	0.067
have come	0.860	0.008
I come	0.046	0.072

1	*2*	*3*
N-Gram	***Colonies***	***ICE***
come here	0.167	0.102
come into	0.037	0.392
and come	0.023	0.389
had come	0.316	0.192
not come	0.713	0.004
we come	0.056	0.206
come up with	0.209	0.047
to come and	0.077	0.185
to come to	0.021	0.766
to come up	0.025	0.227
you come to	0.068	0.389
come back to	0.007	0.544
come out of	0.102	0.396
come to know	0.229	0.026
come to the	0.207	0.784

Appendix IV: ***Pairwise p-values for come- and go-grams that have been shown to be significantly different <u>between</u> the ICE.***

Where $p \leq 0.05$ the number is shown in red (the null hypothesis is rejected).

1	2	3	4	5	6	7
	India-GB	**India-Jamaica**	**India-Canada**	**GB-Jamaica**	**GB-Canada**	**Jamaica-Canada**
gonna go	0.670	0.093	0.010	0.206	0.03	0.368
have come	0.001	0.220	0.015	0.048	0.438	0.228
not come	0.218	0.065	0.000	0.538	0.019	0.085
come up with	0.485	0.051	0.012	0.211	0.068	0.566
come to know	0.005	0.024	0.024	0.591	0.591	1.000

Appendix V: ***Pairwise p-values for come- and go-grams that have been shown to be significantly different <u>within</u> the ICE.***

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Private/ Public	Private/ Unscripted	Private/ Scripted	Private/ Non-Printed	Private/ Printed	Public/ Unscripted	Public/ Scripted	Public/ Non-Printed	Public/ Printed	Unscripted/ Scripted	Unscripted/ Non-Printed	Unscripted/ Printed	Scripted/ Printed	Non-Printed/ Printed
go	0.250	0.271	0.001	0.019	0.000	0.960	0.028	0.230	0.009	0.024	0.211	0.008	0.689	0.162
and go	0.147	0.211	0.003	0.084	0.009	0.841	0.121	0.783	0.240	0.080	0.634	0.169	0.707	0.368
I go	0.057	0.146	0.007	0.104	0.000	0.652	0.438	0.783	0.064	0.220	0.861	0.021	0.282	0.033
they go	0.192	0.920	0.016	0.002	0.011	0.160	0.270	0.064	0.210	0.012	0.001	0.008	0.880	0.546
to go	0.194	0.211	0.001	0.024	0.000	0.960	0.036	0.342	0.012	0.032	0.317	0.011	0.689	0.121
we go	0.367	0.483	0.035	0.003	0.003	0.841	0.230	0.038	0.043	0.035	0.023	0.026	0.409	0.960
you go	0.294	0.271	0.003	0.004	0.001	0.960	0.051	0.004	0.021	0.057	0.072	0.024	0.726	0.653
go back	0.484	0.803	0.003	0.034	0.007	0.653	0.026	0.154	0.045	0.007	0.061	0.014	0.822	0.565
go in	0.616	0.499	0.007	0.023	0.002	0.861	0.027	0.075	0.008	0.042	0.109	0.013	0.652	0.381
go into	0.822	0.499	0.240	0.012	0.064	0.653	0.161	0.006	0.038	0.064	0.001	0.012	0.499	0.499
go on	0.368	0.409	0.094	0.014	0.001	0.940	0.438	0.121	0.011	0.495	0.104	0.009	0.076	0.317
go out	0.134	0.024	0.011	0.024	0.002	0.453	0.294	0.453	0.099	0.764	1.000	0.368	0.548	0.368
go through	0.726	0.920	0.068	0.230	0.005	0.802	0.029	0.121	0.001	0.054	0.193	0.003	0.317	0.104
go to	0.194	0.057	0.000	0.089	0.000	0.549	0.021	0.689	0.028	0.089	0.841	0.110	0.920	0.072
have to go	0.341	0.127	0.007	0.002	0.002	0.565	0.084	0.038	0.033	0.250	0.133	0.121	0.689	0.960
to go to	0.121	0.014	0.000	0.177	0.002	0.368	0.040	0.841	0.134	0.250	0.014	0.548	0.582	0.089
go to the	0.499	0.051	0.002	0.017	0.003	0.202	0.015	0.089	0.020	0.250	0.670	0.293	0.920	0.532
can go	0.437	0.281	0.067	0.003	0.005	0.764	0.293	0.027	0.042	0.452	0.057	0.084	0.328	0.861
just go	0.213	0.169	0.003	0.001	0.005	0.897	0.082	0.029	0.119	0.108	0.040	0.153	0.856	0.533
go and	0.161	0.395	0.000	0.012	0.001	0.582	0.020	0.271	0.068	0.004	0.099	0.017	0.617	0.468
go for	0.270	0.176	0.001	0.006	0.013	0.802	0.020	0.098	0.168	0.038	0.161	0.259	0.341	0.783
to go and	0.635	0.546	0.004	0.006	0.002	0.763	0.044	0.066	0.029	0.021	0.032	0.013	0.860	0.725
you go to	0.276	0.417	0.001	0.004	0.002	0.780	0.033	0.076	0.043	0.016	0.040	0.021	0.919	0.800
go down	0.644	0.472	0.009	0.016	0.035	0.797	0.031	0.051	0.100	0.057	0.090	0.165	0.607	0.758
go there	0.246	0.055	0.008	0.016	0.001	0.449	0.130	0.208	0.030	0.449	0.614	0.158	0.512	0.364
going to go	0.797	0.878	0.072	0.051	0.031	0.918	0.040	0.027	0.016	0.057	0.040	0.024	0.719	0.837
want to go	0.132	0.040	0.000	0.132	0.001	0.581	0.033	1.000	0.067	0.114	0.581	0.201	0.763	0.067

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Private/ Public	Private/ Unscripted	Private/ Scripted	Private/ Non-Printed	Private/ Printed	Public/ Unscripted	Public/ Scripted	Public/ Non-Printed	Public/ Printed	Unscripted/ Scripted	Unscripted/ Non-Printed	Unscripted/ Printed	Scripted/ Printed	Non-Printed/ Printed
can't go	0.227	0.198	0.003	0.003	0.006	0.938	0.072	0.072	0.123	0.085	0.085	0.143	0.797	0.797
go ahead	0.270	0.109	0.548	0.176	0.146	0.007	0.088	0.014	0.011	0.316	0.802	0.880	0.394	0.920
go up	0.799	0.721	0.005	0.028	0.053	0.541	0.002	0.014	0.028	0.041	0.066	0.114	0.386	0.799
I go to	0.047	0.007	0.012	0.056	0.002	0.475	0.610	0.939	0.251	0.838	0.429	0.665	0.524	0.221
to go in	0.497	0.407	0.006	0.006	0.005	0.880	0.039	0.039	0.033	0.056	0.056	0.047	0.940	0.940
to go into	0.379	0.258	0.466	0.063	0.167	0.802	0.108	0.006	0.024	0.063	0.003	0.012	0.514	0.633
when you go	0.511	0.247	0.014	0.014	0.009	0.617	0.073	0.073	0.051	0.197	0.197	0.148	0.875	0.875
go into the	0.821	0.291	0.598	0.015	0.138	0.407	0.451	0.008	0.087	0.113	0.000	0.011	0.339	0.339
come	0.549	0.617	0.008	0.080	0.000	0.920	0.040	0.250	0.004	0.032	0.211	0.003	0.395	0.080
they come	0.821	0.393	0.102	0.066	0.022	0.530	0.063	0.039	0.012	0.013	0.007	0.002	0.513	0.651
to come	0.424	0.317	0.014	0.582	0.001	0.841	0.099	0.803	0.012	0.147	0.653	0.021	0.395	0.006
you come	0.393	0.599	0.009	0.051	0.001	0.745	0.080	0.271	0.017	0.038	0.154	0.007	0.531	0.202
come and	0.367	0.745	0.004	0.140	0.002	0.565	0.045	0.565	0.028	0.010	0.250	0.005	0.841	0.104
come back	0.368	0.653	0.004	0.021	0.002	0.653	0.046	0.162	0.032	0.014	0.064	0.009	0.881	0.453
come in	0.881	0.438	0.068	0.084	0.020	0.354	0.094	0.115	0.029	0.009	0.012	0.002	0.617	0.548
come on	0.237	0.131	0.000	0.003	0.004	0.744	0.019	0.074	0.087	0.044	0.145	0.167	0.530	0.940
come out	0.940	0.499	0.010	0.006	0.009	0.548	0.012	0.007	0.011	0.057	0.038	0.051	0.960	0.900
come up	0.707	0.565	0.230	0.012	0.005	0.342	0.115	0.004	0.001	0.532	0.051	0.024	0.104	0.764
I come	0.209	0.107	0.002	0.191	0.018	0.725	0.063	0.960	0.268	0.131	0.763	0.450	0.450	0.291
come into	0.130	0.208	0.650	0.166	0.632	0.801	0.290	0.004	0.046	0.420	0.008	0.082	0.351	0.364
and come	0.783	0.707	0.024	0.267	0.006	0.920	0.047	0.408	0.013	0.060	0.467	0.017	0.616	0.098
to come to	0.582	0.423	0.089	0.381	0.020	0.802	0.249	0.153	0.075	0.367	0.093	0.127	0.531	0.001
to come up	0.920	1.000	0.841	0.016	0.035	0.920	0.920	0.021	0.045	0.841	0.016	0.035	0.057	0.764
come back to	0.980	0.430	0.056	0.126	0.018	0.415	0.059	0.133	0.019	0.007	0.020	0.002	0.647	0.401

Appendix VI: ***ICE file designations***

PRIVATE	**S1A**
Direct Conversations	S1A-001 *to* S1A-090
Telephone Calls	S1A-091 *to* S1A-100
PUBLIC	**S1B**
Class Lessons	S1B-001 *to* S1B-020
Broadcast Discussions	S1B-021 *to* S1B-040
Broadcast Interviews	S1B-041 *to* S1B-050
Parliamentary Debates	S1B-051 *to* S1B-060
Legal Cross-examinations	S1B-061 *to* S1B-070
Business Transactions	S1B-071 *to* S1B-080
UNSCRIPTED	**S2A**
Spontaneous Commentaries	S2A-001 *to* S2A-020
Unscripted Speeches	S2A-021 *to* S2A-050
Demonstrations	S2A-051 *to* S2A-060
Legal Presentations	S2A-061 *to* S2A-070
SCRIPTED	**S2B**
Broadcast News	S2B-001 *to* S2B-020
Broadcast Talks	S2B-021 *to* S2B-040
Non-broadcast Talks	S2B-041 *to* S2B-050
NON-PRINTED	**W1**
NON-PROFESSIONAL WRITING	**W1A**
Student Essays	W1A-001 *to* W1A-010
Examination Scripts	W1A-011 *to* W1A-020
CORRESPONDENCE	**W1B**
Social Letters	W1B-001 *to* W1B-015
Business Letters	W1A-016 *to* W1B-030
PRINTED	**W2**
ACADEMIC WRITING	**W2A**
Humanities	W2A-001 *to* W2A-010
Social Sciences	W2A-011 *to* W2A-020
Natural Sciences	W2A-021 *to* W2A-030
Technology	W2A-031 *to* W2A-040
NON-ACADEMIC WRITING	**W2B**
Humanities	W2B-001 *to* W2B-010
Social Sciences	W2B-011 *to* W2B-020
Natural Sciences	W2B-021 *to* W2B-030
Technology	W2B-031 *to* W2B-040
REPORTAGE	**W2C**
Press News Reports	W2C-001 *to* W2C-020
INSTRUCTIONAL WRITING	**W2D**
Administrative Writing	W2D-001 *to* W2D-010
Skills & Hobbies	W2D-011 *to* W2D-020
PERSUASIVE WRITING	**W2E**
Press Editorials	W2E-001 *to* W2E-010
CREATIVE WRITING	**W2F**
Novels & Stories	W2F-001 *to* W2F-020

Appendix VII: ***Concordances of to go to***

Canada/Private

1 ne and do {2 [2 counselling 2] with cases rather than waiting them to go to {3 [3 ah transition houses 3] $C [1 When I 1] 1} $C [2 Mhhm 2] 2
2 ay that $B (,,) W'I know but (,,) that was before she had a chance to go to Burlington $A (,,) Anyway I'm going to ask Fayed if he can bring
3 n I enjoyed that $A One of the Indians in my group decided (,) to (,) go to a a barber shop to get his hair cut just so he could tell his fr
4 here this year (,) and last year I was part-time so I didn't have to go to those $A Right ((laugh)) Oh that's right Well you've done okay
5 t matter even if we had made plans (,) she would have changed them to go to Burlington anyways (,,) { [So] $A [Oh] } Oh you mean she's n
6 so it keeps me pretty active so And also tomorrow night I'm going to go to the arena and see about starting skating again down here $A Mkay
7 ere's one move left (,) but if I have to go to work I'll just have to go to work $B Well that's why I'm gonna uh (,) do all the big things (
8 [1 Yes 1] 1} $A [2 Yes I have to do that next weekend 2] 2} I have to go to New Brunswick again $B Far $A So it's going to be nine hours plu
9 these breakfasts uh these breakfast things I mean if I'm not able to go to them cos I'm in class { [(,,) you] you know you should probabl
10 d)) I expect the kids are going to go on and on Now they're going to go to Paris for their honeymoon $A Oh I that's what I was going to ask
11 that type of movie so it's just as well we went (,) He didn't want to go to that Greek restaurant did he $B No { [Nei neither] did I actua
12 yway whatever So I'm I'm sort of looking forward to that I'm going to go to the group thing { [(,,)] uhm because I think what he wants to
13 deal $B Hmm $A Uhm (,,) when I was in grade one we had we all had to go to the high school This is quite a traumatic event We all had to go
14 go to the high school This is quite a traumatic event We all had to go to the high school to see a film (,) and I had never seen televisio
15 i { [you know] $B [Mhh hmm] } $A At six he got sent to England to go to school His parents were there until he was like ten $B Mhh hmm $
16 ion where like uh you know there's one move left (,) but if I have to go to work I'll just have to go to work $B Well that's why I'm gonna u
17 to uh (,,) { [Boston] $A [You going to] } Boston $B I'm going to go to Boston I'm going to go work out of Boston for the next five seve
18 hh] } $A This was when my parents were bit better off you know Go to go to eat in the restaurant there at the Sheraton Hotel there's a the
19 kay yeah (,,) ((sigh)) Oh I missed all these notes I didn't want to go to class it's so boring $A What (,) this is biology $B Yeah (,,) So
20 lt I think cos you're kind of in the middle of nowhere $B I'd have to go to St-Jovite { [so] $A [Well] } that wouldn't be too bad $B No
21 ade eight Pretty soon after a while I learned that you didn't have to go to school ((laugh)) So I just wouldn't go in those days You know
22 a do eh (,,) $B Get wrecked $A S Get wrecked ((laugh)) Do I have to go to your meeting on your year (,,) $B Not unless you want to $A Well
23 lunch and then we get to just (,) maybe you know I I would prefer to go to the museum on Sunday Like so Saturday wa'll have been in the car
24 e had four courses in Cree literacy (,,) And now they're (,) ready to go to the next stage which is to (,) which is to develop their own cre
25 t I'd just sell all my furniture and everything I have (,) and try to go to RADA or something in { [England] $A [Ya] } $B Like just hurl
26 ing and driving { [in those] days $A [Okay] } $B But they used to go to the port the Stork Club and go to dances and then go back to Lon
27 what you did wrong $A I think if you want to see the exam you have to go to the professor $B Oh $A But you don't get them back { [at all]
28 y well (,) you know want me to meet you at the hospital cos I have to go to work right now but I can meet you at the hospital And the girl s
29 in maybe at Janette 's or something $B Ya maybe I definitely want to go to Alison 's $A I'm having that Crispy Crunch by the way So { [don
30 e same night as Janette 's party { [((laugh))] $A [You wanted to go to] } { [Janette 's] $B [We're] } such jetsetters ((laugh))
31 so then the next year we said no let's do it And we really wanted to go to like San Fransisco and stuff $B Mhh $A But then on the other han
32 f the if they're already in there it's probably better to go there to go to the schools and talk to the principal It'll probably be easier t
33 many] $B [And who knows] } I might have to go and I might have to go to like some teachers' college out in boonieville </I>
34 one of my best friends her dad worked on the railroad and he used to go to the States all the time { [(,,)] so he's always brought them b
35 $B [Mh hmm] } $A Totally different mindspace And we had planned to go to California the year before and I (,,) like put the brakes on it
36 y shrewd I just hope that the two corroborate cos then she'll have to go to a third { [if the second guy ((a few words)) the first guy ((
37 e second guy ((a few words)) the first guy ((a few words)) has to go to a third] she'll get really confused $B [Oh God really ((laugh
38 I mean I'd much ra it would make infinite more sense to to be able to go to Ottawa U (,,) Uhm $A I want Ottawa U to start a midwifery progra
39 } $B Unless you wanna push another two hundred miles down the road to go to that uh { [(,,)] Gitchigoumi place we were at last time you kn
40 t think a thing] $A [I guess and (,)] } he said he wasn't going to go to the wedding or was thinking not going and apparently she got him
41 t a fluke It was ah (,,) a friend of a friend and we just happened to go to this woman's house and say we we stopped by to visit for a minut
42 me in Kyoto { [(,,)] and uhm $A [Awright] } Well that's a city to go to $B Ya (,) it's a beautiful city And uh (,) we uh (,) decided one

GB/Private

y what I've said before basically but I'm not (,) $B Are you going to go to all uhm the day on of the phonology lectures $A I think I ought
something (,) something (,) about that actually I'd always wanted to go to Australia And I met this Australian in London and I lived with h
er $B Well now I am I think for years and years I've always wanted to go to Australia and I 've always kept (,) very good ties with Jeremy s
want to might he But he might do I think he does actually $A Want to go to bed with you $B No no no no no no Want to go out for a drink Jum
1 but 1] of course (,) presumably if she is saying no I don't want to go to boarding school {2 [2 and yet you see she'd have such a lot of f
2} not $B Probably not $A Yeah Well it depends $C Why do you want to go to Brussels $A Uhm { [it depends] It might not be the seventy-sev
right if you 've got a car but] $D [I mean i it was you have to] } go to Cambridge and get the village bus out And it took over three and
how did it] } happened $A He got a job $B But what made him want to go to Disneyworld for the job $A Well he he rang me up in great excite
now Some certain electrical goods like I mean like I mean if I was to go to Dubai I can get things like electrical { [goods ((unclear-word
e reason that you treat people well though is it cos you're scared to go to Hell (,) Well for some people it is It doesn't seem quite correc
ually worked out very well $B It did because we didn't really want to go to Holland and certainly not to northern Holland which is not near
comes up $B Aha (,) $A But I mean (,,) And I've got so many events to go to I mean ((laughs)) I know that sounds a bit odd but I mean I've
B Well Doncaster uh Concert Hall { [one place and uh he was going to go to I don't think he's at Edinburgh this year $C [Really] $C What
into a lot of opera and if you start giving her a series of events to go to it'll make her shy away $B I see $A She hadn't She hadn't been (
uhm but in August yeah uhm Did did I tell you that I was c hoping to go to Japan to teach [1 Uhm $B [2 Uhm 2] 2} $B Well no It was a vague
ht Well it's now definite (,) {1 [1 uhm 1] and I I definitely want to go to Japan uhm until I {2 [2 uh unclear-words ((two or three words)
B But { [uhm] $A [I'll] } be completely uh $B I mean I'm going to go to Kaye's to see w if he's got anything new to say (,) after thirty
y in the studio {1 [1 says 1] that he's been going on about having to go to London to to cut a record tomorrow so (,) it sounds OK [2 (,) {2
she's got (,) he's got somebody living in his house who's who used to go to Mrs Parsons ' $B So how did you meet up with him then (,) $A Oh
$A That's all you've got to do $B I've got nothing I'd quite like to go to Richmond Park because I was reading about it in this novel I isn
thing and I went 3] Yeah Oliver Oliver Pemb Pemberton $B [1 I used to go to s 1] 1} $B [2 I went to school with Emma 2] 2} $Z $B Didn't I go
hree days so I've just thought I'd have an extra few days just to (,) go to Sainsbury's (,) do my room (,) And then next weekend I'm going d
at I'd pretend to be sick (,) (($A-laughs)) and so I didn't have to go to school I got quite good at like heating up thermometers and stuf
because I wanted to be Prime Minister of England it's just I want to go to school thank you And the second thing that you have to appreciat
d more uncertainty as I left home just because I mean I knew I had to go to school and do things really that were provided [1 (,) {1 1] Bala
oy did I need forty grand who doesn't {4 [4 you know 4] But it was to go to Siberia for six weeks and photograph (,) a car rally and I thoug
$B ((unclear-words)) (,,) $A Hello dear (,) $B I think I'm going to go to sleep now (,) so that horrible bug can not report me any more $A
he fact that Terry is a greengrocer Ray knew me you see and I used to go to Terry and Mary for years [(,) {] ((unclear-words)) $A [Mhm
{ [worries me a little bit] $B [I suppose so Are you going to] } go to that (,,) $A Which $B The first two Palmer { [and Firth Yes] $
))] [That's so] } no that's really unkind $A Sorry But she had to go to the loo $B Didn't didn't $F That's horrible I think that's reall
is that at $B I don't know I can ask her I dou doubt she'd want me to go to the same place as her though because she is a bit possessive abo
eens and twenties I suppose every Saturday one of my pleasures was to go to the local bookshop and buy another volume in the Everyman Librar
ford $A No why are you and { [Richard] $B [Yeah] } we're going to go to (( )) { [this weekend] $A [Are you] } $B Yeah $A Ah $B Right
(a few words))] $B [Well he goes] } up to North London anyhow to go to this therapy group [1 (,) {1 1] uhm and she lives up in North Lo
se dates and they 've only got four days there $C Yeah I'm tempted to go to ((unclear-words)) $A Yeah wish they'd never written the ruddy
ster for this week and next week (,) And after next week I'm going to go to Weymouth Did I tell you that I was [(,) {] going for a job in

India/Private

it's a we have there is something I'm missing and I we're planning to go to Tawan (,,) when it becomes less (,) cold you know sometime in Fe
this person working with you then we found out his address we used to go to him we informed him (,) So (,) then he came (,) to Bombay we tal
uh once it so happened you know that kind of fun as you might have to go to this university to pick up the ((one or two words)) (,) And I
lf (,) $A You have used it $B Yes (,,) $A Accha (,,) so I told her to go to (,) Walawalkar (,) uhm trust $B No no (,) in Chandramukh's (,,)
n if it is true (,) till now there is no intimation $B Do you like to go to see (,) ((two words)) in Bombay to stay there (,) $A Okay I (,
go (,) They will force me to go (,) and after going there I have to go to the (,,) ((two words)) school (,,) and attend that uh (,) yout
an (,) $B They asked me to come to Bombay (,) $A Oh (,) so you had to go to Bombay { [then] $B [Haan] } we had our meeting there (,) $A
get the experience $A Uhn (,) $B In our postal course we have to (,) go to the some school and teach them and then they will give you certi

9 (,) She's personally seen that fellow {2 [2 uhm uhm (,) They used to go to that saint $A [1 I see 1] 1} $A [2 Uhm uhm 2] 2} $A When did he
10 because she was uh (,) bit in hurry (,) today $B Yeah $A She has to go to her place (,,) She may be { [on her] way $B [What about] } $
11 es are out of the scene uh they cannot (,,) have any thing to (,,) to go to the people (,,) uhm I mean again for something else or so (,,) $
12 [uh (,) kidney and I had no time to go the hospital because I had to go to the passport office And I noticed that on fourteenth of (,,) thi
13) So then after fourth standard when I was nine years so (,) I had to go to a nearby town Taluka place $B Uhm uhm (,) $A So (,) three miles
14 and that also in the midnight uh (,) $A Haan $B So they asked them to go to the (,) morgery (,,) whereas (,,) in the morgery even the (,,) f
15 bliophile ((laughter)) aren't you (,,) So I would better ask you to go to the (,) Mahalaxmi temple (,) have you been there $A Yes I've bee
16 have heard about uh (,) one college is that one (,) guy was asked to go to the ladies hostel (,) $A Haan (,) $B Uhn the (,,) warden was (,)
17 second year (,) { [uhn (,) And uh they got to go to (,) they have to go to different industries (,,) where (,,) the students are expected (
18 tional causes exactly and all that (,,) And then I got scholarship to go to Chicago (,) I would say (,) Masica was (,,) some what influentia
19 (,,) Actually I'm (,) I wanted to go (,) early because (,) I want to go to this Holy Cross (,) my daughters school (,) $A [1 Okay 1] 1} $A
20 nths at the end of the second year (,) { [uhn (,) And uh they got to go to (,) they have to go to different industries (,,) where (,,) the
21 $A Yeah (,) $B Okay I'll (,) call him $A He want to go to (,) airport from here to airport $B Ahn (,) $A But the flight was
22 ,) the P G is only at the university centre $B Uhm (,) $A So I had to go to (,) Dharwar for M Sc (,) $B There also you stayed in the hostel
23 beautiful only they may have a $B I should really (,) I feel like to go to Kashmir for ((one word)) (,) Kashmir and Switzerland is same n
24 (,) $B Hello $A I'm not (,) likely to come this thing I'm likely to go to Goa for a conference (,) $B When is that (,) ? $A Uh it is (,) b
25 kept (,,) So (,,) sometimes they (,,) I mean they ask the students to go to the (,,) dissection hall and identify (,,) so many things (,,) A
26 her place was Dharwar (,) So Dharwar was bit away (,) so I { [had to go to] I did my this uh $B [So you did it in Bagalkot] } $B And whe
27 Since in the morning there { [is a great rush] (,) Everyone has to go to their department before nine $A [Great rush yeah] } $A Yes (,)
28 are back from their department But in { [the] morning they have to go to the department $A [Ah ha] } $A Yes yes (,) $B And if they get
29 ans (,) $A I think you both had decided that uh (,,) you both have to go to B D S $B Yeah she had hoped (,) more hopes from this side B D S
30 iversity (,) judge $B We've discussed it we've discussed $A I want to go to the court today (,) $B You (,) ? $A Yeah myself or uh (,) someon
31 local school (,,) in uh fourth standard (,) And my daughter is yet to go to school $C I see $B Where is your (,) son studying Asha ? $C My s
32 y there (,) $A Okay I (,) I don't know I can't say whe that I like to go to Bombay (,) but (,) I wanted to be economically independent (,) i
33 I feel very tired and sleepy $A Yeah (,) ((laughs)) And you have to go (,) to your branch very early I { [suppose] $B [Very] } early y
34 i (,) $B You are going to Madurai (,) ? $C Madurai $B Oh $C I have to go to Madurai (,,) $B Uh what about your (,) father (,,) ? $C That's t
35 Chandramukh's (,,) shop $A I did not know that I have told her { [to go to] Walwalkar $B [Near to] } $B Near to Walawalkar ((one word)
36 y will be more convenient na (,,) $B That's true $A Would you want to go to Monday (,,) $B Look you're going to give me a party (,,) You're
37 uh (,) and (,) uh like uh (,) last month (,) I was told that I had to go to Japan $B [Uh] } $B Uhm $A It's a rare opportunity (,) and uh I
38 ol (,) Today (,) yesterday rickshaw has not come no so (,,) I have to go to (,) collect $B I also go to collect and (,,) ((one word)) drop
39 Out of Tamil Nadu I will like to go to Goa (,,) $A You would love to go { [to Goa] $B [Yeah] } but uh (,) $A Haan (,,) $B But I have se
40 ,) ? $B You are going to Madras ? $C No (,) $B Then (,,) $C I have to go to Madurai (,) $B You are going to Madurai (,) ? $C Madurai $B Oh $
41) And many times I had experienced it that I couldn't get a chance to go to temple (,,) and uh I just pray in heart and my wishes come true
42 t thirty miles away from my place $B Uhm $A So (,) naturally I had to go to Bagalkot (,) And another place was Dharwar (,) So Dharwar was bi
43 d leave his flee away from his home town Lucknow and he was forced to go to Calcutta And there (,) amidst lots of tragic situations and grim
44 A [1 Uh 1] 1} $A [2 Uhn 2] 2} $B My God $C And I was the first man to go to driver and to scold him (,) because see whether you are a driver
45 (,) Haan I forgot to tell you na $A Yeah (,) $B See I was supposed to go to Goa (,) $A Yeah (,) $B But I (,) that pro that was cancelled (,)
46 I had been working (,) uh I in fact was not at all (,) uh willing to go to such a place (,) which is quite a far off from my place $B Yeah
47 ut] of out of $B [Tamil Nad] } $B Out of Tamil Nadu I will like to go to Goa (,,) $A You would love to go { [to Goa] $B [Yeah] } but

Jamaica/Private

1 wn person or somebody>s child because only my child I can tell not to go to the gate $C You see if I'm go to go to the ((word(s))) for my
2 g to Harbour View at this time of morning and people out here have to go to Bull Bay and you have on the Bull Bay route and it I said I'm go
3 y're from upper class {3 [3 (,) 3] families so they have resources to go to extra classes and whatever (,) uhm {4 [4 (,) 4] I guess what we
4 every three week {2 [2 but I 2] go home like every week cos I have to go to the doctor every weekend $B [1 Uh-uhm 1] 1} $B [2 Wow 2] 2} $B T
5 ed the paper and he he gave us back you know the results so we had to go to his office So we were all standing outside He has an open door p
6 wood $A Mhm $B Well personally uhm it was it was a personal choice to go to Well of course we have to speak English I suppose eh cos your in
7 know this $B [No] } $B No Not recently Well my father forced me to go to (,) a college that was you could board you could live on so I de

middle of cane fields So we had to ride three miles on Marle Road to go to school But it was fun People got robbed every so often and raped
ed and getting the right grade and you know cos everybody dream is to go to UWI or some university abroad so I mean getting prepared was a r
essful at both places { [(,)] but for personal reasons I decided to go to (,) Shortwood (,) Uhm $A [Uhu] } $A When you said both places
n addition to school was church the Catholic Church You you had to to go to well we had to go to St Monica's Home because it was close to us
nt to France she went on the France program {1 [1 where 1] you get to go to France {2 [2 to 2] teach in France {3 [3 (,) 3] for a year and s
e that you had to abide to the same rules (,) like (,) uhm you had to go to classes you had to (,,) after class you can't go where you're go
nd I'm in your class and thereafter you just find small items just to go to him and go to You don't have to go every week { [(,)] He has u
Oh really 1] 1} $B [2 That's cool 2] 2} eh That's cool I I I want to go to { [Sweden I want to travel the world] $A [You should ((word
free yeah That was that was good that was very good If I don't get to go to uhm Fully Loaded this year I have to reach the Stages have to It
h you you you { [rode the bike] $B [I used to ride the bike] } to go to school $A Oh So everybody rode their bicycle $B No because norma
oing it $A And then after the bio and the and the English you have to go to n a nursing school $B Yeah then I'll after that I am $A You have
oing to Emerson college $A Uhm even though it's my goal uhm I want to go to Emerson uhm I realise that the whole financial situation will be
y normal {2 [2 (,) 2] enjoy myself {3 [3 (,) 3] Right And I wanted to go to the movies too probably next week $A [1 Okay 1] 1} $A [2 Mhm 2]
r it be the scientists or just like straight technicians they have to go to field at some point in time Before all these people got hired I
'd like to do someone-clears-throat)) I went to (,) I go I'd like to go to Germany visit the place and buy my stuff (,) And I think if you
u're moving off hall $B No I think I staying for the summer I want to go to summer school $A Okay That's why you're here still { [(,)] How
son don't even go to Ewarton but (,) by right is not supposed to (,) go to that school because of the l locale uhm location {3 [3 (,) 3] th
was in economics but then after finishing I decided I didn't want to go to that area and now I'm doing my masters in management information
ith the I D University students get in free {2 [2 (,) 2] So I want to go to that but I'm not sure $A [1 Okay 1] 1} $A [2 Okay 2] 2} $A Right
seven o'clock and you don't stop talking till nine until I wake up to go to school $B [1 No 1] 1} $B [2 Me 2] 2} $B No that is so not true {
rm for lights out at nine $B [Mhm] } $B Hold on You mean you had to go to your bed at nine o'clock $A Right That that is the rule but we d
So we had fun 2] $A [1 Mhm 1] 1} $A [2 Oh 2] 2} So why you choose to go to that school Your mother or it was just near to where you lived $
ight {2 [2 wanting me 2] to tell her back {3 [3 stories for 3] her to go to sleep $B [1 Mhm 1] 1} $B [2 Mhm Mhm 2] 2} $B [3 Mhm 3] 3} $B Mhm
hat was something I wanted since I was in third form I just wanted to go to UWI and I mean getting my acceptance letter I mean I first I got
ied to the J F as well so should I be called in I have to leave in to go to that $B Tell me more about that wanting to become a pilot ultima
't need to go to the lecturer but just for your sake only you need to go to him and say listen my name is X and I'm in your class and therea
ords)) all that you know energetic like you get up like Oh I have to go to work like Jamaicans like ah I have to go to work this morning an
h 1] 1} $B [2 Extra 2] 2} lights Oh Okay ((laughs)) Nine o'clock to go to your bed just like a { [grand-ma of sixty-eight] My God ((lau
] } $A But it has so many opportunities All you see stuff you know to go to $B Yeah I know but then sometimes I have to be doing something l
e you might be understanding the course so you feel you don't need to go to the lecturer but just for your sake only you need to go to him a
m not really a party person but at least you know that if you want to go to a party it's just at the union So you just got to walk upstairs
A Alright $B But we're having serious problems because people have to go to work six o'clock in the morning and you're out by the bus stop f
h $A Okay Yeah ((laughs)) $B Yes uhm personally it was my choice to go to Shortwood really cos I was successful at both places { [(,)] b
{1 [1 is it just 1] so different You want to travel {2 [2 you want to go to all over Yes I see 2] $A [1 Yeah 1] 1} $A [2 And go all over All
t I wanted to do And I described to her the fact that I might want to go to Mass Comm but I didn't want to I didn't want I don't I didn't ex
take care of nobody {8 [8 (,) 8] And and her man him husband used to go to uh it ((((word)) him {9 [9 (,) but 9] him didn't love him dau
eted ((one-or-two-words)) $B [Uhm] } $B Well my sister want me to go to Cayman because she haven't seen me in a while want to bond I don
a do something that makes a lot of money and you don't really have to go to work a lot ((giggling)) But then again I've heard that those t
He was getting something like eight dollars an hour He didn't need to go to college $B To get eight dollars an hour $A He could have gone fr
mandate from the people he wouldn't have had a better opportunity to go to the public now for a mandate Uhm I mean the economy seem to be (
quad that's worship { [(,)] After worship you have prep You have to go to the prep room for uhm two hours (,,) Right And then you go home
nly my child I can tell not to go to the gate $C You see if I'm go to go to the ((word(s))) for my mother it's a problem because he he's s
nery $A Mhm {1 [1 (,) And 1] Sundays we have chapel Everybody have to go to the chapel Well except for me I didn't like church I used to hid
Based on the fact that I need to go back and study (,) like I want to go to my P h D and I'm thinking hey maybe that's an option maybe I sho
ther was very protective and and and you know he wouldn't allow me to go to parties and stuff but then I can say maybe that's what made me u
was church the Catholic Church You you had to to go to well we had to go to St Monica's Home because it was close to us {2 [2 (,) 2] {3 [3 a
You had to go 2] 2} $B [3 Mhm Mhm 3] 3} $B [4 Mhm 4] 4} $B [5 Had to go to ((word)) 5] 5} }$B[2 You had
get up like Oh I have to go to work like Jamaicans like ah I have to go to work this morning and they don't want to come out of bed is like
5] 5} $A Oh my God you know (,) And also I think all of them have to go to extra class $B And they do {1 [1 All of 1] them go to extra clas
u] } from experience A lot of my friends' kids that come up there to go to college and habitually all of them because being here you know t
go to Arawak museum we 2] {3 [3 learn about the Arawak 3] We had to go to here we learn about {4 [4 that 4] Every time you learn something
d afterwards I hope to do ((words)) $B Okay So where do you want to go to study speech therapy $A I'm not sure yet you know uhm but uhm I
owing year I want to take up uhm hopefully I can get a scholarship to go to Mexico to study $A [1 Mhm 1] 1} $A [2 Mhm 2] 2} $A So uhm have y
ourse {2 [2 (,) 2] field methods in linguistics and we're supposed to go to Guyana and Surinam $A [1 Okay 1] 1} $A [2 Mhm 2] 2} $A For how l

me {7 [7 (,) 7] you have an entitlement {8 [8 (,) 8] you need just to go to an attorney {9 [9 (,,) 9] and brief them to take action in the S
t up (,) and then basically I have to buy another one so I had to (,) go to the electrician {3 [3 (,) 3] But (,) uhm basically (,) electroni
he the the issue where Port R the earthquake {1 [1 (,) 1] you have to go to Port Royal {2 [2 (,) 2] So I'm saying that {3 [3 would 3] is it
Jonestown ((words)) I don't know (,,) inner city area is passing to go to Campion and {2 [2 (,) 2] two students actually {3 [3 on Campion
use (,) basically (,) my T V something was wrong with it and I had to go (,) to an electrician to {2 [2 fix it 2] because I don't have the s
ven field trip I remember as a child going ((word)) {2 [2 We had to go to Arawak museum we 2] {3 [3 learn about the Arawak 3] We had to go
to push yourself actually do something actually getting out of bed to go to the class (,,) $B [1 Mhm 1] 1} $B [2 Mhm 2] 2} $B [3 Your outlin
A Oh What are you going to do for the this weekend $B Uhm I wanted to go to the Quad {1 [1 (,) 1] but I'm not sure about that cos I I'm not

Appendix VIII: ***Concordances of come and***

Canada

Private

1 that you're gonna uhm you know write uh up this lease and then he can come and sign it and duh duh duh duh duh $A And then when he phoned ba
2 ow if I don't know if if uh if we can't go if just Helene and Sam can come and we can't fit them in the car (,,) I don't know if they'd be w
3 ves you like less than two weeks $A Ya $B Anyway the deadline's gonna come and go and then you're gonna be kind of disappointed later $C It
4 trollable (,) than this spra the spray that the commercial (,) people come { [(,)] and do $A [Mm hmm] } $B You ever watch them they spra
5 really gonna be able to do it or something But then other times she's come (,) and like (,) obviously everything's together y'know and she a
6 st that { [(,,)] you know I mean I mean if if you're the one that's come and (,,) try you know that that comes to it for the sake of dropp
7 hen maybe I'll get one of the stu like one of the student teachers to come and do the dialogue {3 [3 with me for example 3] $B [1 The introd
8 $A [1 ((word)) 1] 1} $A [2 Sure 2] 2} sure $B See if they wanted to come and (,) have supper with us and Or lunch or something (,) $A Mm h
9 I don't think I'll use it $B Ya $A I think I'll just ask uh Morris to come and get it (,,) remove remove it { [(,,) cos I] don't think we
10 xactly how much Uh (,) the thing is that (,) there will be people who come and will pay whatever you've got on there (,) Uh which is probabl
11 robably the majority of people But also there're a good chunk (,) who come and they wanna bargain So if you put it at a dollar they say oh f
12 d they'll not come So (,,) $A Well it it Mick and Jamie { [will (,,) come and] $B [Ya definitely and Jack and Maisy] } I think { [and]
13 [Right] } $B Well uh Harry who is twelve decided that (,) he would come and get his hair cut { [(,)] and uh then his dad decided he wan
14 not done it for a few years But this would be like a constant And you come and you go and and you use it to touch base and (,) it (,) might

Public

15 children from below Sherbrooke Street in the (,,) apartment complexes come and flock there after school It's like home you know You don't ge
16 ,) as long as the child was healthy (,) uhm they could (,) they could come and and (,) play and participate but the parent would have to obs
17 or three words)) $A Sorry I didn't mean to cut you off You You could come and go as you please though $B Yeah $A Nobody'd ever tell you wha
18 rience that we came across was uhm (,,) a child who was longing to to come and play in the playground $B { [Aw] $A [Uhm] } and so what w
19 g with the community We invited (,) uhm everyone on the street (,) to come (,) and uh meet with us And it was a somewhat antagonistic uh (,)
20 esting uh (,) uh get-togethers with the community and invited them to come and visit and we had an open house and they could come in and (,)
21 inning Uh they sent the uh I guess the the Quebec elections police to come and see us Uh from there they contacted Voyageur bus where Voyage
22 2] and an invitation (,) to (,) uh the parents that were there (,) to come (,) and participate (,) in {3 [3 uh the 3] next uh (,) uh school
23 committee on (,) the twelfth of June All E P P parents are invited to come and the reason for that was school-wide we wanted to (,) plant th

Unscripted

24 And smaller communities they're still (,) not actively encouraged to come and in some places they're still kind of (,) discouraged to leave
25 marks and they got fifty-five they failed That's it Sorry You cannot come and wash my car and buy me chocolates and take me out for dinner

26 t anyway Uhm (,) this is in the Book of Esther (,,) Uh he says to her come and talk to him and see if you can (,) talk him out of this It's
27 ng other activities We'll start working on other activities Then I'll come and ((word)) some other class with words missing you know the o
28 (,) uhm Kay so this is what you spend a lot of time playing You just come and and try out some of these effects uhh (,) ((sound effect in
29 children have very old parents Report card day comes and the parents come and meet you and and you think (,,) is this that woman's child ((
30 or $B Well the artistic marks should be much better (,) and here they come and they are much better (,) Up to a five five from the Hungarian
31 work (,,) And that's what you do (,,) If they have a question they'll come and ask you (,,) That's just for the third type (,,) Do you have
32 ich What a group we have seen (,) Browning (,) Eldredge (,) Bowman to come and right in the middle of it is Elvis Stojko (,) the youngster w
33 the other hand it's really hard to find people who are are willing to come and work in the school and do that (,) uhm So the people you get
34 ce the late seventies the very late seventies And so I've seen trends come and go and different attitudes and whatnot It i (,) In spite of m
35 just in case you do feel (,,) like you want to Cos all too often we come and we look at images and we intuitively ask the right questions
36 he crosses See what I mean (,,) Have most of you seen it Kay so we'll come and we'll look at it much more closely (,,) Uh well when we get u
37 uh Ray Kroc's secretary (,) in the times of high growth (,) he would come and he wouldn't be able to pay her weekly salary (,) because he t

Scripted

38 ld stay at her place She also asked if Miranda Fortino's mother could come and pick her up (,) but Miranda refused The same thing had happen

Non-Printed

39 ove is so strong. She does nothing and waits until the male lover can come and save her. He will come to her and she will " take his hand an
40 can because sometimes I'm away too. Good thing about volunteer jobs; come and go without asking. I always go on my arranged-for day when I'
41 ham , add hundreds of millions of words. You are certainly welcome to come and use the resources of the Language Unit. Yours sincerely, Jenn

Printed

42 ough change, by focusing on elements old and new. Man-made structures come and go in relatively rapid succession. A Halifax city block is sl
43 tly it was time to let it all go; phoned my husband in Cape Breton to come and take my child, and drove me to the Emergency at the Halifax I

GB

Public

1 n (,) uhm { [conflict] And you know it's seen it over the centuries come and go and alas it still sees it today $A [Uhm] } $A Indeed ind
2 in the business a good number of years Haven't you seen these crises come and go before Isn't this an age-old story $C The arts are always
3 (,) uhm meeting in the village hall he simply said uh uh let Kingdom come and my name perish {2 [2 ((laughs)) 2] $A [1 ((laughs)) 1] 1}
4 ow a lot of bridges between let's say Baghdad where the main supplies come and to the south where the Republican Guard are there are thirty-

5 there in London both of home grown Londoners and of the tourists that come and they are being denied access to the arts And that for me is t
6 nly re uh uh hope that they will vote for us we want them actually to come and join us work with us and help shape our policy That is the po
7 more effort more more of an effort to actually encourage students to come and to uh fill the places if we're going to in increase it $A D'
8 e of like-minded people in a rehearsal room whereby they will want to come and and commit themselves to it and they will be prepared if you
9 Uhm we should have a V P where N (,) Sorry Yeah] } $A Why don't you come and draw it because it's much easier than waving your hands which

India

Unscripted

1 radition also (,) though we have migrated from our (,,) districts and come and settled down here Now we've become a part of this place (,,)
2 imatise to keep the head to the ground then easily (,) the confidence come and they can practise shirsana (,) And this is a soothing (,,) bo
3 bring up (,,) our (,,) people (,,) or (,,) the society (,,) which has come and settle down here (,) become a part of Bombay (,) they will de
4 good event (,) that (,,) this coach (,,) as the British teacher have come (,) and (,,) she feels thankful to the (,) arrival of these two t
5 rs have been short finale (,) So he all (,) the customer has to do is come and pick up the application (,) fill up all this (,) and if he fu
6 the house (,,) Then uhm (,,) some people (,,) so (,) some people just come and ask me to paint things (,,) on any object (,,) They bring eit
7 y could (,) the would possibles have to prove (,) the watchman should come and make a statement before the police officer or before the cour
8 morning (,) I think probably the school has given them the day off to come and watch the match (,,) There the hills behind I think Chandigar
9 t time (,) but he didn't (,) Our Gudi was there (,,) I forced Gudi to come and join this institution He was not willing to come (,,) because
10 (,) tell (,,) now like you are there (,) or you are there (,) now to come and tell them (,,) look I don't mind marrying a guy who is an Aut
11 enewed (,,) For the film buffs of course (,) this is the time to (,,) come and meet (,,) some of their favourite actors (,) and singers (,,)
12 to (,) give a written argument (,,) or if he is not (,) then he will come and argue himself (,,) and at the most it will take another two o
13 ou what you are getting (,,) but we have a soft corner for you if you come and sacrifice (,,) and ultimately I had to cut a joke at him (,,)
14 I'd like to uh uh (,) interview her father in fact (,,) Why don't you come and share your views with us (,,) ? We'll do that a bit later bec

Jamaica

Private

1 off the] } recorder cos she coming in here to be bad ((laughs)) $A Come and be bad $B Shouldn't they say they want to interview you comin
2 ee months yeah You can't there is no ambience like this where you can come and sit to have a couple of drinks and go home with your friends
3 re and him come and say you'll take care and make sure that she don't come and bother cos they were afraid {6 [6 because 6] it was bad enoug
4 e the scholarship to pay back the student loan {5 [5 (,) 5] But don't come and depend on the scholarship and if you don't get it you're goin
5 le coming here showing their I D anyway Is not possible for everybody come and show you {1 [1 I D 1] them a go give you bare attitude and st
6 Lord a boat {1 [1 (,) 1] is a a big cook-up you know where everybody come and they they bring something and you just throw it into two pots
7 $A You you you have one more semester left $B Yeah $A What you gonna come and do You going to do five courses $B Mhm $A Why $B Five $A Mhm
8 peare had a brother like named Hubert Shakespeare somebody would have come and said This my origin just as well as his brother you know {2 [
9 ppened to uhm and he c and he felt her and he came down there and him come and say you'll take care and make sure that she don't come and bo
10 them cannot find him And then uh also around maybe one year later him come and kill off all of the all of the family too You know she said s
11 hey use those things to to decide on on your future So you can't just come and say alright I don't have the time or the lecturer not importa
12 accommodate so many managers and leaders you understand and they just come and tell you we're university with no experience but how possible

I want/won't tell about the student loan thing well I guess you know come and they might fund some bank loan that them offering (,) to go b
age] } $A Mhm By yourself $B No yeah by myself by myself come but me come and stay with me mother {1 [1 (,) 1] and then with my mother and
bout big sister big {1 [1 brother The little sister think 1] she must come and this guy was saying that (,) if he was like at the time of Wi
uhm W B sh it's the longest one of {1 [1 the W B shows (,) 1] Others come and go {2 [2 and it's still there 2] $B [1 W B show I have to say
rs just say them students Alright You don't wanna say uhm uhm parents come and talking to me sit down They're playing they're drawing things
y keep you awake all night {3 [3 (,) 3] And as (,) soon as the police come and and they turn it down and as soon as the police drive off the
s killing her and the Master Anyway me no know a few months later she come and she say How come you didn't tell me you were doing a Master's
[Mhm] } $A ((laughs)) Kedisha is the quietest one in my room She come and she o knock on the door and she come and she sit down and she
t right beside me so she come and she realized what was happening she come and she cook my chicken and all and you know It was nice after th
quietest one in my room She come and she o knock on the door and she come and she sit down and she read No $B No Lauri $A ((words)) you k
(,) Because she was always in the kitchen just right beside me so she come and she realized what was happening she come and she cook my chic
aican English $C Yeah $Z $B In Microsoft Word $C There are spe Simone come and sit here darling $D Why do you want me to sit there $C Becaus
hs)) that him him annoyed with him and everything So she uh suddenly come and say him say oh it's okay We love honey honeymoon (,) Come to
ida International University {6 [6 and all 6] these universities that come and set up $A [1 Mhm 1] 1} $A [2 Mhm 2] 2} $A [3 Yeah yes 3] 3} $
pe and stuff {3 [3 that one 3] Or I watch sports in the evenings they come and watch {4 [4 (,) the game 4] $A [1 It's like a 1] 1} $A [2 Oka
d and uh everybody takes this as an indication of g of good things to come and I I am really really l really really looking forward to the r
risk anything and my mother isn't somebody who you know is going to come and quarrel with {1 [1 anybody 1] or you know {2 [2 but 2] $A [1
the Jamaica Labour Party for uhm Mr Seaga's sake and for Mr Seaga to come and say publicly that uhm he doesn't think that there is anyone i
talk to people {4 [4 (,) 4] and that they were willing for people to come and talk to them $B [1 Mhm 1] 1} $B [2 Yes (,) Thanks very much f
the back of the bus pushing it Can you imagine that And we ask him to come and help us Him don't do bus I was most upset Anyway all of that
with them (,) you know Well we question them cause some of them will come and by talking to them (,) you find out that things are not nice
will come and cos they fear humans too because humans sometimes will come and kill them off because them believe all vampire all lots of us
where you are you have to get used to the fact that good friends will come and they will at one point leave {2 [2 (,) 2] leave you alone {3
it not working and the university inefficient but I think better will come and I I think this university will be an exciting place of the fu
rinthians who are like the saviours of their people So like they will come and cos they fear humans too because humans sometimes will come a
ssible We give them sedation {2 [2 (,) you know 2] Not that you won't come and hear them screaming you know (($A-laughs)) We're a third wo
of what is happening outside of that So what you find is that I would come and do the same thing {5 [5 that they advise 5] me to and if some

Unscripted

quite it (,,) Even with all of that when I'm told two three weeks ago come and talk about English it really I I I felt a little bit (,) stun
he areas that you can come and read (,) Okay One of the areas you can come and read and that is where you collect your books If you request
is the R B C reading room area This is one of the areas that you can come and read (,) Okay One of the areas you can come and read and that
ving seen the flag {1 [1 because they 1] wanted to know if they could come and get a reading ((laughs)) $B [1 Okay 1] 1} $Z $A So so it it
ere you purchase the copy card So this is one section where you could come and copy the book But in case you don't want to copy the book you
ick tour {2 [2 (,,) before 2] he gonna speak with the class So I just come and introduce him to you (,) And he has any question him have for
the Jimmy Cliff men they were ((word)) sitting in Limbo Harder will come and all those music now (,) just became (,) the soundtrack And wh
tion But we are grateful (,) that they thought it important enough to come and give us some of our time (,) And (,) I want to explain why Pr
d won't be able to be with us (,) So he asked the deputy principal to come and I haven't seen her as yet (,) Uh I know that there is a very
p Harbour View's approach to this game They were mentally prepared to come and take on any physical challenges and I think the momentum of t
n it is the single most important aspect of his life (,) So for me to come and say boy rasta a idiot I am putting myself now in right into t
lty but we have been still waiting for that still waiting for that to come and I $Z $Z $Z $A [Yeah] } $A I'm not sure uhm uhm the way I un (,,)
think the experience is to come and see it for yourself looks a really beautiful place (,,) Now we
I give him the tour but I just told Mrs Cocker so that he's going to come and finish we come and sit with the upper school fi him talk to S
$E [3 Mhm 3] 3} $E And there has been some interests in the media to come { [and see the Junior Centre] as to carry their performances to
s success Our message to the consumer our message to the consumer was come and experience the real Jamaica Meet our people explore our cultu
r but I just told Mrs Cocker so that he's going to come and finish we come and sit with the upper school fi him talk to So when he's finish
kay Well this is the section if they'll R B C books this is where you come and you borrow the books to take out Alright ((long- break-with-
in You enter This if you need (,) open shelf books this is where you come and get the books to go out If you are borrowing the books (,) Op

Appendix IX: ***Concordances of and go***

Canada

Private

1 tell Hamish again too but they should disappear at twelve thirty and go and get changed (,) So that people it's a signal to the (,) you kno
2 ike less than two weeks $A Ya $B Anyway the deadline's gonna come and go and then you're gonna be kind of disappointed later $C It doesn't h
3 ere we'd (,) stop and go to the fish market Pick up some lobsters and go back to the Big Lobster and and eat it $A Mhh { [hmm] $B [Eat]
4 ,) person or you know (,) a bunch of people that who get together and go do B and E's you know It was like this is a huge vast network of cr
5 e'll uhm sort of find a way to separate himself from (,) the kids and go he'll have a chat with her at times $A Oh well that's good {1 [1 Th
6 words))] $B [((laugh))] } $A { [((words)) gonna stand up and go hey that's not true I feel guilty as heck like whoops ((laugh))]
7 ther people $B ((laugh)) $A Well $B { [Just wander up] to you and go hey what are you doing $A [((laugh))] } $B It's a baboon behind
8 $A [((laugh))] } $B Probably uh stay ((word)) (,) camping and go hiking Or my dad had a good idea {1 [1 (,,) we should 1] get into c
9 oh Barry she's such a jerk $B Ya ((laugh)) And of course I call and go oh god I called you again I was calling my dad And Barry goes oh we
10 exist (,) but it's not mostly (,) a piece of art that you look at and go oh isn't that beautiful { [and] $B [Although] } Although there'
11 gram (,,) $B [Program for you] } You're really going to go over and go out and sweat $A I don't know ((laugh)) Well I if I'm going to jo
12 it's uh seem to to feel that it's a necessity to leave from work and go to a club { [or whatever] and then come home (,) later $A [Mhh h
13 ice cream $B Mhh $A And after that uh (,,) more playing the piano and go to bed and up to (,) Montreal again this morning $B Hmm $A So I mad
14 s $A [Okay] } $B But they used to go to the port the Stork Club and go to dances and then go back to London afterwards And they could drin
15 our lives changing that much except we may have an apartment (,) and go to the $B Do and can't wait to get there (,) and then come home bec
16 [there] $A [Yes] } $B Pretty little town Where we'd (,) stop and go to the fish market Pick up some lobsters and go back to the Big Lob
17 with you (,) if you want So you can just go up like one guitarist and go up alone and have the whole band back him $B Oh { [wow] $A So but
18 nd the other (,) {1 [1 side where 1] it's not quite so steep and (,,) go up that {2 [2 path 2] $B [1 Mhh uh 1] 1} $B [2 No 2] 2} I've just b
19 uhm $A I'll just put that over here $B I'll have to take Katy (,) and go up to the General ((phrase)) the only place to park but I'll just
20 ell her the same thing ((laugh)) And then she'll agree with him and go with whatever (,) you know she's been told before by him anyway so

GB

Private

1 derneath the Royal Oak and the big chute affair and then come out and go down in the gutter [5 (,) {5 5] And the ducks used to come out ther
2 use it 's a nice chord uhm but I'll (,) maybe we can establish it and go away from it [(,) {] But uh we'll see about that OK (,) Anybody g
3 1 Really was a great success You often sort of pull up your roots and go away 1] in the best part of the summer and {2 [2 you miss the garde
4 going to snitch or I've already taken down out down the West End and go and see (( )) something fantastic $A I am taking her out down the W

Public

5 { [conflict] And you know it's seen it over the centuries come and go and alas it still sees it today $A [Uhm] } $A Indeed indeed But u

6 ve to proceed with the pregnancy and uh get it over and done with and go and start again (,) on the next one like a baby machine (,,) $A But
7 do you think the answer is to def define it for one's own purpose and go and use that $B What frightens me also is defining it for your own
8 Well let's leave that elegiac note on which the curtain came down and go back if we may to the beginning Roy Jenkins could I ask you in nine
9 usiness a good number of years Haven't you seen these crises come and go before Isn't this an age-old story $C The arts are always in crisis
10 you had to force yourself for every minute of it to go on writing and go on working (,) And on other days it was coming and you didn't want

Unscripted

11 ot a single cohort more than is absolutely necessary should leave and go into employment without training which is sufficient for their work
12 es one though even that has seen some large corporations (,) come and go since since since its founding Uh (,) ((clears-throat)) and one s
13 us about the cycle of vibration how the vocal folds come together and go apart (,) And I'll show you some examples from four different voice
14 ,) Uhm but uh it's there's such a close uhm similarity Let me try and go back on the (,) right to convince you uhm that so there can be no d
15 of a woman's second age (,) is to (,) watch the children grow up and go away (,,) ending (,) the second experience from her point of view (
16 example when a tree dies the nutrients in it rot away uhm (,) and (,) go back into the soil and provide the uh (,) ((clears-throat)) the f
17 ait (,) And only when it's clear look behind with a lifesaver (,) and go And once you're past the parked vehicle pull over quickly to your s
18 st (,) And I thought oh I'm going to get out of here I'll go and (,,) go and find a hotel (,) And I was reluctant to leave the Old City cos
19 t the beneficial effect that more and more healthy babies survive and go through childhood (,) But this tends to be at a at a cost And the c

Scripted

20 its advantages in an intensely volatile region where regimes come and go even if in the end time cannot halt the flow of advanced knowledge
21 our best suit on and you and Cranley Onslow and th a third person and go in and say Margaret we've sen sed the Party's opinion We've come to
22 en hit (,) All citizens were instructed to put on their gas masks and go into sealed rooms as a protection against chemical weapon attack (,

Non-Printed

23 months alone . If I get this I will buy them an answering machine and go to the beach . Anyhow I'm now trying to organize my own placement w

Printed

24 ity of her situation . The girls all knew that they could pack up and go any time , that they had money enough to cope , and realised too th
25 l placating Dr Carson , reluctant to end one awkward conversation and go back to another . Even after replacing the receiver , she hesitated
26 , claim your Retirement Pension , and then cancel your retirement and go back to work . You cannot get Unemployment Benefit : after the age
27 ket . Keep volleying well out in front of your body . Be positive and go to meet the ball . Remember : volleying is fun . VOLLEY VARIATIONS

India

Printed

1 're going to watch a beautiful sunset. Then I'll pick up Kristine and go dancing at the country club." " Sure you're ready for it?" Shannon
2 is an unreasonable request. How can he leave all his urgent work and go for a stroll beside the Ganga with this chit of a girl? But Monolin

e is in jail. Keep away from the big cities, boy. Earn your money and go home." " I'll do that, uncle. My mother and sister will expect me t
f while life lasts. Poet Gray said ` Eat drink, be merry yourself and go the deuce - - if there be a deuce" . He was not sure that the it wa
s or assign causes to the events described that they part company and go their opposite ways. Omar denies human responsibility for this wret
en a tough decision for him when he had to leave his mother alone and go to the USA for specialisation in neuro-surgery. " I will stay here

Jamaica

Private

they-both-laugh)) $B Oh my God Yeah So that's I usually do that And go visit some friends who live in the area you know { [bec] $A [So
{1 [1 him him go the dir 1] (,) It's easier to go that direction and go up in the hill {2 [2 (,) 2] than to go Red Hills {3 [3 (,) 3] proba
sit with them in meetings et cetera So you can't just come here and go to your classes and not do anything You have to be involved especia
class and thereafter you just find small items just to go to him and go to You don't have to go every week { [(,)] He has uhm the lec lec
een applying for jobs and so on $B Yeah I've started that process and go to ((words)) What else German embassy was really rude They sent m
laughter)) and the doctor look about it and sound me everything and go to the pharmacy we buy medic alright ((laughter)) yeah and everyt
5} $A [6 Alright mhm 6] 6} $A [7 But then to 7] 7} leave Spanish and go to I R it's not easy though $B I realized because {1 [1 I started d
or you for us to teach you (,) but usually you have to leave here and go to another institution although and do your masters (,) to basicall
I'm thinking hey maybe that's an option maybe I should just leave and go study (,) and then later on maybe after doing the P h D three years
are taught that they are the providers So guys will leave school and go straight into a job start working cos you see a guy picks up a girl
situation thing So that was one thing Let's change the topic now and go on to something else ((laughs)) Yes so hm the point I was making
[7 action 7] {8 [8 don't it 8] And when she did it and she go on and go on and she was in the wrong ((words)) {9 [9 (,) 9] So everybody s
) I'm supposed to be packing up my stuff And I probably watch T V and go look for my friend and get another book to read ((laughter)) Not
lasses on that day or we had a class in the morning and then left and go into the production After that I left in ninety-seven and worked fo
account for these people being able just to walk right up and say and go in to to Red House and to T T T and and establish themselves How do
e like this where you can come and sit to have a couple of drinks and go home with your friends yeah uhm There are pubs yes but it's not the
s] knee (,,) giving him problem and him say I could not stay (,) and go home at the time $A [Mhm] } $A Mhm okay by_ _ hisfeet b
't just say that we're going to to tie up the whole of this thing and go home and forget that it took place and you go back to your commune
(,) 1} You have that but my girl she take her shoes off She go on and go flip flop {2 [2 (,) 2] ((words)) and I thought and I I was this i
Uhm you're up all day into the night Five o'clock I leave campus and go down to run a race at five o'clock and then you come back up for sc
have to stay here you have to stay in here I said take the child and go Because if this is your attitude other parents there's about twenty
ne { [(,)] Well (,) one would need more detail Did it just turn and go back to Tripoli or did it land somewhere else Was it really en rout
scholarship and if you don't get it you're going have to pack up and go back home You have to plan ahead Apply if you're expecting money fr
ds (,) right around and past the picket fence where the lines are and go around (,) and then you see it on your right hand $B Do they have a
[2 you want to go to all over Yes I see 2] $A [1 Yeah 1] 1} $A [2 And go all over All over It you know 2] 2} I wonder for where we living {
being the stubborn person I am always like trying to go in there and go {3 [3 out and 3] stuff ((((words)) But far apart from that still
$A Oh yeah for true {1 [1 yeah 1] {2 [2 And you didn't 2] get up and go {3 [3 Mhm 3] $B [1 Right 1] 1} $B [2 And 2] 2} $B [3 And then 3] 3}
g I said your child will not be accepted in school Take the child and go {3 [3 (,,) So 3] she she said uhm uhm I'm not taking the child to w
sh it's the longest one of {1 [1 the W B shows (,) 1] Others come and go {2 [2 and it's still there 2] $B [1 W B show I have to say cos 1] 1

Appendix X: ***Co-selection components of the* come-** ***and* go-*grams.***

Node	***Collocation***	***Colligation***	***Structural Preference***	***Semantic Preference***	***Discourse Preference***	***Semantic Force***
come (deictic)				locations specific and non-specific to the participants		movement associated with the participants or towards their location
go (deictic)				locations specific and non-specific to the participants		movement from a specific location towards a non-specific location
Post Pre-set						
come and		post verb	surplus			movement to undertake an action that will happen after the utterance with emphasis on the participant(s) being physically present in a place specific to the participants
		post verb	surplus	post social interaction		movement to undertake a social action that will happen after the utterance with emphasis on the participant(s) being physically present in a place specific to the participant(s)
	ante *to*	post verb	surplus	post social interaction Ante (not) want/desire		desire for movement to undertake a social action that will happen after the utterance with emphasis on the participant(s) being physically present in a place specific to the participant(s)
go and		post verb	surplus			movement to undertake and amplification of the following action

Node	***Collocation***	***Colligation***	***Structural Preference***	***Semantic Preference***	***Discourse Preference***	***Semantic Force***
come back			list	previous/usual/routine location specific point in time		movement to return to a previous/usual/routine state or place at a specific time
go back				vague/non-specific location vague/non-specific point in time		a return to somewhere or something that is slightly vague in scale
come from				roots, home	ICE-India	indication of movement from an original location
				source rather than location		indication of making something better or worse for whoever, or whatever moves from source
				roots	ICE-GB	it is what one contributes not where one comes from that counts
come in		[*in*] adverbial or adprep		specific location		movement inside a specific location
go in		[*in*] adverbial or adprep		non-specific location		movement inside a non-specific location
come in			elliptic (no second *in*)	objects/types of objects		placing the objects into type or orders of types
go in	*there*			non-specific location		emphasis of movement inside a non-specific location
go into				something important about the place/time/state, or and important event		movement into somewhere important

Node	***Collocation***	***Colligation***	***Structural Preference***	***Semantic Preference***	***Discourse Preference***	***Semantic Force***
go into				reluctance trivial place/time/state		movement into somewhere trivial or unimportant
come on			post qualification	a correction to/reformulation of the previous utterance	dialogue spoken/written	exhortation to other participant(s) to re-think, or think like them with a requirement for evidence
			backchannel		dialogue spoken/written	you don't say
go on			noun phrase	non-specific location		movement to or from a non-specific location
		adverb or preposition				continue
			verb phrase or nothing			continue
come out				positive outcome non-physical movement		positive change in state
go out				further significant action		movement from a location to undertake something significant
go through				resultant neutral or negative state		change from one state to another that is at the very least neutral but more likely negative
come to			post noun phrase	end state		movement towards an end state
		post verb				progression towards a time something will occur

Node	***Collocation***	***Colligation***	***Structural Preference***	***Semantic Preference***	***Discourse Preference***	***Semantic Force***
	post *know*				ICE-India spoken	progression towards a time when something is known
go to			post noun phrase	location non-specific to participant(s)		physical movement to a location non-specific to participant(s)
go to the		post noun		location non-specific to participant(s)		physical movement to a location non-specific to participants
come up				class/gender	ICE-India	movement that produces advancement of class or gender
	post *with*			trying better plans, schemes, ideas		attempted movement that, and this is by no means guaranteed, creates a change of state for the better
Ante Pre-set						
and go	*leave*			leave		depart from/leave behind in order to move into something new and unrelated – an emphasis of end before a new beginning
they come			eg: post pre-set collocates	*as post pre-set collocates*		*as post pre-set collocate*
they go			eg: post pre-set collocates	*as post pre-set collocates*		*as post pre-set collocates*
to come, to go	*want*			ante (not) want or (not) desire		(no) desire for movement
				ante measurement of time/distance		time or distance remaining before state is attained

Node	***Collocation***	***Colligation***	***Structural Preference***	***Semantic Preference***	***Discourse Preference***	***Semantic Force***
have to go				necessity, obligation problem/difficulty		obligated movement that has the potential to be difficult
you come				[*you*] the other participant or participants of the exchange		movement of the other participant or participants of the exchange
you go				[*you*] a random person or persons that could exist somewhere		movement of a random person or persons
Ante and Post Pre-set						
to go to	*have*		post noun phrase	location non-specific to participant(s) necessity, obligation problem/difficulty		obligation with some problem/difficulty to move towards a location that is non-specific to the participants
	want		post noun phrase	location non-specific to participant(s) ante (not) want or (not) desire		desire for physical movement to a location non-specific to participant(s)
come and go				time and/or cycles unhindered		unhindered repetitious movement from a non-specific location/state to a specific and back to a non-specific location/state over a time period
			repeated pronouns eg: *you come and you go*	time and/or cycles unhindered		unhindered repetitious movement from a non-specific location/state to a specific and back to a non-specific location/state over a time period
Discourse Managing						

Node	***Collocation***	***Colligation***	***Structural Preference***	***Semantic Preference***	***Discourse Preference***	***Semantic Force***
come back, come in, come on, come to, come up				specific subject/topic in the discourse and time (often in the near future)	spoken language	indication that the discourse will move at a specific time to specific subject/topic
go back, I go, we go				non-specific subject/topic in the discourse and time	spoken	an indication that the discourse will move at a non-specific time to a non-specific subject/topic
go back, I go			hesitation	specific subject/topic in the discourse and time (often in the near future)	spoken	a non-assertive indication that the discourse will move at a specific time to specific subject/topic
Sports' Reporting						
come				winning position sporting terms	live radio	movement towards winning location
go				winning position sporting terms	live radio	movement away from winning location
Speech/thought Replacement						
and go, I go, they go, we go, you go		post discourse markers	post reported speech or thought		ICE-Canada	this is what was said or thought