



University of Dundee

Paradigmatic influences on lexical bundles in research articles in the discipline of education

Candarli, Duygu; Jones, Steven

Published in:
Corpora

DOI:
[10.3366/cor.2019.0170](https://doi.org/10.3366/cor.2019.0170)

Publication date:
2019

Document Version
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education. *Corpora*, 14(2), 237-263. <https://doi.org/10.3366/cor.2019.0170>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.euppublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

Paradigmatic influences on lexical bundles in research articles in the discipline of education

Abstract

Lexical bundles are pervasive in English academic writing; however, little scholarly attention has been paid to how quantitative and qualitative research paradigms influence the use of lexical bundles in research articles. In order to investigate this, we created two equal-size corpora of research articles in the discipline of education. Four-word lexical bundles were examined in terms of their structural characteristics and discourse functions in the quantitative and qualitative research articles published in international English-medium journals. We attribute intra-disciplinary variations in the use of lexical bundles to the knowledge-making practices that are specific to quantitative and qualitative research articles. This paper provides further evidence that the research article is not a unitary construct. The results have implications for academic writing, and corpus building and design in academic discourse. One of the key implications of this study is that L2 novice writers need to take into account the influences of research paradigm on the use of lexical bundles when writing research articles for English-medium journals in the discipline of education.

Keywords: lexical bundles; academic writing; paradigmatic influences; education.

1. Introduction

Phraseological patterns are one of the most essential aspects of linguistic production, enhancing coherence, writing fluency, and group membership (e.g., Hyland, 2012; and

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.euppublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

Wray, 2002). A wide range of terms have been used to describe phraseological patterns, including ‘formulaic sequences’ (Wray, 2002), ‘lexical bundles’ (Biber *et al.*, 1999), and ‘lexical phrases’ (Nattinger and DeCarrico, 1992). The present paper focuses on lexical bundles (LBs), i.e., ‘sequences of word forms that commonly go together in natural discourse’ (Biber *et al.*, 1999: 990), in English academic writing. This study extends the knowledge of LBs in English academic writing by examining the influences of quantitative and qualitative research paradigms within a single discipline - education - that embraces quantitative, qualitative, and mixed-methods research (Cohen *et al.*, 2013). A research paradigm, which can be defined as ‘a shared belief system or set of principles on what problems are to be investigated and how to investigate them’ (Cohen *et al.*, 2013: 13), can influence the construction of disciplinary knowledge-making practices in social science research articles (RAs). This definition of a research paradigm is consistent with those given in previous studies that investigated paradigmatic influences on metadiscourse in RAs (Cao and Hu, 2014; and Hu and Cao, 2015). As Gray (2015: 6) noted, ‘little attention has been paid to the possibility that research articles themselves are not a monolithic concept’. Hence, this paper investigates the paradigmatic influences on the structural types and discourse functions of LBs in RAs within a single discipline, in order to reveal intra-disciplinary variations in the use of lexical bundles.

Lexical bundles, which are identified according to their frequency of occurrence and dispersion criteria (Biber *et al.*, 1999; and Cortes, 2004), reflect the ‘phraseological tendency of language’, which suggests that words ‘go together and make meanings by

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

their combinations’ (Sinclair, 2004: 29). LBs are usually not idiomatic in meaning or complete structural units (Biber *et al.*, 2004), but constitute ‘important building blocks in discourse’ (Biber and Barbieri, 2007: 270) of English academic writing.

Lexical bundles have been shown to play a central role and fulfil a wide range of functions in English academic writing (e.g., Biber *et al.*, 2004; and Cortes, 2004). The three major discourse functions of LBs can be summarised as follows: (1) *referential expressions* frame and specify attributes (e.g., ‘in the context of’); (2) *discourse organisers* organise the text and establish links between elements in the text (e.g., ‘on the other hand’); (3) *stance expressions* convey writers’ (un)certainty and attitudes (e.g., ‘it is possible to’). LBs reflect how academic writing is framed by a discourse community, reveal disciplinary membership, and enhance effective communication in academic writing (Biber *et al.*, 2004; Cortes, 2004; and Hyland, 2008, 2012).

2. Review of previous studies

A large and growing body of literature focuses on LBs in published English academic writing (Biber, 2009; Biber and Barbieri, 2007; Cortes, 2004, 2013; and Hyland, 2008, 2012). Most of these studies have examined four-word LBs and their discourse functions in academic prose. The general trend that has emerged from the study of LBs is reliance on noun and prepositional phrases, which primarily function as referential expressions in published English academic writing.

Disciplinary variations are evident in the use of LBs in academic writing (Cortes, 2004; Durrant, 2017; and Hyland, 2008, 2012). For example, Hyland (2008)

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

noted that only five four-word bundles were shared across the disciplines of biology, electrical engineering, applied linguistics, and business studies. It has also been found that texts written in the hard sciences contain more bundles that highlight research procedures and practices than social science texts, which feature a comparatively greater reliance on stance expressions (Hyland, 2008, 2012). The present study focuses on the discipline of education, since it is relatively under-researched in phraseology research. The discipline is characterised as soft (i.e., based on human experiences rather than abstract concepts) and applied (i.e., directed towards practical implications) (Becher, 1989).

Drawing on Bernstein's (1999) characterisation of academic disciplines in relation to the knowledge structures they represent, Maton (2007) argued that disciplines can be conceptualised along a continuum from those that represent a knowledge code and those that represent a knower code. Disciplines that are characterised by a knowledge code, such as the natural sciences, value 'coherent, explicit, systematically principled' (Bernstein, 1999: 161) knowledge-making practices that rely on empirical enquiry rather than an interpretive approach (e.g., Hu and Cao, 2015; and Maton, 2007). Humanities disciplines, on the other hand, are characterised by a knower code, and value knowledge-making practices that rely on interpretation and evaluation, which may reflect the knower's personal insights (e.g., Hu and Cao, 2015; and Maton, 2007). Social science disciplines occupy the middle ground in the continuum, and 'there is always a kind of dynamic tension between the science and the social in the discourse' (Wignell, 2007: 202). Trowler (2014: 1727) notes that

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

disciplinary practices in social sciences tend to be ‘situationally contingent’. This suggests that these practices vary according to contextual factors, one of which can be the research paradigm.

There are multiple research paradigms in the social sciences; however, the quantitative and qualitative paradigms are considered the two major paradigms in the discipline of education (Cohen *et al.*, 2013; and Creswell, 2009). Quantitative and qualitative paradigms govern the methodological procedures applied in research, and influence the discursal and rhetorical conventions of RAs (Cao and Hu, 2014; Cohen *et al.*, 2013; Gray, 2013, 2015; and Hu and Cao, 2015). Hu and Cao (2015) found that the quantitative RAs in the disciplines of applied linguistics, education, and psychology included significantly more stance markers than the qualitative RAs; however, their focus was mostly on single words (e.g. ‘perhaps’). Similarly, Gray (2013, 2015) found that nominalisations and passive voice verbs occurred more frequently in quantitative RAs than in qualitative RAs in the disciplines of applied linguistics and political science. Based on these three studies, it was hypothesised that there would be paradigmatic influences on the frequencies of each structural and discursal type of LBs, specifically that the quantitative RAs would include more bundles in each category of LBs than the qualitative RAs. The specific research question addressed in this study is as follows: To what extent are there paradigmatic differences in the structural types and discourse functions of LBs between the quantitative and qualitative RAs in the discipline of education?

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.euppublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

3. Methodology

This section describes the corpora used in this study, the methods used to extract LBs, the functional taxonomy adopted to analyse the sequences, and the statistical analyses.

3.1 Corpora

In order to answer the above stated research question, we created two comparable corpora of RAs in the discipline of education, as shown in Table 1: quantitative RAs in English in international journals (EDU-QUANT-RA); and qualitative RAs in English in international journals (EDU-QUAL-RA).

Following the principle of *tertium comparationis*, defined as shared similarity between texts (Connor, 2004), we sought to construct corpora that were alike in terms of size, subject matter, and date. Accordingly, all the empirical RAs, which analysed data, were extracted from the period of 2011-2015 from three SSCI-indexed journals: the *British Educational Research Journal* (BERJ); *British Journal of Educational Studies* (BJES); and *The Journal of Educational Research* (JER). These three journals can be classified as ‘generalist journals’ that publish RAs from different sub-disciplines of the discipline of education (Gray, 2015: 42). In the EDU-RA corpus (see Table 1), we included both L1 and L2 writers of English, as Swales (2004: 57) encourages the detachment of ‘communicative performance in research settings from mother tongue status per se’ in English-medium international publications. Previous studies have also taken a similar approach to corpus construction by including papers by broadly English-proficient writers (Hu and Wang, 2014; and Römer, 2009). For this study, the

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

references, tables, figures, acknowledgements and appendices were removed from all the RAs.

| | RAs in English-medium international journals (EDU-RA) | |
|--------------------------------------|---|-------------|
| | Quantitative | Qualitative |
| Tokens | 1,001,099 | 1,001,342 |
| Types | 21,087 | 27,275 |
| Type/token ratio (TTR) | 2.12 | 2.73 |
| Standardised type/token ratio (STTR) | 35.35 | 40.88 |
| No. of RAs | 142 | 141 |
| Average length of RAs in tokens | 7,050 | 7,102 |

Table 1: Corpora used in this study.

At the next stage, RAs were coded according to their research methodology paradigms, as in previous studies (Gray, 2013, 2015; and Hu and Cao, 2015). This coding procedure informed our corpus design, which in turn increased the likelihood of representing quantitative and qualitative RAs in the discipline of education in our corpora (see Biber, 1993; Gray, 2015; and McEnery *et al.*, 2006). The first author selected 30 RAs using a random number generator and coded them according to Creswell's (2009) descriptions of quantitative, qualitative, and mixed methods research paradigms in education research after reviewing the methods section of each article. In cases where the RAs did not fit clearly into one of these three categories, they were coded as 'other'. A PhD candidate in the discipline of education then independently coded the same RAs. The results of the two coding procedures were the same, except for one article; in light of this degree of agreement, the first author coded all the other RAs into the abovementioned categories. Finally, a balanced number of quantitative and qualitative RAs from each journal were selected for inclusion in the two corpora. All

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

RAs that were coded as ‘mixed methods’ or ‘other’ were excluded from our data. The two corpora of RAs included approximately the same number of tokens and texts, in order to enhance the comparability of the corpora.

3.2 Data analysis

3.2.1 Extraction of lexical bundles

This study examined only four-word bundles in the two corpora of RAs, since in English ‘many four-word bundles hold three-word bundles in their structures’ (Cortes, 2004: 401). Five-word and longer LBs are comparatively rare, and they often subsume shorter sequences (Hyland, 2012). Previous studies have also restricted their analysis to four-word LBs in English (e.g., Ädel and Erman, 2012; Cortes, 2004; and Hyland, 2008).

The two other criteria that were applied to extract LBs were frequency threshold and range. The frequency cut-off point for four-word LBs in English has ranged from 10 to 40 times per million words (pmw) in the literature (Hyland, 2012). Following Cortes (2013), the cut-off point was set to 20 times pmw for four-word LBs in this study. However, it should be noted that the frequency cut-off points for identification of LBs are ‘somewhat arbitrary’ (Biber *et al.*, 2004: 376). LBs were extracted, using WordSmith Tools Version 6.0 (Scott, 2012). Regarding the range criterion, a sequence must occur in at least five different texts in the corpus (Biber *et al.*, 2004; and Cortes, 2013), or 10% of the texts (Hyland, 2008) to mitigate against idiosyncratic usage by

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

individual authors. In this study, the cut-off point for range was set at 5% of the texts, i.e., at least in seven texts in each corpus.

After we had identified the LBs, those containing proper nouns were manually removed (e.g., ‘in England and Wales’), as in previous studies (e.g., Ädel and Erman, 2012). Similarly, overlapping LBs with the same frequency were combined as appropriate, as in the case of ‘the extent to which’ and ‘extent to which the’. In other cases, when two LBs formed part of a five-word bundle, a subtractive method was applied (Chen and Baker, 2010). For instance, when ‘at the end of’ and ‘the end of the’ were subsumed into ‘at the end of the’, the frequency of ‘at the end of the’ was subtracted from that of these two four-word LBs to determine their exact frequency.

3.2.2 Analysis of the structural types of lexical bundles

The lexical bundles were classified structurally, drawing on the taxonomy of previous studies (Biber *et al.* 1999; Biber *et al.*, 2004; and Chen and Baker, 2010, 2016). The categories with examples from our corpora are as follows:

1. Prepositional phrase-based (PP-based) bundles, which include phrases that begin with a preposition followed by a noun phrase.

1.a. Prepositional phrase with embedded of-phrase fragment: *as a result of, at the end of*

1.b. Other prepositional phrase: *at the same time*

2. Noun phrase-based (NP-based) bundles, which refer to noun phrases with of-phrase fragments or other post-modifier fragments.

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.euppublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

- 2.a. Noun phrase with of-phrase fragment: *the nature of the, the size of the*
- 2.b. Noun phrase with other post-modifier fragment: *the fact that the, the difference between the*
3. Verb phrase-based (VP-based) bundles refer to any phrases with a verb component.
 - 3.a. (Noun phrase +) copula be + noun phrase/adjectival phrase: *is one of the, there is a need*
 - 3.b. Verb phrase with active verb: *research has shown that, has the potential to*
 - 3.c. Anticipatory it + verb phrase/adjectival phrase + (complement clause)/to-clause fragment: *it is possible to, it is clear that*
 - 3.d. Passive verb + (prepositional phrase fragment): *are presented in table, can be seen as*
 - 3.e. That-clause fragment: *that there is a*
4. Others: *as well as the*

All of the LBs identified in this study were coded according to this structural taxonomy.

3.2.3 Analysis of the discourse functions of lexical bundles

Several functional taxonomies have been developed to describe the discourse functions of LBs (e.g., Biber *et al.*, 2004; and Hyland, 2008). Biber *et al.* (2004) proposed a functional taxonomy of LBs, arguing that they fall into three main discursal categories: (1) referential expressions, which specify propositions, frame quantities, time, and place, and introduce abstract and concrete entities and ‘determine [the author’s] way of

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.euppublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

looking at things' (Cortes, 2004: 401); (2) discourse organisers, which establish relationships between different parts of the text and indicate introductory, transitive, causative, and inferential textual relations; (3) stance expressions, which express the writer's (un)certainty and commitment (epistemic stance), and convey their attitudes towards propositions or express obligations, directives or ability (attitudinal/modality stance). Within epistemic stance expressions, 'hedges' mitigate the writer's certainty/commitment towards propositions, while 'boosters' increase the writer's certainty and commitment (Hyland, 2005).

Hyland's (2008) classification of discourse functions of LBs was developed from Biber *et al.*'s taxonomy (2004), and the classification included three main categories: research-oriented, text-oriented, and participant-oriented bundles. As Hyland (2008) noted, these three categories can be considered as equivalent to referential expressions, discourse organisers, and stance expressions, respectively.

As Biber *et al.*'s (2004) taxonomy of discourse functions of LBs has been widely adopted in a large number of studies in the literature (Cortes, 2013), this taxonomy was also used in this study. We adapted the taxonomy in several ways. First, we added the subcategory of descriptive bundles, which signal concrete and abstract entities and qualities (e.g., 'the role of the'), within referential expressions, as in Cortes' study (2004). Second, Hyland's (2008) resultative signals (text-oriented bundles) were incorporated into the subcategory of discourse organisers as inferential/resultative bundles, which signal inferential or causative relations between propositions. Finally, we added the subcategory of subject-specific bundles (Cortes, 2004), which refer to the

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

specific topic of the RAs (education) within the corpus (e.g., ‘teaching and learning strategies’). As subject-specific bundles do not serve as discourse building blocks, we will present the results for the first three categories of LBs. The taxonomy used in this study is presented in Table 2, alongside examples of bundles from our corpora.

| Categories | Bundles |
|---|---|
| 1. Referential expressions | |
| 1.a. Identification/focus | <i>one of the most</i> |
| 1.b. Specifying attributes (quantifying/framing) | <i>in terms of</i> |
| 1.c. Descriptive | <i>the role of the</i> |
| 1.d. Place-/time-/text-deictic | <i>at the end of</i> |
| 2. Discourse organisers | |
| 2.a. Topic introduction/focus | <i>in this paper we</i> |
| 2.b. Topic elaboration/clarification/ transition | <i>on the other hand</i> |
| 2.c. Inferential/resultative signals | <i>as a result of</i> |
| 3. Stance expressions | |
| 3.a. Epistemic stance | <i>more likely to be</i> |
| 3.b. Attitudinal/modality stance | <i>it is important to/to be able to</i> |
| 4. Others (Subject-specific bundles) | <i>for free school meals</i> |

Table 2: The functional taxonomy of lexical bundles.

Several researchers have noted the multi-functionality of LBs, and the functional overlap between categories (Ädel and Erman, 2012; Biber *et al.*, 2004; Chen and Baker, 2010; and Cortes, 2004). Therefore, while coding the functions, context was taken into consideration. Accordingly, when LBs had multiple functions, the predominant function of each LB was coded after the analysis of concordance lines (Biber *et al.*, 2004; and Chen and Baker, 2016). In order to ensure consistency in coding, two researchers separately coded the discourse functions of all the identified LBs. Where the two coders

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

disagreed, the relevant LBs were sent to a third coder. When agreement was obtained between two of the three coders, that category was considered to be the discourse function.

After coding all the structural types and discourse functions of LBs, a script was written in the Python language to record the token frequencies of each category of LBs (i.e., referential expressions, discourse organisers, stance expressions, NP-based, PP-based, and VP-based bundles) for each text (per 1,000 words of each text). We used robust inferential methods; as in most cases these provide more statistical power and accuracy in comparison to parametric and non-parametric tests when there are outliers and non-normal data (Larson-Hall and Herrington, 2010), as was the case in this study. One such robust method involved bootstrapping, which is a procedure where many samples are taken from the observed data to estimate a sampling distribution from which inferential analysis can be conducted. For non-normal data, bootstrapped analyses ‘are much more powerful and more likely to either detect statistical significance when present or to reveal a statistical relationship as spurious’ (Plonsky *et al.*, 2015: 605). Bootstrapped confidence intervals (CIs) for the differences in means for each category of LBs between the quantitative and qualitative RAs were calculated using the bias-corrected-and-accelerated (BCa) method¹, based on 10,000 non-parametric bootstraps, with replacement from our data within each of the groups. These were calculated using the `boot.ci` function in the `boot` package (see Laflair *et al.*, 2015) in R (R Core Team, 2017). We also performed a bootstrapped independent samples t-

¹ This method corrects for both bias and skewness in the distribution, and is generally more accurate than other methods (see Laflair *et al.*, 2015).

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

test on the basis of 10,000 bootstraps using the *yuenbt* function² in the WRS2 package, without any trimming of means (Wilcox, 2016). The alpha level was set at .05.

4. Results

In total, 108 LBs (4,459 tokens) in the EDU-QUANT-RA corpus and 80 LBs (3,275 tokens) in the EDU-QUAL-RA corpus were coded for their structural types and discourse functions after the refinement of bundles. The results show that LBs were more frequent, both in terms of types and tokens, in the quantitative RAs than in the qualitative RAs. This section reports the quantitative and qualitative analyses of the structural types and discourse functions of the LBs across the two corpora.

4.1 The structural types of lexical bundles

The quantitative and qualitative RAs in the discipline of education contained more PP-based bundles than NP-based or VP-based bundles, as can be seen from Figure 1. While PP-based bundles were followed by VP-based bundles and then NP-based bundles in the quantitative RAs, PP-based bundles were followed by NP-based bundles and then VP-based bundles in the qualitative RAs in terms of frequency.

² This function uses a bootstrap-t method, which is more effective than other methods when means are not trimmed (Wilcox, 2016).

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

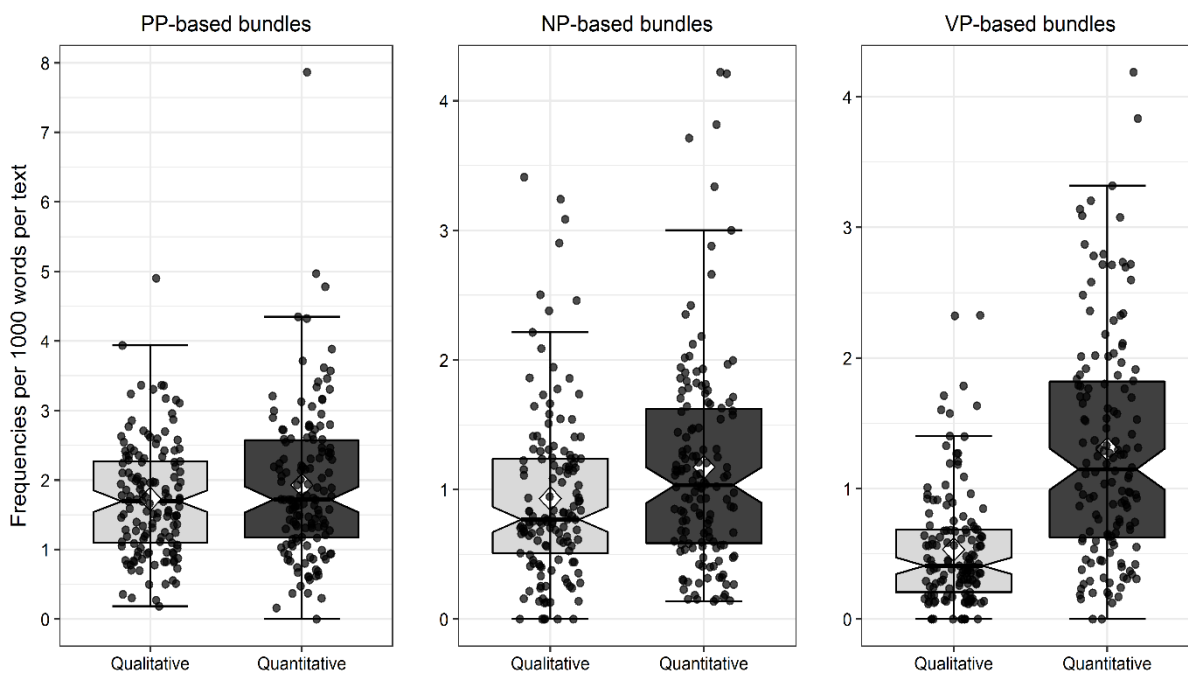


Figure 1³: The distribution of structural types of lexical bundles across the two corpora.

4.1.1 PP-based bundles

There was no statistically significant difference between the quantitative and qualitative RAs with regard to the frequency of PP-based bundles, $t = 1.22$, $p = 0.22$, 95% CI [-0.02, 0.42], $\xi = .11^4$. Furthermore, 66%⁵ and 70% of the PP-based bundles were formed with embedded of-phrase fragments in the EDU-QUANT-RA and EDU-QUAL-RA corpora, respectively.

³ In Figures 1 and 2, \diamond indicates mean frequencies.

⁴ As an effect size, we report ξ , a robust analog of Cohen's d , which allows heteroscedasticity in the data. This explanatory measure of effect size was proposed by Wilcox and Tian (2011), and the values 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes, respectively (Wilcox and Tian, 2011).

⁵ All the subsequent percentages are reported for token frequencies.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

Across the two corpora, 19 shared PP-based bundles were identified. Four of these ('on the other hand', 'on the one hand', 'as a result of', 'in addition to the') indicated relationships between prior and coming discourse in the RAs. Others (e.g., 'in terms of the', 'in the case of', 'in relation to the', 'in the context of') specified the entities that the authors were referring to in their RAs. Additionally, 22 and 21 PP-based bundles were distinctive⁶ for the quantitative and qualitative RAs, respectively.

| EDU-QUANT-RA | Raw freq. | No of texts | EDU-QUAL-RA | Raw freq. | No of texts |
|----------------------|-----------|-------------|-----------------|-----------|-------------|
| in the present study | 123 | 43 | at the heart of | 49 | 32 |
| for each of the | 59 | 37 | as a means of | 43 | 26 |
| over the course of | 50 | 11 | in the light of | 40 | 25 |
| at the beginning of | 48 | 28 | on the part of | 40 | 25 |
| at the start of | 44 | 30 | as a form of | 39 | 24 |

Table 3: Top-five most frequent distinctive PP-based bundles in the two corpora.

Table 3 shows the top-five most frequent distinctive bundles in the two corpora; these distinctive bundles reveal the knowledge-making practices specific to quantitative and qualitative research. The distinctive PP-based bundles in the EDU-QUANT-RA corpus referred to the study being reported on in order to formulate hypotheses, data, methods, and statistical techniques ('in the present study'), the duration and time of the study (e.g., 'over the course of'), and specified entities ('for each of the') for which arguments were presented, as the examples illustrate below.

(1) Included within oral language is vocabulary breadth, which, although not included *in the present study* as a predictor, forms a subscale within the WOLD criteria. (EDU-QUANT-RA100.JER2013)

⁶ The distinctive bundles in the results section refer to ones that were unique to the EDU-QUANT-RA or EDU-QUAL-RA corpus.

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.euppublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

(2) Pooling the two surveys, we estimated probit specifications *for each of the* indicators... (EDU-QUANT-RA106.JER2013)

(3) ...students who display stable school engagement *over the course of* their school careers are less likely to drop out. (EDU-QUANT-RA9.BERJ2012)

(4) ... the data are collected *at the beginning of* schooling and again at the end of the first year allowing analyses of progress...(EDU-QUANT-RA30.BERJ2015)

(5) This ... parallels the results of an earlier analysis of the relationship between young people's intentions for post-16 participation expressed *at the start of* secondary school. (EDU-QUANT-RA126.BJES2013)

The distinctive PP-based bundles in the EDU-QUAL-RA corpus indicated the importance of the topic or findings ('at the heart of'), put interpretations into perspective (e.g., 'in the light of'), and elaborated on arguments (e.g., 'as a means of'), as the examples show below.

(6) *At the heart of* this issue lies uncertainty about the nature and status of 'expertise' in interprofessional settings. (EDU-QUAL-RA53.BJES2012)

(7) Teachers of qualitative methods also use active learning *as a means of* making the research process visible. (EDU-QUAL-RA26.BJES2014)

(8) ...it seems salient to revisit some of the issues she raised, particularly *in the light of* UK Coalition education policy... (EDU-QUAL-RA23.BERJ2013)

(9) ...they elaborate how this would require additional explanation *on the part of* the teacher and engagement with students' learning processes. (EDU-QUAL-RA21.BERJ2014)

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

(10) Bourdieu therefore casts dispositions *as a form of embodied cultural capital*...(EDU-QUAL-RA28.BJES2014)

4.1.2 NP-based bundles

NP-based bundles were significantly more frequent in the quantitative RAs than in the qualitative RAs, $t = 2.70$, $p = 0.008$, 95% CI [0.08, 0.42], $\zeta = .24$. NP-based bundles that were formed with of-phrase fragments constituted 76% and 61% of the NP-based bundles in the EDU-QUANT-RA and EDU-QUAL-RA corpora, respectively.

There were 12 shared NP-based bundles between the EDU-QUANT-RA and EDU-QUAL-RA corpora. Most of these referred to concepts (e.g., ‘the role of the’, ‘the nature of the’) and framed propositions (e.g., ‘the ways in which’) that were introduced in the articles. Whereas 20 distinctive NP-based bundles were found in the EDU-QUANT-RA corpus, only 10 were identified in the EDU-QUAL-RA corpus.

| EDU-QUANT-RA | Raw freq. | No of texts | EDU-QUAL-RA | Raw freq. | No of texts |
|---------------------------------|-----------|-------------|---------------------------|-----------|-------------|
| the results of the | 85 | 44 | one of the most | 34 | 25 |
| the effect of the | 45 | 26 | the development of a | 32 | 23 |
| the use of the | 36 | 12 | the rest of the | 28 | 18 |
| the effects of the ⁷ | 35 | 26 | the development of the | 25 | 21 |
| the results of this | 35 | 26 | the implementation of the | 25 | 11 |

Table 4: Top-five most frequent distinctive NP-based bundles in the two corpora.

As shown in Table 4, the distinctive bundles in the EDU-QUANT-RA corpus served to report the findings of the studies (‘the results of the’, ‘the results of this’), signalled cause-effect relations revealed in the studies (e.g., ‘the effect of the’), and

⁷ ‘The effect of the’ and ‘the effects of the’ were identified as two different LBs since our corpus was not lemmatised, and the identification of LBs was based on orthographic word units (Biber *et al.*, 1999).

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

indicated characteristics of the data, methods, and statistical procedures (‘the use of the’). The following examples illustrate these functional realisations:

(11) ...*the results of the* study indicated that inquiry-based science instruction had an effect on science achievement as well as on interest in science. (EDU-QUANT-RA52. JER2012)

(12) ...*the effect of the* academic composition is significantly larger in schools from the Dutch-speaking community...(EDU-QUANT-RA123.BJES2014)

(13) Training involved project leads demonstrating *the use of the* running record, coding and analysis. (EDU-QUANT-RA36.BERJ2015)

(14) For that reason, this group’s gains cannot be confidently attributed to *the effects of the* course. (EDU-QUANT-RA110.BERJ2014)

In the EDU-QUAL-RA corpus, the distinctive NP-based bundles established the centrality of themes explored in the studies (‘one of the most’), explained processes (e.g. ‘the development of a’), and referred to groups and participants (‘the rest of the’), as can be seen in the following examples.

(15) Superhero play has indeed been identified by other studies as *one of the most* favourite play themes. (EDU-QUAL-RA106. BERJ2013)

(16) Similarly, *the development of a* mathematics-teaching identity is ongoing. (EDU-QUAL-RA123.JER2012)

(17) However, they also raised their hand more often to answer questions, relative to *the rest of the* class. (EDU-QUAL-RA20.BERJ2014)

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

(18) At that point and into 2012, further structural changes were taking place, as well as *the implementation of the toughest cuts to public sector expenditure*. (EDU-QUAL-RA47.BJES2013)

4.1.3 VP-based bundles

The quantitative RAs included significantly more VP-based bundles than the qualitative RAs, $t = 8.54$, $p < 0.001$, 95% CI [0.61, 0.93], $\zeta = .69$. The two corpora shared only seven common VP-based bundles ('are more likely to', 'it is important to', 'more likely to be', 'that there is a', 'it is possible to', 'can be seen as', and 'it is clear that'). These bundles introduced a proposition ('that there is a'), indicated tentativeness, possibility (e.g., 'it is possible to'), or certainty ('it is clear that'), and highlighted the importance of claims ('it is important to') that were put forward in the articles. 47% and 54% of the VP-based bundles were anticipatory *it* clauses or *to*-clause fragments in the EDU-QUANT-RA and EDU-QUAL-RA corpora, respectively. Whereas 28% of the VP-based bundles in the EDU-QUANT-RA corpus were passive verbs + (prepositional phrase fragments), this figure was just 14% in the EDU-QUAL-RA corpus.

Table 5 shows the top-five most frequent distinctive VP-based bundles in the two corpora. While there were 20 distinctive VP-based bundles in the EDU-QUANT-RA corpus, only seven distinctive VP-based bundles were found in the EDU-QUAL-RA corpus.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

| EDU-QUANT-RA | Raw freq. | No of texts | EDU-QUAL-RA | Raw freq. | No of texts |
|------------------------|-----------|-------------|----------------------|-----------|-------------|
| were more likely to | 68 | 41 | to be able to | 47 | 34 |
| are presented in table | 53 | 36 | to make sense of | 30 | 15 |
| are less likely to | 45 | 29 | there has been a | 29 | 25 |
| it should be noted | 43 | 19 | be seen as a | 24 | 21 |
| it is possible that | 42 | 33 | has the potential to | 21 | 17 |

Table 5: Top-five most frequent distinctive VP-based bundles in the two corpora.

In the EDU-QUANT-RA corpus, as shown in the examples below, the distinctive bundles were used to report the study findings and statistical models used (‘were more likely to’, ‘are less likely to’), avoid complete commitment to a proposition (‘it is possible that’), refer to data and numbers in tables (‘are presented in table’), and to highlight the limitations of the studies or emphasise the author’s point (‘it should be noted’).

(19) ...rural/island local authority classroom assistants *were more likely to* assess learning and more likely to be unsupervised. (EDU-QUANT-RA120.BERJ2014)

(20) Results of this model *are presented in Table 6*. (EDU-QUANT-RA108.JER2014)

(21) Evidence on the effect of part time workers is mixed and, unlike several other European countries, those in the UK *are less likely to* be trained. (EDU-QUANT-RA125.BJES2013)

(22) *It should be noted* that ... no conclusions are drawn as to the relative ‘impact’ of different outputs based on these metrics. (EDU-QUANT-RA127.BJES2012)

(23) ...*it is possible that* other types of trajectories may be found in other samples. (EDU-QUANT-RA107.JER2014)

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit [https://www.eupublishing.com/loi/cor](https://www.euppublishing.com/loi/cor) for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

The distinctive bundles in the EDU-QUAL-RA corpus expressed ability ('to be able to'), presented the aims of the study ('to make sense of'), introduced a proposition ('there has been a'), and offered interpretations of claims made in the articles (e.g., 'be seen as a'), as illustrated in the examples below.

(24) ...those enjoying greater affluence tend to believe more that people ought *to be able to* convert class/income advantage into educational advantage for their children. (EDU-QUAL-RA41.BJES2013)

(25) The next section will examine more deeply both Dutch and British education context in order *to make sense of* the perception of the students. (EDU-QUAL-RA12.BERJ2014)

(26) ...*there has been a* widespread concern among both academics and practitioners about student engagement... (EDU-QUAL-RA23.BJES2014)

(27) It was consistently observed that an emphasis on assessed work often came at the expense of commitment to CAS, which could *be seen as a* 'box-ticking exercise'. (EDU-QUAL-RA25.BJES2014)

(28) ...the ACCT framework *has the potential to* support professional development...(EDU-QUAL-RA18.BERJ2014)

4.2 The discourse functions of lexical bundles

Both the quantitative and qualitative RAs in the discipline of education included referential expressions, which were followed by stance expressions and discourse organisers in terms of frequency, as shown in Figure 2.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

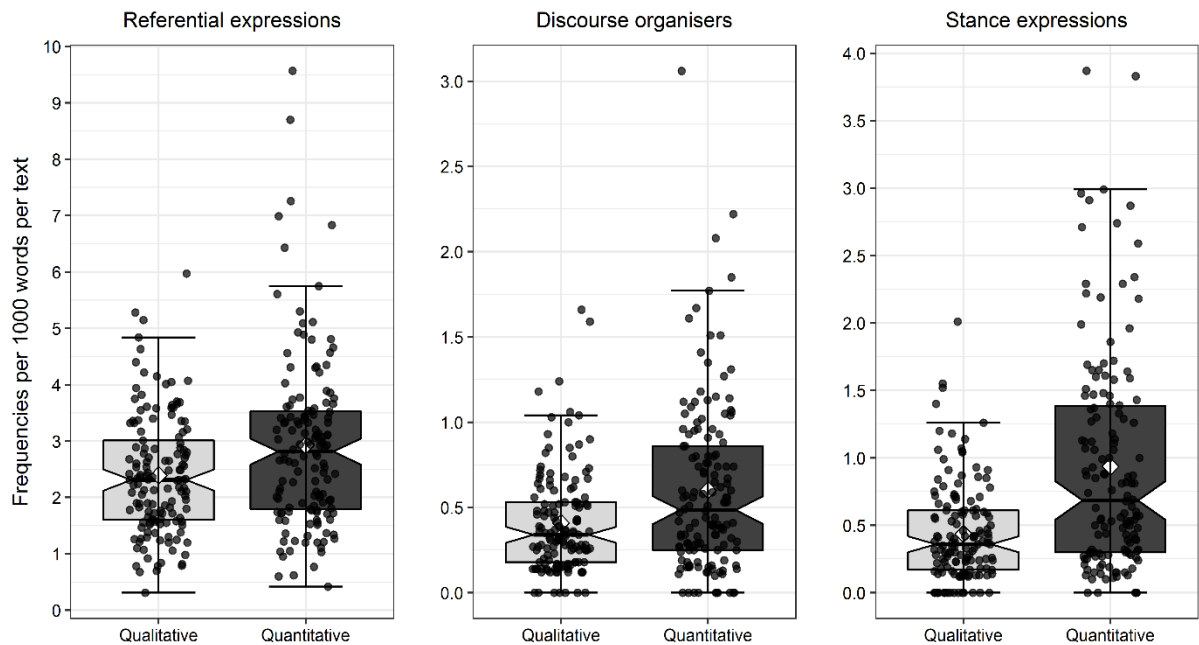


Figure 2: The distribution of discursal categories of lexical bundles across the two corpora.

4.2.1 Referential expressions

Referential expressions were significantly more frequent in the EDU-QUANT-RA corpus than in the EDU-QUAL-RA corpus, $t = 2.89$, $p = 0.004$, 95% CI [0.08, 0.42], $\xi = .25$. There were 25 referential expressions shared across the two corpora. The majority of these referential expressions served to frame propositional content (e.g., ‘on the basis of’, ‘in the context of’, ‘in relation to the’), and referred to entities or concepts (e.g., ‘the role of the’, ‘the nature of the’). Forty-nine and thirty-one distinctive bundles were identified in the EDU-QUANT-RA and EDU-QUAL-RA corpora, respectively.

As seen in Table 6, the distinctive referential expressions in the EDU-QUANT-RA corpus served to report the findings of the study (‘the results of the’ – see example

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

11), provide details of data, methods and statistical procedures (‘for each of the’ – see example 2 and ‘are presented in table’ – see example 20), and refer to the duration of the study (e.g., ‘at the beginning of’ – see examples 3 and 4). On the other hand, the distinctive bundles in the EDU-QUAL-RA corpus were primarily used to frame arguments and put them into perspective (e.g., ‘in the light of’ – see examples 8, 9 and 10), as shown in the examples below.

(29) Similar sentiments about the moral and ‘real world’ value of hardship, work ethic and success *in the face of* considerable odds are echoed by other participants... (EDU-QUAL-RA112.BERJ2012)

(30) A child’s picture might be interpreted *as part of a* ‘child-led activity’, and used to follow up learning... (EDU-QUAL-RA91.BJES2012)

| EDU-QUANT-RA | Raw freq. | No of texts | EDU-QUAL-RA | Raw freq. | No of texts |
|------------------------|-----------|-------------|-----------------|-----------|-------------|
| the results of the | 85 | 44 | in the light of | 40 | 25 |
| for each of the | 59 | 37 | on the part of | 40 | 25 |
| are presented in table | 53 | 36 | as a form of | 39 | 24 |
| over the course of | 50 | 11 | in the face of | 38 | 25 |
| at the beginning of | 48 | 28 | as part of a | 34 | 29 |

Table 6: Top-five most frequent distinctive referential expressions in the two corpora.

The majority of referential expressions in both corpora specified attributes (54% - EDU-QUANT-RA corpus; 64% - EDU-QUAL-RA corpus). The quantitative RAs made frequent use of quantifying bundles (e.g., ‘a large number of’, ‘a higher proportion of’, ‘a high level of’), whereas the qualitative RAs included more framing bundles (e.g., ‘within the context of’, ‘in a way that’, ‘as a way of’). Descriptive bundles in both corpora (24% - EDU-QUANT-RA corpus; 17% - EDU-QUAL-RA corpus) referred to entities or abstract constructs (e.g., ‘the results of this’, ‘the effects of the’ –

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

EDU-QUANT-RA corpus; ‘the development of the’, ‘an analysis of the’ – EDU-QUAL-RA corpus). The descriptive bundles that referred to results and analyses in the RAs (e.g., ‘the results of the’, ‘an analysis of the’) were primarily referential expressions; however, their secondary function was to construct implicit stance. Although these bundles helped the authors to retain impersonality on the surface (Charles, 2006), they also played a role in the construction of knowledge claims and implicitly expressed stance (see example 31). The same phenomenon occurred to a greater extent in the EDU-QUANT-RA corpus (e.g., ‘the results of the’, ‘the results of this’) than in the EDU-QUAL-RA corpus (‘an analysis of the’).

(31) ...*the results of this* study provide important implications for educational research and practice. (EDU-QUANT-RA25.BERJ2015)

The place-/time-/text-deictic bundles occurred more frequently in the quantitative RAs (19%) than in the qualitative RAs (11%), due to the references to tables, visualisations, and time (e.g., ‘are shown in table’, ‘at the beginning of’). Identification and focus bundles (e.g., ‘one of the most’, ‘is one of the’) were the least frequently used referential expressions in both corpora (3% - EDU-QUANT-RA corpus; 8% EDU-QUAL-RA corpus). These bundles were used to claim the centrality of the topic (see Cortes, 2013), review previous prominent studies, and identify the characteristics of the study findings, and these were slightly more frequent in the qualitative RAs than in the quantitative RAs.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

4.2.2 Discourse organisers

The quantitative RAs included significantly more discourse organisers than the qualitative RAs, $t = 3.28$, $p = 0.001$, 95% CI [0.10, 0.29], $\zeta = .31$. Both corpora included discourse organisers, which functioned as transition markers (‘on the other hand’, ‘as well as the’, ‘on the one hand’ and ‘in addition to the’) and inferential signals (‘as a result of’). Only four distinctive discourse organisers were found in the EDU-QUAL-RA corpus, as shown in Table 7. In the EDU-QUANT-RA corpus, eight distinctive bundles were identified.

| EDU-QUANT-RA | Raw freq. | No of texts | EDU-QUAL-RA | Raw freq. | No of texts |
|-----------------------|-----------|-------------|--------------------|-----------|-------------|
| in the present study | 123 | 43 | as a means of | 43 | 26 |
| was found to be | 31 | 22 | to make sense of | 30 | 15 |
| been shown to be | 30 | 23 | in contrast to the | 20 | 18 |
| the present study was | 30 | 20 | when it comes to | 20 | 18 |
| in this paper we | 27 | 13 | - | - | - |

Table 7: Top-five most frequent distinctive discourse organisers in the two corpora.

In the EDU-QUANT-RA corpus, the distinctive discourse organisers were either topic introduction/focus bundles (e.g., ‘in the present study’ – see example 1) that served to introduce a certain aspect of the study, or inferential/resultative signals that were used to present the main conclusions of the study and the author’s inferences from those conclusions (e.g., ‘was found to be’), as the examples show below. A passive construction helped the writer to downplay their role and draw conclusions with an objective tone, as can be seen in example 32.

(32) Although some high school students were mobile, nonattendance *was found to be* more of an issue in these grade levels. (EDU-QUANT-RA53.JER2012)

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

(33) These differences have *been shown to be* strongly associated with outcomes such as types of employment and future income... (EDU-QUANT-RA126.BJES2013)

(34) ...*the present study was* limited by the grade range because the data available did not go beyond Grade 2. (EDU-QUANT-RA73.JER2013)

(35) *In this paper we* report our attempts to measure the students' experiences of transition... (EDU-QUANT-RA4.BERJ2012)

The two distinctive discourse organisers functioned as topic elaborations ('as a means of' – see example 7 and 'to make sense of' – see example 25). The other two discourse organisers were transition markers, as the examples illustrate below.

(36) *In contrast to the* 1998 arrangements, students were eligible for income-contingent loans to cover tuition fees in addition to loans for maintenance. (EDU-QUAL-RA118.BJES2015)

(37) *When it comes to* music, which is often an important reference point for people of this age, the music culture of Roma girls was at odds with mainstream music... (EDU-QUAL-RA119.BJES2015)

There were notable differences between the token frequencies of the subcategories of discourse organisers between the two corpora. While the token frequencies of topic elaboration/clarification/transition markers (e.g., 'on the other hand') accounted for 46% of the discourse organisers in the EDU-QUANT-RA corpus, this proportion was 87% in the EDU-QUAL-RA corpus. Furthermore, 33% of the discourse organisers were inferential/resultative signals ('been shown to be') through which the authors drew conclusions for the study in the EDU-QUANT-RA

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

corpus, whereas only one bundle, ‘as a result of’ (13% of the discourse organisers), was categorised as an inferential/resultative signal in the EDU-QUAL-RA corpus. No topic introduction/focus bundle was found in the EDU-QUAL-RA corpus. On the other hand, in the EDU-QUANT-RA corpus, 21% of the discourse organisers (e.g., ‘in this paper we’) belonged to this subtype.

4.2.3 Stance expressions

Stance expressions were significantly more frequent in the quantitative RAs than in the qualitative RAs, $t = 5.21$, $p < 0.001$, 95% CI [0.36, 0.65], $\zeta = .51$. There were eight shared stance expressions between the two corpora. These bundles withheld the author’s commitment to a proposition (e.g., ‘it is possible to’), indicated certainty (e.g., ‘the fact that the’), or emphasised the importance of a proposition (e.g., ‘it is important to’).

Thirteen and seven distinctive stance expressions were identified in the EDU-QUANT-RA and EDU-QUAL-RA corpora, respectively.

| EDU-QUANT-RA | Raw freq. | No of texts | EDU-QUAL-RA | Raw freq. | No of texts |
|---------------------|-----------|-------------|------------------------|-----------|-------------|
| were more likely to | 68 | 41 | at the heart of | 49 | 32 |
| are less likely to | 45 | 29 | to be able to | 47 | 34 |
| it should be noted | 43 | 19 | be seen as a | 24 | 21 |
| it is possible that | 42 | 33 | important to note that | 21 | 16 |
| more likely to have | 40 | 18 | it could be argued | 21 | 15 |

Table 8: Top-five most frequent distinctive stance expressions in the two corpora.

As shown in Table 8, most of the distinctive stance expressions functioned as hedges that downplayed the certainty of knowledge claims and interpretations of the results (see examples 19, 21, 23 and 38), or emphasised a certain aspect of interpretations offered in the study (‘it should be noted’- see example 22). In the EDU-

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit [https://www.eupublishing.com/loi/cor](https://www.euppublishing.com/loi/cor) for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

QUAL-RA corpus, the distinctive stance expressions were used to indicate the importance of the topic ('at the heart of'- see example 6), express ability ('to be able to' – see example 24), provide clarifications in the text ('important to note that'- see example 39) and offer interpretations of findings presented in the study ('be seen as a' – see example 27 and 'it could be argued that' – see example 40).

(38) ...multi-component approaches were significantly *more likely to have* encountered implementation problems... (EDU-QUANT-RA113.BERJ2014)

(39) *It is important to note that* we used this instrument to assess what events the students recalled and not their levels of knowledge or understanding about these events. (EDU-QUAL-RA124.JER2012)

(40) *It could be argued* that increasing the number of publications is akin to a 'zero sum game'... (EDU-QUAL-RA103.BERJ2012)

With regard to the subtypes of discourse organisers, the EDU-QUANT-RA corpus was markedly different from the EDU-QUAL-RA corpus in that 77% of the stance expressions, including 'it is likely that', 'more likely to be', and 'it is possible that', served as epistemic stance expressions, and 92% of those epistemic stance expressions were hedges that withheld commitment to a proposition. In the EDU-QUAL-RA corpus, on the other hand, only 36% of stance expressions were categorised as epistemic stance expressions.

The other key difference between the two corpora concerned attitudinal stance (e.g., 'it is important to', 'there is a need'). Unlike the EDU-QUANT-RA corpus, the majority of stance expressions (64%) expressed an attitudinal stance in the EDU-

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit [https://www.eupublishing.com/loi/cor](https://www.euppublishing.com/loi/cor) for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

QUAL-RA corpus. Except for ‘to be able to’, which expressed ability, all the other attitudinal stance expressions in the EDU-QUAL-RA corpus conveyed writers’ attitudes towards propositions.

5. Discussion

In this section, we discuss paradigmatic influences on the structural types and discourse functions of LBs in RAs in the discipline of education.

Paradigmatic influences on the frequencies of each structural and discursive type of LBs were observed in RAs in the discipline of education, except for PP-based bundles. The knowledge-making practices in quantitative and qualitative research can provide explanations for these influences.

There was no paradigmatic influence on the frequencies of PP-based bundles in the discipline of education. However, there was a difference between the two corpora in regard to the use of PP-based bundles. The quantitative RAs contained PP-based bundles (e.g., ‘over the course of’) that referred to the duration and time of the study being discussed, and to figures and tables presented in the RAs, since quantitative research includes time as a variable and makes frequent use of figures and tables for data presentation (Cohen *et al.*, 2013). On the other hand, the qualitative RAs made use of PP-based bundles (‘in the light of’) that contextualised cases and situations as qualitative research is context-bound, and thus involves description of the research context (Creswell, 2009).

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit [https://www.eupublishing.com/loi/cor](https://www.euppublishing.com/loi/cor) for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

The more frequent use of referential expressions in quantitative RAs can be explained in reference to the procedures that researchers follow, including formulating the theory/hypothesis to be tested, describing statistical testing, and reporting on whether the theory/hypothesis is supported (Cohen *et al.*, 2013). Hence, reporting these procedures is likely to require frequent references to results, findings, analytical procedures, and concrete and abstract entities in RAs. This also explains the higher frequency of NP-based bundles in the quantitative RAs. This finding is in line with that of Gray's (2015) study, which revealed that quantitative RAs included more nominalisations, process and abstract nouns than qualitative RAs. Qualitative research, on the other hand, 'has unique steps in data analysis, and draws on diverse strategies of inquiry' (Creswell, 2009: 206). These diverse strategies may be reflected in the more infrequent use of referential expressions and NP-based bundles in the qualitative RAs.

Quantitative research identifies relations of similarity and difference through comparisons of empirical results, and cause-effect relations in knowledge claims (Creswell, 2009); therefore, discourse organisers were more frequent in the quantitative RAs than in the qualitative RAs. Positivism, which underlies quantitative research, requires description of the relationships between variables and causality (Creswell, 2009); discourse organisers serve this purpose. Additionally, inferential/resultative bundles (e.g., 'was found to be', 'been shown to be'), which served to report the results and statistical findings, were mostly used with a passive construction. This suggests that quantitative research in the discipline of education tends to lean towards the positivist epistemology, which involves knowledge production through observable experiences

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

and facts. As Hyland (2005: 147) notes, in the positivist epistemology ‘the authority of the individual is secondary to the authority of the text and facts should be allowed to “speak for themselves”’. Hence, the authors of the quantitative RAs seemed to disguise their presence, using inferential/resultative bundles in passive constructions, which served as objectivisation.

Paradigmatic influences were greatest for stance expressions and VP-based bundles, most of which were stance expressions, in that the quantitative RAs of the EDU-RA corpus contained more stance expressions than the qualitative RAs, and the bundles that functioned as hedges accounted for 92% of the epistemic stance expressions. A possible explanation for this is that quantitative research requires alternative explanations for findings, and a discussion of the generalisability of the results to a whole population (Cohen *et al.*, 2013). These two premises may lead to the more frequent use of hedges to fulfil the aims of quantitative research. On the other hand, qualitative research seeks to understand experiences, and generalisability of the research findings is not a concern (Cohen *et al.*, 2013). Therefore, EDU-RA writers of qualitative research may not need to use epistemic stance expressions frequently in their articles. This finding supports that of Hu and Cao (2015), who found that quantitative RAs in the disciplines of applied linguistics, psychology, and education made more frequent use of hedges than the qualitative RAs in the same disciplines. In qualitative RAs, on the other hand, most of the stance expressions expressed writers’ attitudes towards propositions (‘at the heart of’, ‘there is a need’) since qualitative research often

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

emphasises the importance of understanding cases within a particular social context (Creswell, 2009).

Based on the distinctive bundles identified in quantitative and qualitative RAs in this study, it can be argued that qualitative education RAs seem to lean towards the knower code, which relies on discursive and evaluative interpretations, while quantitative education RAs tend to be more oriented towards the knowledge code, which emphasises empirical enquiry and research procedures (see Maton, 2007). Hence, it can be said that paradigm plays a role in creating the ‘dynamic tension between the science and the social in the discourse’ of social science disciplines (Wignell, 2007: 202).

6. Conclusion

The findings of this study confirmed our hypothesis that paradigm influences the frequencies of each discursal and structural type of LBs in English academic writing in the discipline of education, except in the case of PP-based bundles. The distinctive bundles identified in the quantitative and qualitative RAs shed new light on the intra-disciplinary variations in knowledge-making practices in the discipline of education. In the discipline of education, research paradigm partly shapes disciplinary knowledge-making practices, which are argued to be ‘contextually-contingent’ (Trowler, 2014: 1723).

The present study has further revealed that quantitative RAs made more frequent use of lexical bundles than qualitative RAs in the discipline of education. The reliance

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit [https://www.eupublishing.com/loi/cor](https://www.euppublishing.com/loi/cor) for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

on NP-based bundles, most of which served as referential expressions, and VP-based bundles, most of which functioned as stance expressions was greater in the quantitative RAs than in the qualitative RAs. These characteristics of quantitative RAs are attributed to knowledge-making practices that value explicit presentation of research procedures, and discussion of the generalisability of findings. These results have implications for corpus building and design in academic discourse. This study has shown that research paradigm is associated with intra-disciplinary variations in the use of LBs. Therefore, in social sciences, in particular, using a balanced number of RAs with quantitative and qualitative research paradigms would increase the representativeness of a corpus, and enable researchers to investigate intra-disciplinary variations in terms of lexico-grammatical characteristics (see Gray, 2013, 2015).

The results of this study also have important implications for academic writing practices and the teaching of academic writing in English. A more specific approach to teaching phraseological patterns for academic writing may need to be utilised in relation to research paradigms in social science disciplines. The heavy presence of bundles that functioned as hedges in the quantitative RAs in the international English-medium publications suggests that L2 novice writers need to take into account the influences of research paradigm on the use of LBs when they write RAs for English-medium journals in the discipline of education. Our findings thus may help teachers of academic writing and L2 novice writers to develop their understanding of salient rhetorical practices across paradigms. In graduate-level classes, pedagogical instruction that focuses on the discourse functions of phraseological items in English academic writing across

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

paradigms can facilitate comprehension and the use of LBs in social science disciplines. Furthermore, consciousness-raising and practical corpus-driven activities can help L2 writers improve their use of phraseological patterns. For instance, L2 novice writers and advanced language learners may construct corpora within a specific discipline and explore the use of LBs through free software.

Several limitations to this study must be acknowledged. First, entire texts were included in our corpora. The discourse functions of LBs may vary across different sections of RAs, as identified in Shahriari's study (2017) in the discipline of applied linguistics. Moreover, the RAs were extracted from generalist journals, and there was no focus on a single sub-discipline. The generalisability of our findings is therefore limited because our corpora of RAs are small, and differences in lexico-grammatical characteristics can be found across the sub-disciplines of a single discipline.

In future research, discourse-based interviews with academic writers might offer insights into their epistemological dispositions towards research within their discipline. Further research might also compare discontinuous sequences (see Gray and Biber, 2013; and Römer, 2009) across paradigms within the discipline of education, and in other disciplines.

References

- Ädel, A. and B. Erman. 2012. 'Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach', *English for Specific Purposes* 31 (2), pp. 81–92.
- Becher, T. 1989. *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. Milton Keynes: Society of Research into Higher Education and Open University Press.
- Bernstein, B. 1999. 'Vertical and horizontal discourse: An essay', *British Journal of Sociology of Education* 20, pp. 157-173.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

- Biber, D. 1993. 'Representativeness in corpus design', *Literary and Linguistic Computing* 8, pp. 243–257.
- 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), pp. 275–311.
- Biber, D. and F. Barbieri. 2007. 'Lexical bundles in university spoken and written registers', *English for Specific Purposes* 26 (3), pp. 263–286.
- Biber, D., S. Conrad and V. Cortes. 2004. 'If you look at ...: Lexical bundles in university teaching and textbooks', *Applied Linguistics* 25 (3), pp. 371–405.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Cao, F. and G. Hu. 2014. 'Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences', *Journal of Pragmatics* 66, pp. 15–31.
- Charles, M. 2006. 'The construction of stance in reporting clauses: A cross-disciplinary study of theses', *Applied Linguistics* 27 (3), pp. 492–518.
- Chen, Y. and P. Baker. 2010. 'Lexical bundles in L1 and L2 academic writing', *Language Learning & Technology* 14 (2), pp. 30–49.
- Chen, Y. and P. Baker. 2016. 'Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1', *Applied Linguistics* 35 (1), pp. 1–33.
- Cohen, L., L. Manion and K. Morrison. 2013. *Research Methods in Education* (7th ed.). London: Routledge.
- Connor, U. 2004. 'Intercultural rhetoric research: Beyond texts', *Journal of English for Academic Purposes* 3 (4), pp. 291–304.
- Cortes, V. 2004. 'Lexical bundles in published and student disciplinary writing: Examples from history and biology', *English for Specific Purposes* 23 (4), pp. 397–423.
- Cortes, V. 2013. 'The purpose of this study is to: Connecting lexical bundles and moves in research article introductions', *Journal of English for Academic Purposes* 12 (1), pp. 33–43.
- Creswell, J. W. 2009. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Los Angeles, CA: SAGE.
- Durrant, P. 2017. 'Lexical bundles and disciplinary variation in university students' writing: Mapping the territories', *Applied Linguistics* 38 (2), pp. 165–193.
- Gray, B. 2013. 'More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles', *Corpora* 8 (2), pp. 153–181.
- Gray, B. 2015. *Linguistic Variation in Research Articles: When Discipline Tells Only Part of the Story* (Vol. 71). Amsterdam/Philadelphia: John Benjamins.
- Gray, B. and D. Biber. 2013. 'Lexical frames in academic prose and conversation', *International Journal of Corpus Linguistics* 18 (1), pp. 109–136.
- Hu, G. and F. Cao. 2015. 'Disciplinary and paradigmatic influences on interactional metadiscourse in research articles', *English for Specific Purposes* 39, pp. 12–25.

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

- Hu, G. and G. Wang. 2014. 'Disciplinary and ethnolinguistic influences on citation in research articles', *Journal of English for Academic Purposes* 14, pp. 14–28.
- Hyland, K. 2005. *Metadiscourse: Exploring Interaction in Writing*. Continuum: London
- Hyland, K. 2008. 'As can be seen: Lexical bundles and disciplinary variation', *English for Specific Purposes* 27 (1), pp. 4–21.
- Hyland, K. 2012. 'Bundles in academic discourse', *Annual Review of Applied Linguistics* 32, pp. 150–169.
- LaFlair, G. T., J. Egbert and L. Plonsky. 2015. 'A practical guide to bootstrapping in descriptive statistics, correlations, t tests and anovas' in L. Plonsky (ed.) *Advancing Quantitative Methods in Second Language Research*, pp. 46-77. Routledge: London.
- Larson-Hall, J. and R. Herrington. 2010. 'Improving data analysis in second language acquisition by utilizing modern developments in applied statistics', *Applied Linguistics* 31 (3), pp. 368–390.
- Maton, K. 2007. 'Knowledge-knower structures in intellectual and educational fields' in F. Christie and J.R. Martin (eds.) *Knowledge, Power and Educational Reform: Applying the Sociology of Basil Bernstein*, pp. 87-108. London: Continuum.
- McEnery, T., R. Xiao and Y. Tono. 2006. *Corpus-based Language Studies. An Advanced Resource Book*. London: Routledge.
- Nattinger, J. R. and J. S. DeCarrico. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Plonsky, L., J. Egbert and G. T. Laflair. 2015. 'Bootstrapping in applied linguistics: Assessing its potential using shared data', *Applied Linguistics* 36 (5), pp. 591–610.
- R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: <https://www.R-project.org>.
- Römer, U. 2009. 'English in academia: Does nativeness matter?', *Anglistik: International Journal of English Studies* 20 (2), pp. 89–100.
- Scott, M. 2012. *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software.
- Shahriari, H. 2017. 'Comparing lexical bundles across the introduction, method and results sections of the research article', *Corpora* 12 (1), pp. 1–22.
- Sinclair, J. McH. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Swales, J. M. 1990. *Genre Analysis: English in Academic and Research Settings*. New York: Cambridge University Press.
- Swales, J. M. 2004. *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press.
- Trowler, P. 2014. 'Depicting and research disciplines: Strong and moderate essentialist approaches', *Studies in Higher Education* 39 (10), pp. 1720-1731.
- Wignell, P. 2007. 'Vertical and horizontal discourse and the social sciences' in F. Christie and J. R. Martin (eds.) *Language, Knowledge and Pedagogy: Functional Linguistic and Sociological Perspectives*, pp. 184-204. London: Cassell.

This is the peer-reviewed version of the article which has been accepted for publication in 2019 by *Corpora*. Please visit <https://www.eupublishing.com/loi/cor> for the most up-to-date version.

Citation: Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education, *Corpora*, 14(2), 237–263. doi: 10.3366/cor.2019.0170

Wilcox, R. R. 2016. *Understanding and Applying Basic Statistical Methods Using R*. Hoboken, NJ: John Wiley & Sons.

Wilcox, R. R. and T. S. Tian. 2011. ‘Measuring effect size: a robust heteroscedastic approach for two or more groups’, *Journal of Applied Statistics* 38 (7), pp. 1359-1368.

Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.