

Academic lexical bundles: How are they changing?

1. Introduction

One of the most noticeable features of academic written texts is the presence of multi-word expressions referred to as clusters, chunks or bundles. While perhaps not strictly formulaic by Wray's (2002) definition, which makes a claim about how sequences are stored in the mental lexicon, these strings are nevertheless 'glued together' in everyday discourse. Simply, 'lexical bundles' are statistically the most frequent "recurring sequences" of words in any collection of texts: extended collocations which appear more repeatedly than expected by chance across a given range of texts (Biber et al. 1999: 990). They are identified automatically using corpus analysis software which retrieves multi-word units with frequency and distribution criteria specified by the researcher. As a result they are neither idiomatic nor, usually, complete grammatical units (Biber 2006), containing strings such as *it was found that* and *in the case of*. They are familiar to text users and have customary pragmatic or discursal functions.

From a psycholinguistic perspective lexical bundles are seen to have "a processing advantage over creatively generated language" for speakers (Conklin & Schmitt 2008: 72) while their role in identifying a text as belonging to particular register or discipline is now well established (Biber, Conrad & Cortes 2004, Chen & Baker 2010, Hyland 2008a & b). It is also the case that bundles have been found to be conspicuous in other languages, such as Spanish (e.g. Butler 1998) and Chinese (e.g. Ma 2009, Wei 2007), and that they lead to improvements in L2 students' writing in English when part of an instructional programme (e.g. Meunier & Granger 2008). But while most of the research into lexical bundles explores English and focuses on academic discourse, we know almost nothing of how they have changed in research writing in the past 50 years. In this paper we address this question. We focus on published research writing as both the most ubiquitous and conventional form of academic discourse and the area which may have the greatest pay-off for junior scholars seeking frequent phrases to make their writing recognisably scholarly. We do so by analyzing a corpus of 2.2 million words compiled from articles in the top journals in four disciplines. Our aim is to discover whether, and to what extent, lexical bundles have changed in their frequencies, forms and functions in different disciplines between 1965 and 2015.

2. Lexical bundles and academic writing

The importance of frequent multiword combinations as a way of assisting communication has long been acknowledged in applied linguistics (e.g. Firth 1951). By making language more predictable to the hearer/reader they function as processing short-cuts, retrieved from memory as chunks rather than generated anew on each occasion (Wray & Perkins 2000). Routinely employed sequences therefore work to facilitate pragmatically efficient communication, and in academic discourse help to reduce processing time by using familiar patterns to structure a discourse by guiding readers through it (*in the next section, we can see that*) or by linking ideas (*is due to the, in contrast to*). In addition, by signalling appropriate use of a disciplinary code, they allow writers to display solidarity with colleagues (Cortes 2006) and to construct a disciplinary competent voice (Hyland 2008a, Pang 2010).

Lexical bundles therefore seem to reflect a very real part of users' communicative experiences. As suggested by Sinclair's (1991) 'idiom principle', there is a phraseological tendency in language use whereby speakers and writers co-select words in routine ways. Sentences are typically made up of interlocking bundles as words are mentally 'primed' for use with other words through our experience of them in frequent associations (Hoey 2005). Everything we know about a word is a result of our encounters with it, so that when we formulate what we want to say, the wordings we choose are shaped by the way we regularly come across them in similar texts. Nor is this a modern phenomenon as the importance of highly fixed and repetitive lexical strings has also been confirmed in historical texts (e.g. Culpeper & Kytö 2002, Kopaczyk 2013). Needless to say, these lexical bundles are pervasive in academic language use and a key component of fluency, marking out novice and expert use in both spoken and written contexts.

In fact, corpus research has identified recurrent patterns which occur significantly more frequently in academic than in other, non-academic registers (Biber et al. 1999, Simpson-Vlach & Ellis 2010). Academic writing, for instance, draws on a much larger stock of prefabricated phrases than either news or fiction in the BNC Baby edition, with over 450 different 4-word clusters occurring more than 10 times in one million words (Hyland 2008a) while Biber et al (1999: 994) suggest that 4-word bundles occur over 5,000 times per million words in academic prose. Within academic contexts these bundles also differ across modes (Biber 2006, Biber, Conrad & Cortes 2004), so that the Academic Formulas List shows completely different items for its spoken and written corpora (Simpson-Vlach & Ellis 2010). They also differ across genre, with classroom teaching containing far more, and a far greater variety of, different bundles than textbooks and academic prose (Biber, Conrad & Cortes 2004, Biber & Barbieri 2007) and

dialogic speech events like tutorials and class discussions using completely different and twice as many lexical bundles as monologic genres such as lectures (Simpson 2004).

This variation is repeated across written genres so that Chen and Baker (2010), Cortes (2004) and Hyland (2008a) have found considerable differences in bundle use in published academic writing and student texts with many high frequency bundles frequently never used by students at all. Lexical bundles also appear to be a strong marker of discipline. Cortes's (2004, 2006) analysis of published and students' academic writings in the fields of biology and history showed large variations while in Hyland's (2008b) study Electrical engineering texts contained both the highest frequency and the greatest range of 4-word bundles compared with biology, applied linguistics and business studies. There was also considerable disciplinary specificity in the 4-word bundles themselves with just four items occurring in all four disciplines in the thirty most commonly used bundles in the four fields.

The above review of the literature has shown that bundle use reflects something of the argument patterns and purposes of speakers and writers, indicating different modes, registers, genres, disciplines and changing contexts. But while it may appear that we know a great deal about lexical bundles in academic writing, almost nothing has been written about how they have changed in recent years (see Kopaczyk 2012 for a review). This is an important question if we wish to understand both the role and frequency of bundles in disciplinary writing and how their use may have responded to contextual changes in research and publishing. This is the issue we address here. First, we describe our corpus and methods and then go on to discuss the grammatical and functional changes which have characterised published research writing in recent times.

3. Corpus and methodology

While formulaic sequences have the advantage of being methodologically clear-cut, identified solely on frequency of occurrence and breadth of use, researchers have used different frequency and distribution criteria. The threshold frequency, which determines the number of bundles to be included in the list, has ranged from 10 (Biber et al. 1999, Biber 2006) through 20 (Cortes 2004, Hyland 2008a & b) to 40 times per million words (Biber, Conrad & Cortes 2004), and even raw frequencies (Chen & Baker 2010). A second identification criterion is that sequences have to occur in a specified number of files in the corpus, such as 3-5 texts (e.g. Biber & Barbieri 2007) or 10% of texts (Hyland 2008a) to avoid the quirks of individual speakers or writers. Analysts must also decide on the length of strings they select. 2-word bundles are extremely common, and are therefore less useful in diachronic corpus research (Kopaczyk 2012) while 5- and 6-grams are comparatively rare and often subsume shorter ones. Four-word bundles seem

to be most often studied, perhaps because they are over 10 times more frequent than 5-word sequences and offer a wider variety of structures and functions to analyze (e.g. Biber et al. 1999).

For the present study, we followed the convention of identifying bundles which meet specified cut-off points for frequency and range (e.g. Biber et al. 1999, Hyland 2008a), while taking a conservative approach by following Hyland (2008a & b) and Cortes (2004) in setting a relatively high frequency cut-off of 20 times per million words. Additionally, we included only those lexical bundles which occurred in at least 10% of texts. A percentage offers a consistent distribution criterion and reduces the possibility that the identified bundles occurring in texts from a single discipline. We also decided to focus on four-word bundles due to their frequency, their variety and their probable prevalence across time. We manually excluded bundles with text-dependent noun phrases (e.g. *in the United States*) and removed overlapping word sequences where two 4-word bundles are actually part of a 5-word string (e.g. *it has been suggested* and *has been suggested that*) (Chen & Baker 2010).

We constructed the corpus from three corpora of research articles taken from the same five journals in four disciplines spaced at three periods over the past 50 years: 1965, 1985 and 2015. We used online versions of the articles and stripped out appendices and reference lists. The different time spans (20 years + 30 years) were chosen in case we wanted to focus on the later period and to see if changes were more pronounced then, although we were concerned with overall changes during the 50 years. We also chose to use 1985 as a mid-point because this seemed to be on the cusp of the move to electronic academic publishing and a turning point at which applied linguistics, sociology and biology recorded a huge increase in interactive markers after a sharp drop (Hyland & Jiang 2016a & b).

Applied linguistics, sociology, electrical engineering and biology were selected as representative of soft and hard science fields, as well as established and emergent fields. We were, then, interested in how disciplines of very different characteristics may have changed their use of lexical bundles. We took six papers at random from each of the five longest-running journals which had achieved the highest ranking in their disciplinary category according to the 5-year impact factor in 2015. That is, 30 articles in total from each discipline from each year. The journals are listed in Appendix 1 and the corpus comprised 360 papers of 2.2 million words (see Table 1).

Table 1: Corpus characteristics

| Discipline | 1965 | 1985 | 2015 | Overall |
|---------------------|-------------|-------------|-------------|----------------|
| Applied linguistics | 110,832 | 144,859 | 237,452 | 493,143 |
| Biology | 244,706 | 263,465 | 237,998 | 746,169 |

| | | | | |
|---------------|----------------|----------------|----------------|------------------|
| Engineering | 92,062 | 97,545, | 235,681 | 425,288 |
| Sociology | 149,788 | 196,232 | 262,203 | 608,223 |
| Totals | 597,388 | 702,101 | 973,334 | 2,272,823 |

Such a sampling process is not without difficulties as the top ranked journals today may not have existed 50 years ago and those that have survived the last half a century may not be among the highest ranked now. Journals come and go, they fragment and become more specialised, they rise and fall in popularity, and they are replaced by new ones. Our selection may not reflect the major journals in the field now and that there is a certain heterogeneity in the choice of journals as a result of our sampling. The decision to include the same journals, however, ensures a certain internal coherence in bundle use and despite variations, these are all robust periodicals with a long history at the top of their respective fields and, we believe, represent both the consistency and diversity of disciplinary practices.

Having created the corpora, we searched for 4-word bundles using AntConc (Anthony 2014) following our criteria of strings with frequencies of at least 20 cases per million words and occurring in at least 10% of texts. The results were transferred into an Excel file where we coded each example for its function and grammatical structure (see Table 4 and Table 7 below). The two authors worked independently to code a 10% sample, refining agreement through successive passes to achieve an inter-rater reliability of 99% (structure) and 97% (functions). We should acknowledge that a single bundle may have multiple functions even in a single occurrence. For example, bundles such as *the beginning of the* and *at the end of* can function as referring to a time or place (*at the beginning of the study we were undecided*) or organising the discourse (*at the beginning of the following section we define our terms*). Most bundles have a primary function but we examined all potentially multi-functional bundles to determine this. When comparing bundles across the corpora, we followed Hyland (2008a & b) and Salazar (2014)'s practice of using log-Likelihood tests to determine differences of statistical significance.

4. Changing bundle use: a quantitative overview

We identified 6798 bundle tokens when we searched the sub-corpora in each year using our criteria, averaging 18.9 cases per paper. These comprised 82 different types in the 2015 corpus with the most frequent being *on the other hand* closely followed by *in the case of*. The analysis suggests that lexical bundles in academic writing, overwhelmingly, include parts of prepositional or noun phrases and that they relate to the discourse itself rather than to referential content or the participants. Table 2, however, shows while both the range of bundles (types) and their frequency (tokens) have risen over the last 50 years, variation in their use has declined (types/tokens) and so has the frequency with which bundles occur as a proportion of total words.

Table 2 Frequency of 4-word lexical bundles over time

| | 1965 | 1985 | 2015 | % Change |
|------------------|-------------|-------------|-------------|-----------------|
| Bundles Types | 78 | 70 | 82 | 5.1 |
| Bundles Tokens | 2239 | 1842 | 2717 | 21.3 |
| Types/Tokens | 3.5 | 3.8 | 3.0 | -14.3 |
| % of total words | 0.4 | 0.3 | 0.3 | -25 |

There was a significant increase in the number of tokens over the 50-year span (*log Likelihood* = 678.51, $p < 0.001$), but the marked drop in the type/token ratio indicates that new bundles are not created endlessly and there is some consolidation of use. It may also reflect an increase in the use of other, less formulaic, ways of expressing the same rhetorical functions suggesting that academic prose is perhaps becoming less rigid in structure¹. Whatever the reason, there may be less reliance on prefabricated language overall, as shown in the fall of bundles as a proportion of all words. So writers have not found a need to add to their repertoire of bundles to any great extent, with just a 5% increase in the types over the period. The type/token ratio, however, which indicates the variety of bundles used, is notoriously sensitive to text length so smaller corpora are likely to be more densely packed with repeated types, meaning that the apparent decline in the proportion of bundles over the period may well be a function of the counting method.

An additional problem is that normalization may not be a reliable procedure when comparing lexical bundles in corpora of different sizes. While it is common practice to norm frequencies of individual target items to 10,000 words, applying such normalization formula has been shown to produce unreliable results when applied to bundles in corpora of different sizes. Cortes (2015: 205) shows that smaller corpora tend to yield many more bundles than larger corpora (after normalization) and that word combinations do not need to repeat very frequently to meet the cut-off point. When frequencies are normalized, phrases that repeat just a few times could be considered a lexical bundle. Because our 2015 corpus is almost 4 times larger than the 1965 one, this could easily skew our results. Because of this we chose to look at the *proportion* of bundles in the sub corpora rather than the frequency.

¹ We are indebted to an anonymous reviewer for this intriguing observation.

While Table 2 shows there has been an increase in the frequency of bundle tokens in our sample overall, the disciplinary variations presented in Table 3 indicate that sociologists and biologists reduced the number they employed, while applied linguists and electronic engineers followed the opposite trend. Although the reasons for this are unclear, Hyland and Jiang (2018) may be correct in arguing that we are witnessing a change in how sociology and biology construct academic arguments.

Table 3 Frequency of 4-word lexical bundles across disciplines over time

| | App Linguistics | | | Sociology | | | Biology | | | Elec Engineering | | |
|----------------|------------------------|-------------|-------------|------------------|-------------|-------------|----------------|-------------|-------------|-------------------------|-------------|-------------|
| | 1965 | 1985 | 2015 | 1965 | 1985 | 2015 | 1965 | 1985 | 2015 | 1965 | 1985 | 2015 |
| types | 187 | 118 | 176 | 284 | 166 | 117 | 277 | 109 | 109 | 232 | 217 | 244 |
| tokens | 901 | 889 | 1418 | 1466 | 1080 | 931 | 1179 | 818 | 836 | 1182 | 1196 | 2272 |
| types/tokens | 20.8 | 13.3 | 12.4 | 19.4 | 15.4 | 12.6 | 23.5 | 13.3 | 13.0 | 19.6 | 18.1 | 10.7 |
| % of all words | 0.8 | 0.6 | 0.6 | 1.0 | 0.6 | 0.4 | 0.5 | 0.3 | 0.4 | 1.3 | 1.2 | 1.0 |

The numbers in Table 3 are higher than the overall figures in Table 2 because of the influence of our range criteria that bundles must occur in 10% of texts to qualify for inclusion. In table 1 there are 120 texts in each-year sub-corpora, and the types must occur in at least 12 texts for inclusion, thus limiting the number of bundles. However, in Table 3, there are 30 texts in each-year corpus of each discipline, so the tokens here must occur in at least 3 texts in each-year corpus in each discipline. So the tokens in table 3 are higher than those in table 2. We also see a fall in the type/token ratios of bundles across all the disciplines studied in Table 3. Engineers make the greatest use of bundles and appear to use many types not found in other disciplines (cf Hyland 2008b), but even here we can see substantial falls. Bundle use has dropped as a proportion of all words in the texts by 25% in Applied linguistics, 60% in sociology, and by 20% in both biology and electrical engineering. Not only has the proportion of bundles making up texts declined in each field, but the number of different bundles (types) meeting the threshold criteria has fallen, except in electronic engineering which shows a slight rise. Sociology and biology now use 60% fewer types than they did 50 years ago, while the applied linguistics and electrical engineering samples have remained relatively steady, the latter actually finding another 6%.

In the next sections we will look at these changes in more detail, both in terms of the patterns that occur and the disciplinary differences, beginning with changes in the preferences for particular structures.

5. Structural changes in lexical bundles

Differences in the formal properties of lexical bundles are a distinguishing feature of academic discourse (Biber et al. 1999). In academic writing bundles are frequently prepositional phrases with -of fragments (*as a result of*) and noun phrase + of fragments (*the nature of the*) (see also Scott and Tribble 2006: 138, Hyland 2008b) or anticipatory *it* fragments (*it is argued that*) (Salazar 2014, Hyland & Tse 2005). Together, these three forms comprise over 70% of 4-word patterns in academic discourse but rarely figure in conversation, where the majority of bundles contain a verb phrase, particularly ‘personal pronoun + verb phrase’ (e.g. *I don’t know what*).

Biber et al. (2004) and Chen and Baker (2014) have sought to combine broad structural patterns into categories, but these combinations fail to show the ways academic writers package information in texts and how these may change over time. For example, Chen and Baker (2014) grouped patterns into noun phrase-based, prepositional phrase-based and verb phrase-based, but from this grouping we cannot see a clear division between phrasal and clausal structuring of messages which is crucial in academic writing (Biber & Gray 2016). Based on Biber et al.’s (1999) 12 categories of written academic bundles and our analysis of the current diachronic corpus, we identify three distinct categories of formal realisation as shown in Table 4.

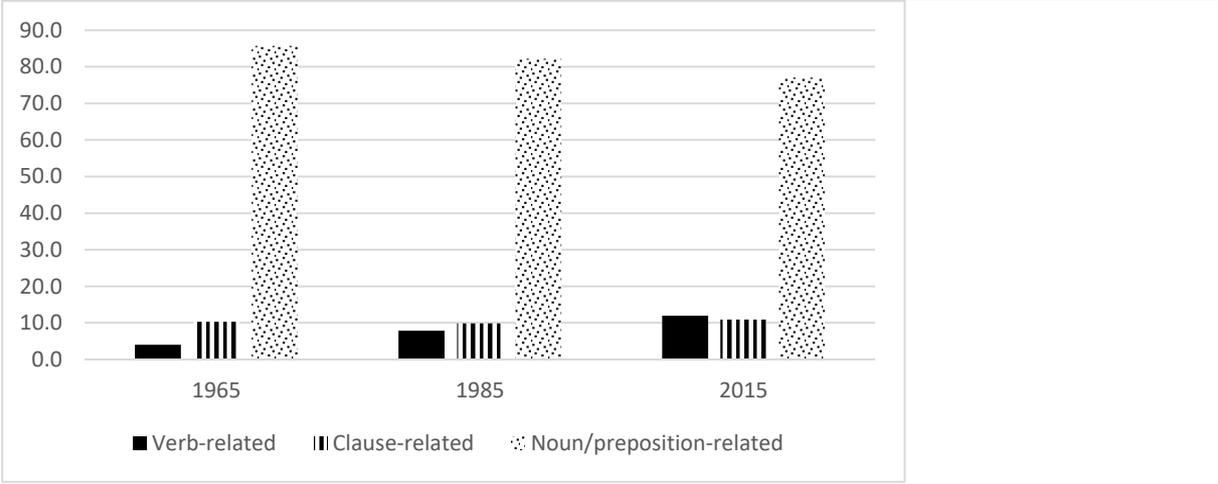
Table 4: Classification of 4-word lexical bundles in academic writing

| |
|--|
| verb phrase-related bundles |
| <ul style="list-style-type: none"> • passive verb (<i>is shown in fig, can be noted that</i>) • copular be (<i>is one of the, is the number of</i>) • imperative (<i>should note that the, let us observe that</i>) |
| clause-related bundles |
| <ul style="list-style-type: none"> • anticipatory <i>it</i> (<i>it is important to, it follows that the</i>) • abstract subject (<i>the goal is to, fig b shows the</i>) • human subject (<i>we shall have to, one should note that</i>) • as-fragments (<i>as can be seen, as shown in fig</i>) • if-fragments (<i>if and only if, if we look at</i>) • there-fragments (<i>there seems to be, there has been a</i>) • wh-fragments (<i>which is to be, which is equivalent to</i>) • that-fragments (<i>that the effect of, that need to be</i>) |
| noun/preposition-related bundles |
| <ul style="list-style-type: none"> • noun phrase with of-phrase fragment (<i>the nature of the, the case of the</i>) • noun phrase with other post-modifier fragment (<i>the fact that the, the extent to which</i>) • prepositional phrase expressions (<i>in terms of the, with respect to the</i>) • comparative expressions (<i>as well as the, as far as the</i>) |

The analysis shows that while noun/prepositional forms continue to account for the overwhelming proportion of bundles, they have also recorded the greatest overall decline in the last 50 years. Figure 1 shows how the percentage of bundles containing a verb phrase has risen significantly (*log Likelihood* =52.00, *p* < 0.001), particularly since 1985, while clause-related

forms have seen a slight proportional rise (*log Likelihood* = 7.20, *p* < 0.05). Within the verb-phrase category preferences for passives and copular verb patterns have doubled overall and together now comprise nearly 12% of all bundles.

Figure 1 Structural change of lexical bundles over time (% of total)



Analysis of the individual disciplines indicates that these changes have not been uniform, with the hard science fields showing big shifts away from verb-phrase bundles towards noun/prepositional types. Table 5 shows the proportional changes in each discipline over the three time points. It is worth pointing out that despite the overall reduction of noun/prepositional bundles seen in Figure 1, individual disciplines actually increased their use of these. Once again, this points to the methodological anomalies thrown up by examining bundles in diachronic corpora. As noted above, setting the distribution criterion at 10% of texts means counting strings as bundles which occur in at least 12 texts in each period and in at least 3 texts in each discipline every period. The overall trends in Figure 1, therefore, are essentially the use of bundles shared across disciplines, rather than a simple numeric adding up of disciplinary figures.

While the largest changes are in the formal sub-categories making up the smallest parts of texts, it is worth mentioning the large falls in the use of bundles containing passives in both biology and electrical engineering, a finding related to the apparent increase in self-mention in these two disciplines (Hyland & Jiang 2016a). Passive bundles have also declined in sociology but increased in applied linguistics, where they comprise nearly 7% of all bundles in the 2015 corpus. Sociology and biology now make greater use of bundles with copular *be*, although numbers are very small. Together these structures often allow writers to offer firm assertions about the statements they are making and are a staple of experimental writing:

- (1) This result *is consistent with the* absence of disagreements ... (Bio)
- (2) These functions *are likely to be* modulated by the sleep/wake cycle... (Bio)
- (3) This result *is consistent with* the theory of reasoned action... (Soc)
- (4) The internal resistance (178-201 Qcm at 12-15°C) *is similar to that* found for the exumbrellar epithelium. (Bio)

Clause related bundles, dominated by anticipatory *it* structures (*it is believed that, it is necessary to*) are also down substantially across all disciplines but sociology. This seems to have fallen from favour most obviously among electrical engineers, despite its value in allowing writers to frontload statements with evaluative meanings:

- (5) But *it is important to* recognize that institutional power is ... (Soc)
- (6) *It is possible that* the decrease in the relative magnitude ... (Soc)

Table 5 Proportional change of structures of 4-word lexical bundles across disciplines over time

| | Applied Linguistics | | | | Sociology | | | | Biology | | | | Electronic Engineering | | | |
|-------------------------------------|---------------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|------------------------|-------------|-------------|--------------|
| | 1965 | 1985 | 2015 | % change | 1965 | 1985 | 2015 | % change | 1965 | 1985 | 2015 | % change | 1965 | 1985 | 2015 | % change |
| (1) verb phrase-related | 15.8 | 14.3 | 15.7 | -0.2 | 14.0 | 10.8 | 11.1 | -20.9 | 24.9 | 17.7 | 14.4 | -42.4 | 34.5 | 29.3 | 26.8 | -22.3 |
| passive verb | 4.1 | 5.5 | 6.9 | 68.3 | 4.9 | 3.3 | 2.4 | -51.9 | 22.0 | 12.5 | 6.1 | -72.2 | 27.7 | 20.4 | 18.9 | -32.0 |
| copular be | 6.5 | 5.2 | 5.4 | -17.1 | 5.5 | 5.4 | 7.6 | 38.0 | 1.9 | 3.2 | 5.3 | 182.1 | 5.8 | 5.9 | 3.9 | -32.7 |
| imperative | 5.1 | 3.6 | 3.4 | -33.7 | 3.5 | 2.1 | 1.1 | -69.7 | 1.1 | 2.1 | 3.0 | 171.2 | 1.0 | 3.0 | 4.0 | 298.9 |
| (2) clause-related | 14.4 | 10.3 | 11.6 | -19.4 | 12.5 | 10.7 | 13.0 | 4.1 | 10.0 | 9.3 | 11.5 | 14.7 | 19.8 | 20.2 | 17.0 | -14.2 |
| anticipatory it | 5.8 | 4.5 | 4.3 | -25.5 | 5.2 | 4.0 | 6.9 | 32.6 | 5.5 | 6.1 | 4.3 | -21.9 | 11.8 | 10.5 | 5.3 | -55.4 |
| abstract entity | 1.0 | 1.5 | 2.4 | 140.0 | 1.4 | 0.7 | 0.0 | -100.0 | 1.2 | 1.2 | 3.8 | 222.4 | 0.0 | 1.3 | 1.6 | 160.0 |
| personal pronoun | 1.9 | 1.0 | 0.4 | -81.3 | 0.0 | 0.5 | 2.8 | 280.0 | 0.0 | 0.0 | 0.7 | 70.0 | 1.8 | 1.4 | 1.2 | -30.6 |
| as-fragments | 1.2 | 1.6 | 2.9 | 136.8 | 0.8 | 2.9 | 2.6 | 214.9 | 0.0 | 0.0 | 1.3 | 130.0 | 3.4 | 3.5 | 5.0 | 47.0 |
| if-fragments | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.4 | 1.8 | 2.6 | 88.6 |
| that-fragments | 2.9 | 1.0 | 0.4 | -85.3 | 2.0 | 2.3 | 0.8 | -63.3 | 0.8 | 0.7 | 0.7 | -15.4 | 0.8 | 0.9 | 1.1 | 38.7 |
| there-fragments | 0.6 | 0.8 | 0.7 | 27.1 | 2.7 | 0.4 | 0.0 | -100.0 | 2.5 | 1.2 | 0.6 | -75.7 | 0.0 | 0.0 | 0.0 | 0.0 |
| where-fragments | 1.1 | 0.0 | 0.6 | -49.2 | 0.4 | 0.0 | 0.0 | -100.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 | 0.8 | 0.3 | -61.0 |
| (3) noun/preposition-related | 69.8 | 75.4 | 72.6 | 4.0 | 73.5 | 78.4 | 75.9 | 3.3 | 65.1 | 73.0 | 74.2 | 14.0 | 45.7 | 50.4 | 56.2 | 23.0 |
| noun phrase with of | 18.9 | 20.1 | 19.0 | 0.5 | 27.6 | 24.6 | 20.4 | -26.1 | 12.0 | 19.2 | 26.0 | 115.5 | 19.8 | 19.8 | 20.3 | 2.7 |
| noun phrase with other | 7.7 | 7.5 | 7.1 | -7.0 | 5.6 | 9.8 | 6.3 | 13.3 | 11.6 | 5.6 | 0.7 | -93.8 | 4.6 | 4.5 | 2.4 | -47.0 |
| prepositional phrase | 39.6 | 44.5 | 44.3 | 11.8 | 36.2 | 37.5 | 42.2 | 16.5 | 39.7 | 46.3 | 41.7 | 5.2 | 20.1 | 24.2 | 30.9 | 53.9 |
| comparative | 3.7 | 3.1 | 2.3 | -38.4 | 4.1 | 6.5 | 7.0 | 70.6 | 1.7 | 1.8 | 5.7 | 238.5 | 1.3 | 1.9 | 2.6 | 104.6 |

Noun and preposition related bundles continue to be the main pattern across these fields and now represent a larger share of bundles in texts than before, particularly in the sciences. Within this category bundles containing a *prepositional phrase* make up 31% of all bundles in electrical engineering and over 40% in the other disciplines we studied, rising in all four fields. This pattern is particularly valuable to writers in the soft knowledge fields as it facilitates the discursive exploration of possibilities and limiting conditions, identifying and elaborating logical relationships in argument:

(7) we were necessarily restricted *with respect to the* number of items... (AL)

(8) *In addition to the* foregoing racial disparities, more frequent single mortgage borrowing... (Soc)

Noun phrases with of comprise the second most salient sub-group, falling in sociology but still comprising about 20% of all the lexical bundles in each discipline. This pattern covers a range of meanings in academic discourse and are often used to specify the attributes of what is being discussed. They are therefore particularly useful to scientists and other experimental researchers to identify quantity, place or size (9), to mark existence (10), or highlight qualities (11) (see also Simpson & Ellis 2010):

(9) *the temperature of the* saline was systematically varied from 6 to 16°C. (Bio)

... feedback loop from the output emitter follower to *the base of the* input transistor...

(EE)

(10) ... further inquiries, ideally in *a wide range of* L2 learning contexts. (AL)

the presence of the injection barrier certainly limits the current flowing. (EE)

(11) This may be due to *the nature of the* cloze activity. (AL)

the total elongation of the distal end (Dist) as *a function of the* total length of the bone

(1)... (Bio)

The preference for these two patterns helps distinguish academic writing from other registers (Biber et al. 1999) and results from the fact that they provide authors with a range of obviously useful academic formula. The ubiquitous presence of these patterns is apparent in the familiarity of particular 4-word bundles themselves, with *on the other hand, on the basis of, in the case of, at the same time, in terms of the, in the presence of, in the form of* and *in the absence of* remaining among the top 20 bundles in academic writing over the past 50 years (Appendix 2 presents the most frequent 20 bundles over time). Table 6 shows the most commonly used bundles in the four fields in frequency order for each year, indicating something of both disciplinary differences and the persistence of high frequency bundles over time.

Table 6 The 10 most frequent 4-word lexical bundles in each discipline over time (recurring bundles within a discipline are shaded)

| Applied linguistics | | | Sociology | | |
|------------------------|----------------------|----------------------|------------------------|---------------------|---------------------|
| 1965 | 1985 | 2015 | 1965 | 1985 | 2015 |
| on the basis of | on the basis of | at the same time | the percent of the | the percent of the | on the basis of |
| at the end of | on the other hand | in the context of | in the case of | on the other hand | are more likely to |
| on the other hand | at the end of | on the other hand | on the basis of | on the basis of | at the same time |
| at the same time | in the case of | a wide range of | at the same time | the extent to which | on the other hand |
| in such a way | the extent to which | in terms of the | on the other hand | in the case of | over the course of |
| in terms of the | the purpose of the | in the present study | only percent of the | in terms of the | in the face of |
| in the case of | the degree to which | at the time of | in terms of the | on the one hand | in the case of |
| it is possible to | the ways in which | on the basis of | the degree to which | at the same time | as well as the |
| the fact that the | in terms of the | it is important to | found that the percent | in the form of | it is important to |
| as well as the | in the area of | was found to be | in view of the | as well as the | as a result of |
| Biology | | | Electronic engineering | | |
| 1965 | 1985 | 2015 | 1965 | 1985 | 2015 |
| the percent of the | in the presence of | in the absence of | as shown in fig | as shown in fig | as shown in fig |
| on the basis of | as a result of | in the case of | in terms of the | it is easy to | in the case of |
| in the case of | as a function of | on the other hand | it is possible to | the total number of | in this section we |
| as a result of | in the absence of | the size of the | on the other hand | as a function of | on the other hand |
| in the presence of | in the case of | in the presence of | if and only if | if and only if | the value of the |
| in the absence of | on the other hand | as a result of | the size of the | it is possible to | with respect to the |
| of the effects of | on the basis of | as well as the | as a function of | the size of the | if and only if |
| the end of the | the end of the | a wide range of | the fact that the | is said to be | can be written as |
| on the other hand | the structure of the | at the level of | it is assumed that | can be used to | as a function of |
| the development of the | the role of the | the total number of | in the case of | to the number of | in the presence of |

Clearly *on the other hand* and *on the basis of* are perennial favourites, with the former in all disciplines in all years and the latter only absent in engineering, which is also alone in the persistent popularity it affords to *if and only if* and *as shown in fig*. *At the same time* and *in terms of the* only occur in the soft fields (especially after 1965) and *as a function of* only in the sciences. *The end of the* does not figure in sociology or engineering and disappears from the other two lists after 1985. Diachronically, *in the area of* and *the structure of* have dropped off the lists and *in the context of* and *at the time of* (applied linguistics), *are more likely to* (sociology) and *a wide range of* (applied linguistics and biology) have recently made an appearance on them. More interestingly, almost all the high frequency bundles in each discipline are noun phrases with of fragments or prepositional phrase expressions. Only *it is possible to* and *it is important to* fall outside these sub-categories. Equally, almost all relate to the organisation of the discourse, offering transitions, results, or limiting the conditions under

which something is true. It is to the changing functions of lexical bundles in these disciplines that we now turn.

6. Changes in functions of lexical bundles

Our categorisation of the functions which bundles are used to perform attempts to avoid the proliferation of types and sub-types found in the work of Nattinger and DeCarrico (1992) and others, and seeks to distil the data into a compact model. Here we follow Biber, Conrad and Cortes (2004) and Hyland (2008a, 2008b, 2012) in grouping bundles into three main functional groups: research-oriented, dealing with referential functions in the real world; text-oriented, concerned with the organisation of the discourse; and participant-oriented, concerned with stance and evaluation. Each category is further subdivided into the main focus of the bundle as shown in Table 7.

Table 7: Functional categories of lexical bundles (Hyland 2008a).

| |
|--|
| Research-oriented |
| <ul style="list-style-type: none"> • location – indicating time and place (<i>at the same time, in the present study</i>); • procedure (<i>the use of the, the role of the, the purpose of the, the operation of the</i>); • quantification (<i>the magnitude of the, a wide range of, one of the most</i>); • description (<i>the structure of the, the size of the</i>); |
| Text-oriented |
| <ul style="list-style-type: none"> • transition signals – establishing additive or contrastive links between elements (<i>on the other hand, in addition to the, in contrast to the</i>); • resultative signals – mark inferential or causative relations between elements (<i>as a result of, it was found that, these results suggest that</i>); • structuring signals – text-reflexive markers which organise stretches of discourse or direct reader elsewhere in text (<i>in the present study, in the next section, as shown in fig.</i>); • framing signals – situate arguments by specifying limiting conditions (<i>in the case of, with respect to the, on the basis of, in the presence of, with the exception of</i>). |
| Participant-oriented |
| <ul style="list-style-type: none"> • stance features – convey the writer’s attitudes and evaluations (<i>are likely to be, may be due to</i>); • engagement features – address readers directly (<i>it should be noted, as can be seen</i>). |

The data in Table 8 show authors have reduced their use of bundles in almost all categories. The only exception being biology which has seen significant increases in both text-oriented and participant-oriented functions ($\log \text{Likelihood} = 15.32, p < 0.001$; $\log \text{Likelihood} = 20.53, p < 0.001$).

The most striking change indicated in Table 8 is in the shifting balance of participant- and research-oriented functions, with the hard sciences reducing their use of bundles which focus on reporting research and adopting more forms which carry epistemic evaluations, attitudes

or modal meanings. These participant forms are interpersonal and concerned with the involvement of writers (mainly) and readers in the text, drawing attention to the dialogic nature of discourse. They represent the writers' annotations to comment on the possible accuracy of a claim, the extent they want to commit themselves to it, or the attitude they want to convey while drawing a reader explicitly into a discourse. These functions are illustrated here:

(12) *It is well known* that statistically independent real-valued Gaussian processes with independent... (EE)

(13) this function *is likely to be* the same throughout development. (Bio)

The option writers have to comment on their attitudes to propositions and readers in this way varies according to discipline and has traditionally been exercised most frequently by those in the social sciences. Here texts are often structured to evoke affinity and engagement while the hard science fields tend to downplay these aspects of rhetorical persuasion (Hyland 2004). However, both applied linguistics and sociology have seen significant falls in the proportion of bundles devoted to this function. Both hard science fields in the corpus, in fact, now have a higher percentage of participant-oriented bundles than the two soft knowledge disciplines. This suggests a more measured epistemic stance and a more circumspect approach to authorial intrusion in applied linguistics and sociology than in the past, perhaps related to what has been perceived as an increasing scientism in the social sciences with a more hard science orientation in their dominant methods and approaches (e.g. Glynos & Howarth 2007).

Electronic engineering and, especially, biology have greatly increased the proportion of participant bundles so that these represent over a third of all tokens in engineering. It seems that personal intrusion to explicitly get behind arguments now plays a far greater part in creating a convincing discourse in the sciences:

(14) it is fairly *easy to see that* consensus may not be achieved ... (EE)

(15) *it is difficult to* relate the two studies directly (Bio)

Efforts to engage readers by explicitly marking their presence have also increased, particularly in engineering. These principally take the form of directives using anticipatory *it* patterns:

(16) In other words, although mixtures of zero al exists, *it is necessary to* carefully optimize the material parameters associated with the rotational viscosity. (EE)

(17) *It should be noted* that the extracted MAPs are associated with the polymerized tubulin. (Bio)

Here the writer pulls the audience into the discourse at critical points to guide them to particular interpretations.

Table 8 Proportional change of functions of 4-word lexical bundles across disciplines over time

| | Applied Linguistics | | | | Sociology | | | | Biology | | | | Electronic Engineering | | | |
|-----------------------------|---------------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|--------------|------------------------|-------------|-------------|--------------|
| | 1965 | 1985 | 2015 | % change | 1965 | 1985 | 2015 | % change | 1965 | 1985 | 2015 | % change | 1965 | 1985 | 2015 | % change |
| Research-oriented | 36.0 | 40.8 | 40.9 | 13.7 | 45.6 | 39.1 | 47.5 | 4.2 | 72.7 | 53.2 | 48.0 | -34.0 | 34.1 | 33.0 | 30.8 | -9.6 |
| location | 13.4 | 11.4 | 8.1 | -39.6 | 6.4 | 5.8 | 11.5 | 79.2 | 5.6 | 9.2 | 11.7 | 109.4 | 4.0 | 3.3 | 2.0 | -50.2 |
| procedure | 5.2 | 8.9 | 11.4 | 119.0 | 6.3 | 13.3 | 16.0 | 155.3 | 40.0 | 19.4 | 9.8 | -75.5 | 8.1 | 7.9 | 9.9 | 21.9 |
| quantification | 6.5 | 8.7 | 9.8 | 49.7 | 18.5 | 10.6 | 10.6 | -42.5 | 4.9 | 7.3 | 13.5 | 174.8 | 6.5 | 8.3 | 10.7 | 63.5 |
| description | 10.8 | 11.9 | 11.6 | 7.4 | 14.4 | 9.3 | 9.3 | -35.1 | 22.1 | 17.2 | 12.9 | -41.6 | 15.5 | 13.6 | 8.3 | -46.4 |
| Text-oriented | 40.6 | 42.5 | 43.7 | 7.6 | 34.1 | 43.8 | 36.7 | 7.7 | 21.1 | 32.0 | 35.9 | 69.9 | 39.8 | 40.2 | 35.2 | -11.7 |
| transition | 12.4 | 8.3 | 6.8 | -45.5 | 9.1 | 9.4 | 8.5 | -6.5 | 5.7 | 9.0 | 9.8 | 72.6 | 5.9 | 5.9 | 6.5 | 9.3 |
| resultative | 1.1 | 4.6 | 5.9 | 433.7 | 2.5 | 4.5 | 3.1 | 26.8 | 1.6 | 2.8 | 6.0 | 271.1 | 4.4 | 3.5 | 3.7 | -17.0 |
| structuring | 3.9 | 6.9 | 8.0 | 105.1 | 1.4 | 2.0 | 0.5 | -60.6 | 1.8 | 1.5 | 0.7 | -59.7 | 12.4 | 13.2 | 8.0 | -35.1 |
| framing | 23.2 | 22.7 | 23.1 | -0.6 | 21.2 | 27.9 | 24.6 | 15.9 | 12.0 | 18.7 | 19.4 | 60.9 | 17.2 | 17.6 | 17.0 | -0.8 |
| Participant-oriented | 23.4 | 16.6 | 15.4 | -34.4 | 19.7 | 17.1 | 14.7 | -25.4 | 6.2 | 14.8 | 16.1 | 160.8 | 26.1 | 26.8 | 34.0 | 30.5 |
| stance | 20.3 | 13.2 | 12.6 | -38.2 | 16.2 | 13.1 | 11.6 | -28.7 | 5.3 | 13.3 | 13.6 | 159.3 | 18.5 | 17.0 | 21.2 | 14.5 |
| engagement | 3.1 | 3.5 | 2.8 | -9.2 | 3.5 | 4.1 | 3.1 | -10.5 | 0.9 | 1.5 | 2.5 | 169.2 | 7.5 | 9.8 | 12.8 | 69.5 |

Interestingly, these changes in participant-oriented bundles mirror those found for other features used to express stance (Hyland & Jiang 2016a) and engagement (Hyland & Jiang 2016b) in academic writing. Hyland and Jiang speculate this may not be unrelated to the need to address audiences beyond an immediate group of informed insiders. A more visible presence in the text signals an overt authorial role in interpretations of data and for claims of novelty which helps to ensure that the writers' contributions do not go unnoticed by tenure and promotion committees and commercial sponsors.

In contrast, electrical engineering and biology have reduced the proportion of bundles devoted to research, biology recording the steepest fall and from the highest point. Research-oriented forms impart a real-world, empirically-focused sense to a text and in 1965 over 70% of all bundles in the biology corpus were of this kind. This has now fallen to below half in biology and to 31% in engineering, largely as a result of a decline in bundles depicting the description or specification of research objects or contexts:

(18) *The output of the voltage reference and that of the temperature sensor are compared ...*
(EE)

(19) *The size of the spikes corresponding to the small depolarizations described ...* (Bio)
Both disciplines, however, appear to be making greater use of bundles related to the quantification of findings. This sub-function is now among the most used bundle tokens in the two hard science disciplines, functioning to clearly spell out, and often emphasise, not only connections between entities but their scale. This offers a way to strengthen the claim being made:

(20) *the size of the inductive network increases dramatically compared to the leadframe structure.* (EE)

(21) *One of the most widely discussed concepts in education and educational reform these days is critical thinking.* (AL)

The biggest rises in research-oriented bundles in the soft knowledge corpora were in those depicting research procedures, showing the ways that experiments and research were conducted:

(22) *in the use of the semi-structured interview protocols with the lead author...* (AL)

(23) *In the course of these searches we identified 10 community advocacy...* (Soc)

This increased focus on the details of research practices helps convey a strongly empirical flavour to a text and the sense of research activity in support of authorial interpretations. This not only demonstrates disciplinary competence in control of resources and research practices, but the growth in their use, and away from participant-oriented types, may also reflect a greater emphasis on empirically centred papers in these disciplines. Certainly, applied linguistics has

seen considerable changes during this period. In 1965 it was a young discipline with an undeveloped literature and a greater focus on personal accounts of language teaching. The increase in empirically-oriented studies, the broadening of the discipline to embrace a wider array of topics, and the massive growth of a literature which supports its academic endeavours have all contributed to massive changes in how claims are argued and accepted.

Finally, Table 8 shows the importance of text-oriented bundles in these corpora, representing over a third of all forms in each discipline. These attempt to shape and constrain a text to better ensure that readers will recover the writer's interpretations and goals, organising the discourse with the assumed readers' needs in mind. Perhaps more than other functions, they display the discursively crafted and rhetorically-machined nature of the research article genre, providing warrants, connecting ideas, directing readers around the text, drawing conclusions and specifying limitations. While uses in sociology and electrical engineering show a drop since 1985, only one discipline records a fall overall, pointing to their rhetorical importance in academic persuasion. This indicates they are not merely text organising devices but mark writers' assessments of readers' expectations and knowledge.

Structuring bundles, which provide readers with a cognitive map of the text by referring to its stages and announcing discourse goals, have jumped considerably in applied linguistics as texts have become longer. These help frame, scaffold, and present arguments as a coherently managed and organised arrangement, reflecting writers' awareness of readers' needs to follow a sustained discussion:

(24) Some examples of these differences *are shown in Table 1.* (AL)

(25) *In the next section* we examine L2 reading instruction. (AL)

Despite the reader-friendliness of this function, the increase in applied linguistics has not been followed by the other disciplines. Framing devices, however, make up the largest proportion of text-oriented bundles in each field, growing considerably in biology and comprising almost a quarter of all bundles in the soft knowledge corpora. These also help organize arguments but work more locally by specifying cases and pointing to limitations:

(26) ... think of complexity *in terms of the* number of phenotypic parts. (Bio)

(27) forms of racism that they engendered are clear *in the case of* the Black Cherokee (Soc)

These bundles tend to be preposition + of structures and are used to focus readers on a particular instance or to specify the conditions under which a statement can be accepted, working to elaborate, compare and emphasise aspects of an argument.

The biggest proportional rises, although with lower frequencies, were in the use of resultative markers. These bundles highlight causal connections and in so doing construct an assertive construal of events, underlining the writer's interpretation of data and leading readers towards

a preferred understanding of events (28). But while they highlight the inferences the writer wants readers to draw, they are often conciliatory, as in (29), opening a discursive space in which the reader might feel free to dispute it.

(28) Our *results indicate that the* balance between proximal and distal growth ... (Bio)

(29) The *findings suggest that the* relationship between vocabulary frequency and test performance... (AL)

The increase in bundles in this sub-category may therefore reflect writers' awareness of a need to spell out their interpretations more clearly to ensure readers get the take home message. This is, perhaps, increasingly necessary as credit for original claims becomes ever-more important in a highly competitive culture of institutional assessment and the growing need to persuade outsiders in HR committees and corporate funders of the value of one's work.

7. Conclusions

Lexical bundles, like other features of disciplinary discourses, are not static and invariant markers of professional research writing but change in response to new conditions and contexts. The small study reported here is the first to offer an account of such changes in research writing. While confined to the top journals in just four fields, our results show a significant shift in uses and perhaps how these disciplines understand academic argument and the ways their members seek to persuade peers. Some bundles, such as *on the other hand*, *in the case of*, and *on the basis of*, have remained popular throughout the period, but there has been a strengthening in the use of verb bundles away from those composed of noun/preposition-related forms and to participant-oriented tokens from research-oriented forms. The more interesting changes, however, have been within disciplines, with both biology and electrical engineering now using a greater proportion of noun/preposition bundles, especially noun phrases with *of* and prepositional phrases, forms which facilitate the a more discursive rhetoric and the presentation of possibilities, limitations and logical relationships (*in the case of*, *on the other hand*, *as a function of*).

The most conspicuous changes have been in the predominant functions that bundles are used to perform in these disciplines, with the hard sciences now using fewer bundles which focus on reporting research and adopting more forms which carry interpersonal and evaluative meanings. The greater intrusion of writers into their texts using bundles which emphasise an authorial stance and seek to engage readers helps strengthen claims and ensure readers are in no doubt about their source. Speculatively, this may be in response to radical changes in publishing and research practices in recent decades, with the emphasis on securing outsider funding, interdisciplinary collaboration and knowledge exchange encouraging writers to reach out to

new audiences perhaps unfamiliar with specialist their topics, and with the encroachment of the appraisal culture, requiring. The soft knowledge fields, on the other hand, have been slowly moving in the opposite direction towards more ‘author-evacuated’ forms, replacing participant bundles with research and text-oriented ones. Again, the reasons for this are unclear, but they may not be unrelated to an increase in more empirically grounded and quantitative studies which restrict opportunities for more extensive overt stance-taking.

The changes we have uncovered therefore seem to indicate a strengthening of patterns traditionally associated with hard science writing, being more nominal and prepositional, although passive constructions and bundles built on anticipatory *it*, staples of scientific writing, have declined. In terms of functions they seem to represent a shift in emphasis in disciplinary argument patterns which we believe is indicative of rhetorical practices more generally in academic writing and worth exploring further.

Last but not least, by this study, we may have also drawn attention to the method of identifying lexical bundles in diachronic disciplinary corpus, most importantly by norming frequencies to 20 cases per million words and presenting results as a proportion and by employing a 10% distribution as a cut off criterion. While some may argue with these decisions, we believe they overcome some of the methodological difficulties of bundle identification across corpora of different sizes and open the way for further studies of diachronic change in bundle use and in disciplinary preferences.

References

- Anthony, L. (2014). Antconc (Version 3.4.3 w). Tokyo: Waseda University. Retrieved from <http://www.antlab.sci.waseda.ac.jp/>
- Biber, D. (2006). *University language: a corpus-based study of spoken and written registers*. Amsterdam: Benjamins.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14(3), 275-311.
- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at... : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of written and spoken English*. Harlow: Longman.
- Butler, C. (1998). Collocational frameworks in Spanish. *Journal of Corpus Linguistics*. 3 (1): 1-32.

- Chen, Y-H. & Baker, P. (2010). Lexical bundles in L1 and L2 student writing. *Language, learning and technology*, 14(2), 30-49.
- Chen, Y.-H., & Baker, P. (2014). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays. *Applied Linguistics*, 1-33.
- Conklin, K & Schmitt, N. (2008). *Formulaic Sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?* Oxford: Oxford University Press.
- Cortes, V. (2004) Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes*, 23, 397-423.
- Cortes, V. (2006) Teaching lexical bundles in the disciplines: an example from a writing intensive history class. *Linguistics and Education*, 17, 391-406.
- Cortes, V. (2015). Situating lexical bundles in the formulaic language spectrum: Origins and functional analysis developments. In V. Cortes and E. Csomay (Eds.), *Corpus-based research in applied linguistics: Studies in honor of Doug Biber* (pp. 197-218). Amsterdam: John Benjamins.
- Culpeper, J., & Kytö, M. (2002). Lexical bundles in early modern English dialogues. In T. Fanego, B. Méndez-Naya, & E. Seoane (Eds.), *Sound, words, texts and change* (pp. 45-63). Amsterdam: John Benjamins.
- Firth, J. R. (1951). Modes of meaning. *Essays and studies (the English Association)*, 118-149.
- Glynos, J. & Howarth, D. (2007). *Logics of critical explanation in social political theory*. London: Routledge.
- Hoey, M. (2005). *Lexical priming: a new theory of words and language*. London: Routledge.
- Hyland, K. (2004) *Disciplinary Discourses: social interactions in academic writing*. Ann Arbor: University of Michigan Press
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.
- Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.
- Hyland, K. & Jiang, K. (2016a). Change of attitude? A diachronic study of stance. *Written Communication*, 33(3), 251-274
- Hyland, K. & Jiang, K. (2016b). “We must conclude that...”: A diachronic study of academic engagement. *Journal of English for Academic Purposes*, 24, 29-42.
- Hyland, K. & Jiang, K. (2018). “We believe that...”: changes in an academic stance marker. *Australian Journal of Linguistics*, 38(2).
- Hyland, K. & Tse, P. (2005). Hooking the reader: a corpus study of *evaluative that* in abstracts. *English for Specific Purposes*, 24(2), 123-139.
- Kopaczyk, J. (2012). Applications of the lexical bundles method in historical corpus research. In P. Pezik, *Corpus Data across Languages and Disciplines* (pp.83-95). Berlin: Peter Lang.

- Kopaczyk, J. (2013). *The legal language of Scottish burghs: Standardization and lexical bundles, 1380–1560*. Oxford: Oxford University Press.
- Ma, G. H. (2009). A study of four-word lexical bundles in English major students' timed writing. *Foreign Language Teaching and Research*, 41(1), 54-60.
- Meunier, F., & Granger, S. (2008). *Phraseology in foreign language learning and teaching*. Amsterdam, the Netherlands & Philadelphia, PA: John Benjamins.
- Nattinger, J. & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: OUP.
- Pang, W. (2010). Lexical Bundles and the Construction of an Academic Voice: A Pedagogical Perspective. *Asian EFL Journal*, 47, 1-13.
- Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Amsterdam: John Benjamins.
- Scott, M. & Tribble, C. (2006). *Textual patterns*. Amsterdam: Benjamins.
- Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In T. Upton and U. Connor (eds.): *Discourse in the professions: Perspectives from corpus linguistics* (pp. 37-64). Amsterdam: John Benjamins.
- Simpson-Vlach, R. & Ellis, N. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31, 487-512.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: CUP.
- Wei, N.X. 2007. Phraseological characteristics of Chinese learners' spoken English: Evidence of lexical chunks from COLSEC. *Modern Foreign Languages* 30 (3).
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press
- Wray, A. & Perkins, M. (2000). The functions of formulaic language. *Language and communication*, 20, 1-28.

Appendix 1: Journal list

Applied Linguistics

TESOL Quarterly (1967-)
Language Learning (1948-)
Foreign Language Annals (1967-)
Modern Language Journal (1916-)
College Composition and Communication (1950-)

Sociology

American Journal of Sociology (1895-)
Social problems (1953-)
The British Journal of Sociology (1950-)
American Journal of Economics and Sociology (1941-)
The Sociological Quarterly (1960-)

Biology

The Quarterly Review of Biology (1926-)
Biological Reviews (1923-)
Radiation Research (1954-)
BioScience (1964-)
The Journal of Experimental Biology (1923 -)

Electrical Engineering

Proceedings of the IEEE (1963 -)
Automatica (1963 -)
IEEE Transactions on Automatic Control (1963 -)
IEEE Journal of Solid-State Circuits (1966 -)
IEEE Transactions on Information Theory (1963 -)

Appendix 2: The most frequent 20 4-word bundles over time (normed frequency & range)

| 1965 | | | 1985 | | | 2015 | | |
|------|----|---------------------|------|----|----------------------|------|----|---------------------|
| F | R | bundles | F | R | bundles | F | R | bundles |
| 2.4 | 25 | the percent of the | 1.3 | 52 | on the other hand | 1.1 | 55 | on the other hand |
| 1.7 | 51 | on the basis of | 1.1 | 34 | on the basis of | 1.0 | 43 | in the case of |
| 1.5 | 36 | in the case of | 0.9 | 41 | in the case of | 0.8 | 29 | on the basis of |
| 1.5 | 46 | on the other hand | 0.7 | 26 | as a result of | 0.7 | 39 | in the context of |
| 1.0 | 34 | at the same time | 0.7 | 19 | as a function of | 0.6 | 26 | with respect to the |
| 0.9 | 34 | in terms of the | 0.7 | 22 | the extent to which | 0.6 | 32 | at the same time |
| 0.8 | 29 | at the end of | 0.6 | 25 | the end of the | 0.6 | 41 | as well as the |
| 0.8 | 26 | the end of the | 0.5 | 26 | in terms of the | 0.5 | 33 | as a result of |
| 0.8 | 24 | it is possible to | 0.5 | 28 | as well as the | 0.5 | 15 | as shown in fig |
| 0.7 | 28 | the nature of the | 0.5 | 25 | in the presence of | 0.5 | 31 | in terms of the |
| 0.7 | 20 | the size of the | 0.5 | 24 | in the form of | 0.5 | 24 | in the absence of |
| 0.7 | 30 | the fact that the | 0.5 | 23 | at the same time | 0.5 | 38 | it is important to |
| 0.7 | 23 | as a result of | 0.5 | 22 | can be used to | 0.5 | 29 | as a function of |
| 0.7 | 27 | in view of the | 0.5 | 23 | the structure of the | 0.5 | 34 | in the form of |
| 0.6 | 19 | in the presence of | 0.4 | 23 | the fact that the | 0.5 | 31 | a wide range of |
| 0.6 | 20 | is shown in fig | 0.4 | 14 | the size of the | 0.5 | 21 | in this section we |
| 0.6 | 19 | the case of the | 0.4 | 17 | in the absence of | 0.5 | 27 | a large number of |
| 0.6 | 23 | in the form of | 0.4 | 26 | it is important to | 0.4 | 21 | in the presence of |
| 0.6 | 18 | with respect to the | 0.4 | 22 | the nature of the | 0.4 | 16 | is shown in fig |
| 0.5 | 24 | the basis of the | 0.4 | 19 | it is possible to | 0.4 | 15 | the value of the |

Author's address

Ken Hyland
 School of Education and Lifelong Learning
 University of East Anglia
 Norwich, NR4 7TJ
 UK
 K.hyland@uea.ac.uk

Feng (Kevin) Jiang
 School of Foreign Language Education

Jilin University
#2699 Qian Jin Street
Chang Chun, Jilin, 130012
P. R. China
kevinjiang@jlu.edu.cn