

**THE ACQUISITION OF LEXICAL BUNDLES IN ENGLISH
FOR ACADEMIC PURPOSES: A MULTIDISCIPLINARY
STUDY OF NOVICE AUTHORS**

Doctoral Thesis

Diana Balasanyan

Supervisor: Dr. Sally Burgess



Departamento de Filología Inglesa y
Alemana

Universidad de La Laguna

June - 2017

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

ACKNOWLEDGEMENTS

*To the unfading memory of my grandfather,
Professor Levon Harutyunyan,
who would have been proud of me,
and who taught me that learning is
the most sublime form of happiness.*

It has been a privilege and tremendously good luck to work under the supervision of Sally Burgess, a person of outstanding wisdom and serenity combined with extraordinary efficiency. I would like to thank her for her generous, patient and tactful guidance, for her support and encouragement, and for all the faith she had in me.

I would like to express my gratitude to all the undergraduate and PhD students of the Universidad de La Laguna who participated in the workshop on lexical bundles for their keen participation and receptiveness, and for their feedback which was extremely helpful, generous, and constructive.

My doctoral journey at ULL began thanks to Dr. Garik Israelian, founder and President of Starmus Festival, who was so kind as to take care of all the admission paperwork for me before I even arrived on this beautiful island. Many thanks to him

1

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

and his wife Emma whose support and company during my first years in Tenerife were priceless.

I would like to thank all the members of my family for their support, care, and commitment: my grandmother Silva who took care of my toddler Maria while I was writing my thesis, for her unconditional dedication and never-ending patience; to my brother Eduard, his wife Christina and their amazing children, Anahit and Gagik, whose love and support are vital for me; to my father Gagik, for his quiet and considerate reassurance; to my aunt Anna, and her sons, David and Christ, who provided much-needed help and presented even more-needed cheerful moments; to my cousins Silvana and Ada, for their unconditional love and kindness; to my husband Quique, whose sense of humour and patience have saved me from many potential nervous breakdowns; to my 19-month-old daughter Maria, for being such a good girl and letting mummy write her thesis; and above all, to my mother Anahit, my cheerleader, my guardian angel, my all... I have been looking for the right words to thank her for 37 years now, but it looks like I am going to need some more time to find them. For the time being: Mummy, I love you!

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

TABLE OF CONTENTS

Abstract	7
Resumen	11
CHAPTER ONE. INTRODUCTION	15
1.1. English for academic purposes	16
1.2. The lexical bundle approach	20
1.3. Contextualising the present study: lexical bundles in EAP	24
1.4. Overview of the thesis	27
CHAPTER TWO. LITERATURE REVIEW	29
2.1. Phraseology: from periphery to the heart of linguistic research	30
2.2. Terminology and definitions of formulaic language	51
2.3. Lexical bundles: the most frequent sequences in a register	81
2.4. Lexical bundles in the EAP classroom	112
CHAPTER THREE. CORPUS AND METHODOLOGY	159
3.1. Contextualising the present study	160
3.2. Method	161
3.3. Corpus	169
3.4. Participants and workshops	175
CHAPTER FOUR. LEXICAL BUNDLES IN PSYCHOLOGY, LITERARY STUDIES, AND LINGUISTICS	
RESEARCH ARTICLES	179
4.1. Introductory remarks	180

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.2. Lexical bundles in psychology research articles	189
4.3. Lexical bundles in literary studies research articles	214
4.4. Lexical bundles in linguistics research articles	233
4.5. Lexical bundles across disciplines: frequency, structure, and use of lexical bundles in psychology, literary studies, and linguistics	247

CHAPTER FIVE. LEXICAL BUNDLES IN NOVICE ACADEMIC WRITING 262

5.1. Introductory remarks	263
5.2. Lexical bundles in l2 novice academic writing	271
5.3. Introducing lexical bundles to learners	279
5.4. Lexical bundles in the writing of the workshop participants	288
5.5. A case study of linguistic evolution	294

CHAPTER SIX. DISCUSSION AND CONCLUSIONS 298

6.1. Lexical bundles in psychology, literary studies, and linguistics	299
6.2. The frequency, structure, and functions of lexical bundles	300
6.3. Conclusions	313

REFERENCES 319

Appendix 1. The PhD and TFG'15-16 subcorpora	340
Appendix 2. Lexical bundles extracted from the PhD corpus	343
Appendix 3. Lexical bundles extracted from the expert corpus	345
Appendix 4. Materials of the introductory session and the workshop	348

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

LIST OF TABLES

Table 1.1.	Definitions and descriptions of formulaic sequences	55
Table 1.2.	Functions of formulaic sequences	61
Table 1.3.	Nattinger & DeCarrico's (1992) classification of lexical phrases according to four dimensions	68
Table 1.4.	Discourse functions of FEIs	77
Table 1.5.	Structural taxonomy of lexical bundles	88
Table 1.6.	Functional classification of lexical bundles	93
Table 1.7.	Functional taxonomies of lexical bundles	95
Table 1.8.	Classification of lexical bundles	96
Table 1.9.	The structure and function of lexical bundles	98
Table 3.1.	Frequency and range cut-off points	166
Table 3.2.	The PhD dissertations subcorpus	170
Table 3.3.	Research Articles (Expert) corpus	172
Table 4.1.	The most frequent lexical bundles in psychology RAs	190
Table 4.2.	The most frequent four-word bundles in psychology Ras	194
Table 4.3.	Five-word lexical bundles in psychology RAs	196
Table 4.4.	Structural classification of lexical bundles in psychology RAs	198
Table 4.5.	Functional classification of lexical bundles in psychology RAs	203
Table 4.6.	Lexical bundles unique to psychology RAs	211
Table 4.7.	The most frequent lexical bundles in literary studies RAs	216
Table 4.8.	Five-word lexical bundles in literary studies RAs	217
Table 4.9.	Structural classification of lexical bundles in literary studies RAs	219
Table 4.10.	Functional classification of lexical bundles in literary studies RAs	224
Table 4.11.	Lexical bundles found exclusively in literary studies RAs	230
Table 4.12.	The most frequent lexical bundles in linguistics RAs	233
Table 4.13.	The most frequent four-word bundles in linguistics RAs	234
Table 4.14.	The five-word bundles in linguistics RAs	235
Table 4.15.	Structural classification of lexical bundles in linguistics RAs	237
Table 4.16.	Functional classification of lexical bundles in linguistics RAs	241
Table 4.17.	Lexical bundles found only in linguistics RAs	244

Table 4.18.	The number of lexical bundle types in the three disciplines	249
Table 4.19.	Type / token ratios for the lexical bundles in the three disciplines	251
Table 4.20.	The most frequent lexical bundles in the three disciplines	253
Table 4.21.	List of the bundles shared by the three disciplines	257
Table 5.1.	The most frequent lexical bundles in the L1 novice corpus	272
Table 5.2.	Bundles underused in the L2 corpus	274
Table 5.3.	Bundles used more frequently in L2 corpus	275
Table 5.4.	Bundles used with similar frequencies in L1 and L2 subcorpora	276
Table 5.5.	Bundles used in Activity 1	280
Table 5.6.	Change of the underused bundles in workshop participants	289
Table 5.7.	Change of the overused bundles in workshop participants	290

LIST OF FIGURES

Figure 1.1.	Parameters of classification of formulaic sequences along continua	64
Figure 1.2.	Classification of word-combinations according to their degree of fixedness	69
Figure 1.3.	The functional taxonomy of lexical bundles	91
Figure 1.4.	The relationship between expertise, nativeness and proficiency	119
Figure 4.1.	Concordance lines for <i>as a result of</i> in linguistics	186
Figure 4.1.	The distribution of structural types of lexical bundles in psychology	200
Figure 4.3.	The number of types of lexical bundles for each function in psychology	205
Figure 4.4.	Concordance lines for <i>support for the</i>	207
Figure 4.5.	The distribution of structural types of lexical bundles in literary studies	221
Figure 4.6.	The number of types of lexical bundles for each function in literary studies	225
Figure 4.7.	The number of types in each structural category of lexical bundles in linguistics	230
Figure 4.8.	The number of types representing each function of lexical bundles in linguistics	243
Figure 4.9.	Lexical bundle types and tokens in the three disciplines	250

ABSTRACT

Lexical bundles, also termed *clusters* and *N-grams*, have been described as sequences of three or more words which occur more frequently than expected by chance in a register (Biber et al., 1999). Despite their structural incompleteness and semantic transparency, these extremely frequent formulaic sequences have been shown to perform important discourse building functions. It has been argued that formulaic sequences are stored and reproduced holistically as unanalysed chunks by proficient speakers (Wray, 2002; Schmitt & Underwood, 2004) and play an important role in constructing fluent and coherent discourse (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983).

Although lexical bundles are more frequent in conversation than in other registers, they have been shown to perform crucial communicative functions in academic discourse (Biber, 2006). Moreover, each genre employs a characteristic set of lexical bundles to construct meaning, and these characteristic clusters are a distinctive feature of each genre (Hyland, 2012).

A wide range of studies have addressed the use of lexical bundles in academic speech and writing (Biber & Barbieri, 2007; Biber & Gray, 2010; Nesi & Basturkmen, 2006; Sánchez, 2013), the writing of L1 and L2 authors (Ädel & Erman, 2012; Chen &

Baker, 2010; Salazar, 2014), and in expert and novice academic writing (Biber et al. 2004; Chen & Baker, 2010; Cortés, 2004, Hyland, 2008b; Staples et al., 2013) showing the existence of substantial differences in the use of lexical bundles in L1 and L2 writing (Ädel & Erman, 2012; Byrd & Coxhead, 2010; Pan, Reppen & Biber, 2016). Other studies have contrasted the use of lexical bundles across different disciplines (Allen, 2001; Biber, 2005, Hyland, 2008a) or addressed one or two particular disciplines (Ädel & Erman, 2012; Chen & Baker, 2010; Cortés, 2004, Eriksson, 2012; Farvardin, Afghari & Koosha, 2012; Sánchez, 2014; Salazar, 2011).

It has been argued that each field of knowledge possesses a different set of lexical bundles which help to construct knowledge and are an important part of discourse practices and generic conventions of the field (Hyland, 2008a; Durrant, 2009; Eriksson, 2012).

Novice L2 researchers have been shown to overuse certain lexical bundles and underuse others, thus producing texts that are not prototypical of their target genres and disciplines (Ädel & Erman, 2012; Chen & Baker, 2010; Salazar, 2014). Several pedagogically-oriented studies have aimed to generate lists of lexical bundles for novice researchers (Ackermann, 2013; Simpson-Vlach & Ellis, 2010) and to teach the use of lexical bundles in EAP / EFL classes (Cortés, 2006; Erman et al., 2013; Jones & Haywood, 2005; Peters & Pauwels, 2015).

The methodological approach adopted in most studies aimed at teaching lexical bundles is based on vocabulary-teaching techniques (Alali & Schmitt, 2012; Cortés,

2006; Erman et al., 2013; Jones & Haywood) which do not seem to be the most appropriate for teaching lexico-grammatical patterns. As a result, the learners acquire the theoretical knowledge regarding lexical bundles and mark high scores on tests, but their post-training writing does not seem to improve considerably.

The present study examines lexical bundles from a pedagogical / EAP perspective as a resource for improving novice L2 researchers' writing. The study focusses on three disciplines, namely, psychology, literary studies, and linguistics, in order to determine their frequency, structure, and use in each discipline. The second part of the study attempts to devise a methodology for teaching lexical bundles to advanced L2 users of English and provide them with tools and instruction for further self-directed learning.

The present study involves two stages: the analysis of lexical bundles in the three disciplines and instructing novice L2 researchers to compose their own personalised mini-corpora and extract a set of lexical bundles characteristic of the prototypical texts of their discipline, subject, and topic.

During the first stage, lexical bundles were extracted from a corpus of 2,1 million words using AntConc (Anthony, 2010) and classified according to their structures and broad functional categories suggested by Biber, Conrad and Cortés, 2003, and Hyland, 2008b.

The second stage of the study consisted in a short introductory session and a workshop directed to upper-undergraduate students of the Department of English and

German Philology of the University of La Laguna and to Doctoral students of the Faculty of Psychology of the same university.

The findings of the present study suggest that each discipline possesses its own set of lexical bundles which are part of established disciplinary practice and generic conventions, and novice researchers should acquire these practices in order to get accepted in their target discourse communities. The study demonstrates the importance of adopting a discipline-specific approach in addressing the needs of the learners. Although the post-workshop texts submitted by the participants are not sufficient to make strong claims in favour of this approach, the results of the workshop on lexical bundles and the student feedback collected from the participants show that this kind of instruction raises the awareness of the learners and is well-received by them.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

RESUMEN

Los paquetes léxicos (lexical bundles), también conocidos como “racimos” (clusters), and N-grams, se definen como secuencias de más de dos palabras que ocurren en un registro determinado con una frecuencia mayor de la esperada por azar (Biber et al., 1999). A pesar de que son estructuralmente incompletas y semánticamente transparentes, estas secuencias frásticas altamente recurrentes desempeñan importantes funciones en la construcción del discurso (Biber et al., 1999; Biber, 2006). Se ha demostrado que los hablantes competentes almacenan las frases formulaicas y las reproducen enteramente, sin analizar sus componentes (Wray, 2002; Schmitt & Underwood, 2004), y que estas unidades léxico-gramáticas desempeñan un papel importante en la construcción de un discurso coherente y fluido (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983).

A pesar de que los paquetes léxicos son más frecuentes en conversación que en otros registros, se ha demostrado que estas secuencias frásticas desempeñan importantes funciones comunicativas también en el discurso académico (Biber, 2006). Cada género posee un conjunto de paquetes léxicos característicos que contribuyen a la creación de significado y son un rasgo distintivo del género (Hyland, 2012).

Se han realizado numerosos estudios de los paquetes léxicos en el discurso académico escrito y hablado (Biber & Barbieri, 2007; Biber & Gray, 2010; Nesi & Basturkmen, 2006; Sánchez, 2013), contrastando los textos de autores nativos y no

nativos (Ädel & Erman, 2012; Chen & Baker, 2010; Salazar, 2014), así como los de académicos expertos y novatos (Biber et al. 2004; Chen & Baker, 2010; Cortés, 2004, Hyland, 2008b; Staples et al., 2013). Según estos estudios, existen diferencias significativas entre el uso de los paquetes léxicos de los autores nativos y no nativos (Ädel & Erman, 2012; Byrd & Coxhead, 2010; Pan, Reppen & Biber, 2016). Además, se han realizado estudios contrastivos de los paquetes léxicos en diferentes disciplinas (Allen, 2001; Biber, 2005, Hyland, 2008a), y estudios enfocados en disciplinas o áreas académicas concretas (Ädel & Erman, 2012; Chen & Baker, 2010; Cortés, 2004, Eriksson, 2012; Farvardin, Afghari & Koosha, 2012; Sánchez, 2014; Salazar, 2011).

Según estas investigaciones, cada área de conocimiento cuenta con su propio conjunto de paquetes léxicos para construir conocimiento. Estas secuencias frásticas recurrentes constituyen una parte crucial de las prácticas discursivas y convenciones genéricas de cada campo (Hyland, 2008a; Durrant, 2009; Eriksson, 2012).

Se ha demostrado que los investigadores novatos tienden a abusar de algunos paquetes léxicos y hacer menor o ningún uso de otros, produciendo textos que no son prototípicos para su campo y género (Ädel & Erman, 2012; Chen & Baker, 2010; Salazar, 2014). Varios estudios con enfoque pedagógico se han dedicado a generar listas de paquetes léxicos para investigadores novatos (Ackermann, 2013; Simpson-Vlach & Ellis, 2010), o han incorporado la enseñanza de los paquetes léxicos en

clases de Inglés para Fines Específicos o Inglés como Lengua Extranjera (Cortés, 2006; Erman et al., 2013; Jones & Haywood, 2005; Peters & Pauwels, 2015).

En la mayoría de estos estudios enfocados a la enseñanza de los paquetes léxicos, la metodología adoptada se basa en las técnicas de enseñanza de vocabulario (Alali & Schmitt, 2012; Cortés, 2006; Erman et al., 2013; Jones & Haywood), las cuales no parecen del todo adecuadas para la enseñanza de patrones léxico-gramáticos. El resultado de estas metodologías es que los discentes adquieren conocimiento teórico sobre paquetes léxicos y obtienen buena puntuación en las pruebas finales, pero su producción escrita no parece haber mejorado sustancialmente después de la formación.

El presente trabajo estudia los paquetes léxicos desde una perspectiva pedagógica, como un recurso para mejorar los textos escritos por académicos novatos no nativos. El estudio recoge tres disciplinas, a saber, psicología, lingüística y estudios literarios, con el fin de analizar la frecuencia, la estructura y el uso de los paquetes léxicos en cada una de ellas. La segunda parte del estudio consiste en proponer un método de enseñanza de los paquetes léxicos a usuarios avanzados de inglés y ofrecerles herramientas y metodología para aprendizaje independiente.

De ahí que el estudio conste de dos fases: el análisis de los paquetes léxicos en las tres disciplinas arriba mencionadas, y la formación de académicos novatos no nativos con el fin de ofrecerles una metodología para construir sus propios mini-corpus y

extraer los paquetes léxicos a partir de textos prototípicos característicos de su campo, disciplina y objeto de estudio.

Durante la primera fase, se extrajeron los paquetes léxicos de un corpus de aproximadamente 2,1 millones de palabras usando AntConc (Anthony, 2010). A continuación, los paquetes léxicos se clasificaron según la taxonomía funcional propuesta por Biber, Conrad y Cortés (2003) y Hyland, (2008b).

La segunda fase del estudio consiste en una breve sesión introductoria y un taller dirigido a los discentes de la Universidad de La Laguna, alumnos del Departamento de Filología Inglesa y Alemana del último año de la carrera y estudiantes de Doctorado de la Facultad de Psicología.

En el presente estudio se demuestra que cada disciplina posee un conjunto de paquetes léxicos que son parte de las prácticas discursivas establecidas, consecuentemente, los académicos novatos deberían adquirir estas prácticas para ser admitidos en la comunidad del discurso de dicha disciplina. Asimismo, se demuestra que la enseñanza de paquetes léxicos debe basarse en las disciplinas concretas de los discentes para corresponder a sus necesidades. Aunque los textos producidos por los participantes del taller después de la formación no tienen suficiente volumen para afirmar con seguridad la viabilidad de esta metodología, los resultados y la retroalimentación obtenidos de los participantes del taller indican que este tipo de formación aumenta el conocimiento de los discentes sobre los paquetes léxicos y es bien recibida por ellos.

CHAPTER ONE

INTRODUCTION

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

1.1. ENGLISH FOR ACADEMIC PURPOSES

As a result of the dominant position of English as the language of scientific communication the language has on occasion been described as a “Tyrannosaurus Rex” (Swales, 1997). More charitable term that has been used is “Lingua Franca” (Kuteeva & Mauranen, 2014). Regardless of how one might regard the current position of English in the academy, there has been an increasing demand for high-quality training materials for researchers of different cultural backgrounds. Second language (L2) researchers in all fields need efficient strategies for communicating their research and obtaining the status of membership and later expertness in their target discourse communities.

Research shows that L2 researchers get high rejection rates in journal publication submissions partly due to the fact that they fail to comply with the linguistic norms and rhetorical conventions of the international discourse community (Flowerdew, 2008; Lillis & Curry, 2015). There are also cases where scholars turn to ‘literacy brokers’ (Lillis & Curry, 2006) only to find that their submission to journals still fail to gain acceptance.

Both L1 and L2 researchers experience the constant ‘publish or perish’ pressure, although for the latter group there is one more pressing dimension, namely gaining a

knowledge of English, since in the globalised academic world publishing research almost inevitably implies publishing in English (Flowerdew, 2008). The general concept of “knowledge of English” in EAP (English for Academic Purposes) is described more precisely as “academic literacy” (Hyland & Hamp-Lyons, 2002, p. 4), a concept encompassing the genre, community, cultural and disciplinary knowledge related to a discourse community. The desire to join their target discourse communities leads L2 researchers to adopt a range of strategies aimed at understanding and observing the rhetorical conventions of their target discourse community.

Over the past fifty or so years, the field of ESP / EAP research has responded to this need of L2 researchers in order to provide efficient methodologies for publishing and presenting research internationally, and to equip them with strategies for integration into the international academic discourse community, addressing the issue from a variety of perspectives, such as English as a Lingua Franca (Canagarajah, 2007; Kuteeva & Mauranen, 2014; Mauranen & Ranta, 2009), Intercultural Rhetoric (Connor, 2004; Connor & Rozycki, 2013;), Genre Analysis (Bhatia, 1993, 2008; Dudley-Evans, 2000; Swales, 1990; 2004), and so on.

Two fundamental issues should be mentioned in this respect. EAL users tend to perceive English as homogeneous and as operating a single set of ‘rules’ which are applicable in all contexts. In addition, even proficient L2 users of English tend to regard the language as somehow ‘belonging to native speakers’ (Thompson, 2013)

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

and adopt or feel they have been assigned a position of inferiority with respect to the privileged L1 user ‘owners’. (Lillis & Curry, 2015). This situation is all the more surprising given the fact that L2 or EAL users of English by far outnumber native speakers (Graddol, 2006).

While it is true that L2 / EAL texts show great variation across lexico-grammatical, socio-pragmatic and discursal dimensions, EAP research increasingly argues that this variation should not necessarily be stigmatised as deviation (Cargill & Burgess, 2017; Flowerdew, 2008; Kuteeva & Mauranen, 2014). This notwithstanding, both EAP scholars and L2 researchers acknowledge their need for teaching and training materials designed specifically for novice or international researchers seeking to publish their work in English (Cargill & O’Connor, 2009).

The rapidly growing field of EAP has responded to the needs of L2 learners with an extensive body of high-quality research addressing a variety of linguistic features in a full range of text types. This research might focus on broadly defined registers as in the case of Biber’s (2006) study of university language embracing many if not all the possible communicative events one might find in a university setting. At the other end of the spectrum one also finds minute analyses of particular segments of texts in a given genre, such as the rhetorical structure of research article abstracts (Martín Martín, 2003). Advances in the field of English for Specific Purposes have highlighted the need to address the many varieties of English, a full range of genres, many varied features, and learner populations both comparatively and separately.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The present study is an attempt to contribute to the knowledge of novice L2 writer lexico-grammatical usage patterns and to suggest a methodology for developing writing strategies using a corpus-based approach.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

1.2. THE LEXICAL BUNDLE APPROACH

The increasing demand for authentic data, the advances in EAP and corpus linguistics, and the growing number of learners of English as a Second / Foreign Language have led to a paradigm shift in linguistic research in recent years (Hyland & Hamp-Lyons, 2002). It has been increasingly argued that the content taught to learners at all levels should reflect authentic language use rather than relying on artificially constructed materials (Biber et al., 1999; Sinclair, 1991). The personal experience of EFL (English as a Foreign Language) learners of a variety of backgrounds – my own included - reflects the common situation of ‘scoring high on language tests but being unable to produce a coherent sentence’, ‘being a successful student, but showing poor written and spoken communication skills’ (personal communications from different groups of learners), and the notorious difficulty of “native-like selection and native-like fluency” (Pawley & Syder, 1983). Without this ability, grammatically correct but uncommon collocations characterise much writing that is produced and identify writers as non-native or novice (Ellis, 2002).

Many of the difficulties faced by learners do not lie in the realm of lexis or grammar, but are instead related to the native-like selection of lexico-grammatical patterns (Nattinger & DeCarrico, 1992). These are commonly described as “recurrent word-

20

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

combinations” (Altenberg, 1998), “lexical phrases” (Nattinger & DeCarrico, 1992) or “formulaic sequences” which are stored, retrieved and processed holistically by L1 speakers (Wray, 2002). The extent to which such formulaic sequences are treated holistically by L2 speakers is not clear though there seems to be some evidence that certain strings of lexical items are processed more quickly than others by L1 and L2 users alike (Underwood, Schmitt & Galpin, 2004).

Native-like fluency has been defined as the ability to combine or collocate lexical items in commonly accepted ways (Pawley & Syder, 1983; Römer, 2009), and as the knowledge of the company that words keep (Firth, 1962), or the knowledge of their primings (Hoey, 2005).

The analysis of massive corpora has provided unexpected insights into the nature of the lexico-grammatical relationships of linguistic elements and has resulted in the emergence of new categories of formulaic language. One such category are the lexical bundles, first described by Biber, Johansson, Leech, Conrad, and Finegan (1999) as structurally incomplete, semantically transparent multi-word sequences which occur more frequently than expected by chance in a given genre.

It should be noted that the condition of genre-sensitivity is crucial for defining lexical bundles, since in a sufficiently large corpus any string of words may occur more frequently than expected by chance (Sinclair, 1991). The important point is that certain strings occur in certain genres very frequently, thus constituting a reliable indicator for genre definition (Hyland, 2012). Studies have shown that certain sets of

lexical bundles are characteristic not only of different genres, but also of separate disciplines (Eriksson, 2010; Farvardin, Afghari & Koosha, 2012; Oakey, 2002).

These findings have triggered a whole body of research concerned with teaching the ‘correct’ (i.e., standardised and common) use of lexical bundles in EAP (Ackermann & Chen, 2013; Biber & Barbieri, 2007; Biber, Conrad & Cortés, 2003, 2004; Hyland, 2008b; Simpson-Vlach & Ellis, 2002) and, to a lesser extent, in separate disciplines (Ädel & Erman, 2012; Chen & Baker, 2010; Cortés, 2004, Eriksson, 2012; Farvardin, Afghari & Koosha, 2012; Sánchez, 2014; Salazar, 2011, 2014). Contrastive analyses of the use of lexical bundles in learner and expert, L1 and L2 writing have shown great variation across these groups (Ädel & Erman, 2012; Allen, 2011; Chen & Baker, 2010; Staples et al., 2013, etc.).

Despite the extensive body of research in the field of phraseology in general and lexical bundles in particular, few studies have addressed their use in the social sciences and the humanities. Although linguistics and psychology have been included in many cross-disciplinary and multi-disciplinary studies of formulaic sequences (Ädel & Erman, 2012; Allen, 2011; Biber, 2006; Durrant, 2009; Hyland, 2008a; Oakey, 2002; Simpson-Vlach & Ellis, 2010), no study has addressed these disciplines separately in the same way as science disciplines such as physics (Farvardin, Afghari & Koosha, 2012) or health sciences (Salazar, 2014), which have been given exclusive attention. Literary studies is a disciplinary area that has not been

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

addressed in any study that I am aware of, aside from Johnston (2017) Master of Arts dissertation which examines lexical bundles in linguistics and literary studies.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

1.3. CONTEXTUALISING THE PRESENT STUDY: LEXICAL BUNDLES IN EAP

The purpose of this study is to raise the awareness of University of La Laguna upper-undergraduate and post-graduate students of the humanities and social sciences with regard to the use of lexical bundles in the disciplines of applied linguistics, literary studies, and psychology. These disciplines were selected because of the relative scarcity of studies of lexical bundles in these fields as was mentioned above

It has been shown that the arts and humanities disciplines employ a very different set of formulaic sequences (Durrant, 2009) which make these fields stand apart from other disciplines, making them all the more worthy of study. The suggested tentative explanation for this phenomenon is that the humanities researchers are expected to create novel utterances instead of relying on formulae. One of the goals of this study is to determine the extent to which this explanation holds true for the above-mentioned disciplines, particularly for literary studies. Since linguists too are expected to demonstrate their command of the language through original and creative expression of their ideas, then those writing in literary studies might be required to employ even more creative language, hence fewer lexical bundles would be found in texts representing this discipline. Psychology, on the other hand, lying closer to the natural sciences can be regarded as a discipline in which one might expect to find

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

more lexical bundles used much as they are in the sciences to indicate and establish group identity (Wray, 2002), or discourse community membership, and to reflect a depersonalised, objective stance where the personality of the researcher is of little or no importance (Hyland, 2008a).

As far as the pedagogical / EAP dimension is concerned, my assumptions are based on previous studies aimed at teaching experimental courses to native and non-native learners (Alali & Schmitt, 2012; Cortes, 2006; Jones & Haywood, 2004), which while succeeding in raising awareness and teaching the theory, have all been based on standardised lists of lexical bundles. The purpose of this study, on the other hand, was to research lexical bundle use across the target disciplines with a view to providing the learners with tools and methodologies which would enable them to work through the acquisition of lexical bundles, in line with learner-centred (Matukhin & Bolgova, 2015; Thompson, 2013) teaching principles and the idea of learner autonomy (Boyadzhieva, 2016). It was expected that the students' written production would improve significantly as a result of this approach.

To this end, I posit a corpus-based methodology designed for advanced L2 speakers of English which consists in collecting personalised small-scale corpora containing texts which learners consider prototypical of their discipline and analysing them using corpus-analysis software, AntConc (Anthony, 2010).

The purpose of the present study is thus to examine the lexical bundles in research articles drawn from three disciplines: psychology, applied linguistics, and literary

studies, and to propose a corpus-driven methodology for teaching lexical bundles to advanced learners of English in order to improve their academic writing.

The study consists of two parts: 1. identifying the lexical bundles characteristic of each of the three disciplines, and 2. conducting a workshop aimed at raising the awareness of students regarding lexical bundles and instructing them to work with their own corpora to extract the lexical bundles relevant to their work.

The research questions for the first part of the study are the following:

1. Which are the most frequent formulaic sequences in psychology, literary studies, and linguistics research articles?
2. What structure do they have and what functions do they perform?
3. Which bundles are exclusive to each discipline?
4. Which bundles are common to all?

The research questions for the second part of the study are the following:

5. To what extent can a corpus-based methodology help the learners to improve their writing?
6. How will the use of lexical bundles vary across the groups of learners who had received this kind of instruction with respect to a comparable group who have not received such instruction?
7. Can variation in the use of lexical bundles be attributed exclusively to the instruction received?

1.4. OVERVIEW OF THE THESIS

This **first Chapter of the thesis** describes the lexical bundle approach in the field of EAP providing a rationale for a study of lexical bundles in different disciplines and indicating the need for research in this field. It also outlines the importance of lexical bundles in academic writing and their role as indicators of group identity and important building blocks of discourse.

Chapter two summarises the research in the field of phraseology, tracing its development from its origins to the present day (Section 2.1), focussing particularly on such aspects as the confusing terminology used in this field (Section 2.2), explaining the lexical bundle approach (Section 2.3), and discussing the pedagogically-oriented research on lexical bundles (Section 2.4).

Chapter three describes the methodology and the corpus used in this study (Sections 3.1 and 3.2), as well as giving an account of the introductory seminar on lexical bundles, the workshop, and its participants (Section 3.3).

Chapter four begins with a brief introduction to the analysis of lexical bundles (Section 4.1) and discusses their frequency, structure, usage, and the number of unique types in the disciplines of psychology (Section 4.2), literary studies (Section

4.3), and linguistics (Section 4.4). The chapter concludes with an analysis of the bundles which are common to the three disciplines (Section 4.5).

Chapter five lays out an introduction on the use of lexical bundles in novice academic writing (Section 5.1) and the results of the introductory session (Section 5.2) and the workshop conducted for the learners (Section 5.3). It concludes with a case study of one participant's development as a writer of academic text (Section 5.4).

Chapter six summarises the findings of this study and offers conclusions, along with a reflection on the implications and limitations of the study and offers suggestions for further research.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

CHAPTER TWO

LITERATURE REVIEW

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

2.1. PHRASEOLOGY: FROM PERIPHERY TO THE HEART OF LINGUISTIC RESEARCH

Despite the immense variety of linguistic forms and resources, and apparently infinite number of choices available to a speaker or writer when producing spoken utterances or written texts, there exists a category of lexico-syntactic units which appear in speech or writing as unaltered formulae, reproduced every time exactly in the same way by different speakers / writers in different texts, registers and genres. Hunston (2002) describes them as “phrases, which are not fixed lexically but are not random either” (175). This category is far from being peripheral or marginal, as was thought previously: in fact, studies show that it pervades all registers and genres (Biber & Barbieri, 2007; Biber & Gray, 2010; Granger & Meunier, 2008b; Howarth, 1996; Stubbs, 2001; Wray, 2002), and may constitute up to 80% of all the words in certain registers (Altenberg, 1998).

This section will discuss the previous research focussing on key characteristic features of formulaic sequences which confer on them a crucial role in language studies. These characteristics include the pervasive nature of prefabricated language in all genres and registers, their role in achieving idiomaticity and native-likeness in language production, the advantages of using them instead of relying on single lexical

units, among them processing advantage. I will also discuss the factors that brought phraseological research to the centre of linguistic studies and the publications which fostered this promotion.

Fixed or semi-fixed patterns and expressions have received different names in the literature, including “lexical phrases” (Nattinger & DeCarrico, 1992), “sentence stems” (Pawley & Syder, 1983), “phraseologisms” (Gries, 2008), “recurrent word combinations” (Ädel & Erman, 2012; Altenberg, 1998; Conrad & Biber, 2004; Granger & Lefer, 2016), “phrasal lexemes” (Moon, 1998) “phraseological units” (Cowie, 1998), “prefabricated expressions / language / patterns”, or “prefabs” (AlHassan & Wood, 2015; Coxhead & Byrd, 2010; Granger, 1998), “set expressions” (Gläser, 1998; Howarth, 1998), “idioms” (Wulff, 2010), “restricted collocations” (Howarth, 1996, 1998; Kuiper, 2004; Siyanova & Schmitt, 2004), etc. The diversity of terms for formulaic sequences will be discussed in Section 1.2, but the umbrella terms “formulaic language”, “formulaic sequences”, and “formulae” will be used throughout this study to refer to any multi-word sequence which:

- a). is institutionalised and stored holistically in the memory of (native) speakers, whether or not it presents any kind of semantic or grammatical irregularity, and / or
- b). occurs more frequently than expected by chance.

Other terms will be used to discuss previous research and will refer to the sense in which a particular author or authors use it.

Formulaic sequences have been the focus of attention of a variety of linguistic disciplines for several decades, covering a wide range of perspectives, theories and frameworks, notably at the intersections of two or more fields of linguistics, among them corpus linguistics (Ädel & Erman, 2010; Biber et al, 1999; Biber, 2009; Cortés, 2004, 2006, 2008; Chen & Baker, 2010; Erman & Warren, 2000), psycholinguistics (Ellis, 2002; Ellis et al., 2008), SLA, or second language acquisition (Simspon-Vlach & Ellis, 2010; Wray, 2002), discourse analysis (Ädel, 2014; Cortés, 2013, Kuteeva & Negretti, 2016; Le & Harrington, 2015), construction grammar (Wulff, 2010), as well as the field of phraseology itself (Cowie, 1998; Granger & Meunier, 1998; Moon, 2008), EAP, or English for Academic Purposes (Cortés, 2002, 20004, 2006, 2008, 2013; Biber, 2006; Biber et al, 1999; Biber, Conrad & Cortés, 2004; Howarth, 1996, 1998; Hyland, 2008a, 2008b, 2012; Nesi & Basturkmen, 2006). Gries (2008) discusses the concepts of phraseology and phraseologism in the frameworks of Transformational Generative Grammar, Cognitive Linguistics / Construction Grammar and Corpus Linguistics, showing that the latter approach is the one which reflects the nature of formulaic sequences most adequately.

Corpus Linguistics has played a major role in phraseological research over the past decades: at the end of the 90s, corpus-driven data created a number of opportunities and raised several fundamental research questions, one of which was designing and providing adequate and up-to-date learning materials based on corpus analysis. As a result, phraseology has recently gained increasing centrality in linguistic research (Granger & Paquot, 2008). Modern phraseological research is based largely on

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Firthian concepts and notably his famous phrase: “You shall know a word by the company it keeps!” (Firth, 1962: 11), but before this branch of linguistics began to experience a rise in the West, considerable advances in phraseology were made in the Soviet Union, although from a rather different perspective. This section will briefly outline the development of the studies of collocation and their transition into modern phraseological research.

The term ‘phraseology’ comes from the Russian tradition where the textual relationships between single lexical items were first described and studied back in the Soviet period by Vinogradov (1947), Amosova (1963), and others. The Eastern European and particularly Russian tradition was the first to focus on these lexico-syntactic items well before the 1980s, when phraseological units (another term that comes from the Russian tradition) began to arouse the interest and attention of Western linguists.

It should be noted that Soviet and post-Soviet phraseological research is largely theoretical, focussing on the *langue* rather than *parole* in Saussurean terms, and leading to complex and abstract theoretical constructions while seldom addressing the extent to which these theories might be applied by language teachers, learners, or users in general. Even in the case of pedagogic genres, it is the characteristic feature of Soviet linguistics to dwell on prescriptive rather than descriptive approaches. Nevertheless, the contribution of the Soviet and post-Soviet Russian linguists to setting the foundations of the study of fixed expressions should be acknowledged:

“Classical’ Russian theory, with its later extensions and modifications, is probably the most pervasive influence at work in current phraseological studies and is unrivalled in its application to the design and compilation of dictionaries.”
(Cowie, 1998: 2).

The study of formulaic language goes back over a century (Altenberg, 1998), but the real advances happened with the advent of computers and information technology (IT), when it became possible to objectively determine the most frequent lexical sequences in a language or register and use advanced technologies and specialised software to analyse them.

One feature that was of great interest to researchers at the beginning was the idiomaticity of formulaic sequences. Expressions were classified across a continuum from free to bound collocations, in which the free collocations were those whose meaning was transparent and whose components could be replaced more or less freely by other words, thus corresponding roughly to the Firthian concept of “the free narrative of customary observance” (1962: 11), or the open choice principle suggested by Sinclair (1991), while the bound phraseological units were those consisting of fixed components with opaque meaning: the representation of the idiom principle (Sinclair, 1991). It was noted that the most opaque collocations (i.e., idioms) were the least frequent, although at the beginning of phraseological research they received the most attention, while free combinations which were much more frequent in all types of texts were not a popular object of research (Cowie, 1998; Granger & Paquot; 2008 Moon, 1998).

It is not surprising that of all types of phraseological units, idioms attracted most attention (Sabban, 2008): they are both the most conspicuous and the most irregular part of the lexis. However, the advances in corpus linguistics made it evident that idioms were among the least frequent and the least practical items for language users and learners. Since the 1980s, the focus of phraseology has shifted towards more frequent and less figurative lexical sequences. It was then that phraseology ceased to be a marginalised area that occupied the attention of a limited number of trained linguists and became a major trend in language studies (Cowie, 1998; Granger & Paquot, 2008). Since the 1990s, the study of phraseological units (under many different names and from many different perspectives) has attracted the attention of a large number of researchers.

It is generally agreed that formulaic sequences constitute a substantial proportion of discourse in any register and genre, be it spoken or written (Biber et al., 1999; Schmitt & Carter, 2004; Stubbs, 2001). Moreover, the most frequent words in a language show “strong phraseological tendencies” (Stubbs, 2001: 60). Given the all pervasiveness of formulaic language across all registers and genres the fact that it has remained a comparatively neglected area of study is something of a mystery. It has in fact been demonstrated that at least 30% of academic discourse is composed of different kinds of prefabricated units (Cowie, 1998; Howarth, 1998). Prefabricated language plays an even more prominent role in everyday conversation: according to Altenberg (1998), constituting as much as 80% of all the words. Erman and Warren (2000) reported that different types of word combinations made up 58.6% of the

spoken corpus analysed in their study, and 52.3% of the written corpus. These discrepancies may be surprising, but it must be borne in mind that there are many different methods and approaches to studying phraseology which suggest different definitions and different ways of identifying formulae. It is for this reason that Schmitt and Carter (2004) call for methodological triangulation (the use of more than one method) in the identification of formulaic sequences where possible in order to obtain more objective results.

Many authors point out that there are no clear-cut boundaries between formulae and non-formulae, referring to a cline or continuum rather than clearly defined groups (Altenberg, 1998; Cowie, 1998; Erman & Warren, 2000; Nattinger & DeCarrico, 1992; Wray, 2002). The theory of lexical priming (Hoey, 2005) convincingly demonstrates that no collocation is absolutely free, since each word, even minor words belonging to closed classes, are primed for certain uses and at the same time avoid other, less typical combinations. The study carried out by Cosme and Guilquin (2008) confirms that even function words have their ‘company preferences’ and are not as straightforward and simple as it might seem.

Before the advances in corpus linguistics, language teachers and materials designers relied mainly on their intuition for the selection of vocabulary and other content to be taught, thus putting strong emphasis on the most salient but not necessarily the most frequent and relevant items. However, the availability of large corpora and the possibility to search them instantly for any kind of recurrent features made it obvious

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

that other units, more frequent and intuitively non-salient, present major interest due to their crucial role in any register, and a better understanding of the nature of these expressions will be useful both for language learners and native speakers. While intuition should not be discarded altogether (Fiedler, 2017; Howarth, 1996; Meunier & Granger, 2008), as it may be useful at a certain stage of analysis, especially the intuition of a trained linguist (Simpson-Vlach & Ellis, 2010), other, less subjective methods are required to provide adequate learning materials for learners (Biber, Conrad & Cortés, 2004; Biber & Barbieri, 2007; Cowie, 1998; Ellis, 2002; Moon, 1998; Salazar, 2014; Tognini-Bonelli, 1996) using “the evidence of the long text” (Sinclair, 2004: 17). In a thorough study of first and second language acquisition mechanisms and aphasic speakers, Wray (2002) draws a Heteromorphic Distributed Lexicon model which integrates much of the knowledge of formulaic language and proposes an insight into how and why formulaic language is stored and retrieved in the memory of the user, and how it could be taught to different types of learners. Wray also challenges the role of intuition in determining and dealing with formulaicity, observing at the same time that it may be viable for other purposes related to language teaching and acquisition. Thus, intuition appears to be a useful tool, though its role is diminished by other, more powerful and objective tools, such as corpus-driven data and corpus-based analysis.

Access to large digitalised corpora and tools for processing them marked the future of phraseology as a field relying mostly on the analysis of long running texts and the patterns emerging from them, rather than on intuition for identifying what is worth

studying and teaching (Biber et al, 1999; Simpson-Vlach & Ellis, 2010; Sinclair, 1991). Corpus studies increasingly support the idiom principle confirming that the language user is not entirely free to select words: rather, words are co-selected to construct meaning in context (Hoey, 2005; Sinclair, 1991). Tognini-Bonelli (1996) observes that “[w]ords do not often exist as single choices; they entertain close relationships with other items: because of the cumulative effect of repeated occurrences, they end up carrying a meaning that goes beyond their traditional dictionary definitions” (p. 216). Modern phraseology relies on many resources to obtain more objective results in studying the function and meaning of multi-word sequences.

One of the first monographs on modern phraseology, edited by Cowie (1998), highlights the three major traditions of research in phraseology, i. e., the "classical" Russian tradition with its classifications and definitions of phraseological units, the anthropological strand, and the frequency-based approach which is at the moment the most popular in supra-lexical studies (Cowie, 1998).

The Western school of phraseology stems from the concept of collocation, commonly attributed to Firth, although its use has been documented well before Firth’s time: Hoey (2005) finds its occurrences as early as the first decades of the twentieth century, while Nation (2001) states that it was noted in the 1750 Oxford English Dictionary. It is nevertheless true to say that the popularity of the concept of collocation owes much to the work of Firth (1962). A collocation is a linguistic

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

phenomenon in which one word or lexical item tends to occur in company of another particular word and not its synonyms, without any apparent reason that could be attributed to semantic, syntactic, or conceptual factors. It is a matter of usage rather than anything else (Granger, 1998). Or, as Stubbs puts it,

“[Collocation is] a lexical relation between two or more words which have a tendency to co-occur within a few words of each other in running text.” (2001: p. 24).

Nattinger and DeCarrico (1992) define collocation in terms of frequency:

“If it is the case that the node word occurs with a span of particular words at a frequency greater than chance would predict, then the result is collocation.” (p.20)

Collocations are among the most frequent linguistic phenomena occurring in all types of texts and registers, but their systematic study, like that of other phraseological units, began relatively recently. This despite the fact that, as Hoey (2005) observes, it is virtually impossible to construct or even imagine a text without any collocations. That said it is also extremely difficult to find one composed exclusively of collocations or prefabs (Erman & Warren, 2000).

Two seminal publications in the late twentieth century shaped the future of phraseology in Western linguistics. The first was Pawley and Syder’s (1983) study where they noted that the difference between a native and non-native speaker (or the “puzzle of nativelike selection”, p. 193) lies in the use of institutionalised phrases (“sentence stems”) which make the speech of native speakers natural and fluent: they

are stored in the native speakers' memory as preconstructed chunks, while non-native speakers tend to collocate words in an unnatural and non-nativelike way because they create utterances out of single words instead of using prefabricated sequences. Non-native speakers, according to Pawley and Syder, do not have this stock of formulae at their disposal and hence cannot reproduce them automatically from memory. Thus, idiomaticity, manifests through a body of "lexicalised and institutionalised sentence stems" (Pawley & Syder, 1983: 192) and becomes central to nativelikeness and fluency; indeed, the two concepts are often treated as parallel, if not synonymous: "[conventionalised language forms] make language performance appear 'native-like', hence the notion of 'idiomaticity'." (Yorio, 1989:56).

The second influential publication was Nattinger and DeCarrico's study of lexical phrases (1992), in which they argue that 'lexical phrases' are ubiquitous and pervasive linguistic phenomena that are stored holistically in memory. They argue strongly that language teaching should focus on these 'lexical phrases' instead of the sporadic or anecdotal occurrences of picturesque and rare idioms. Nattinger and DeCarrico classify lexical phrases according to four dimensions: length and grammatical status, canonical or non-canonical shape, variability or fixedness, continuity or discontinuity (see Table 1.4 in the following section). They proposed a method of teaching English using the 'lexical phrases' they had extracted from a corpus of authentic spoken academic English, thus rendering their study a pioneering piece of research which inspired much of the phraseological investigation that followed it.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Thus, in the last decades of the twentieth century, several researchers (Granger, 1998; Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Sinclair, 1991) called for a more thorough investigation of ‘prefabs’ or formulaic language highlighting the importance of several issues in this area, such as the need to describe and classify the English phraseological stock, the importance of exploring learner collocations in English as a foreign language, and exhaustive descriptions of the learners’ mother tongues (Granger, 1998). This trend became powerful in linguistics through the work of many outstanding scholars whose contribution will be discussed below.

Further development of the study of collocations was largely led and inspired by John Sinclair who made outstanding contributions both to phraseology and corpus linguistics. Sinclair’s (1991) idiom principle had a significant impact on the focus and shape of a number of linguistic disciplines, including phraseology, corpus linguistics, and language teaching. The idiom principle suggests that speakers and writers draw on “a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.” (Sinclair, 1991: p. 110). Sinclair contrasts this with the open choice principle, i.e., the assumption that each item is chosen freely and placed into a ‘slot’ within the utterance depending on the will of the speaker or writer (p. 110). Many studies support the idiom principle by showing that idiomaticity or formulaicity are not just marginal and negligible phenomena, as had been thought previously (Granger & Paquot, 2008). The idiom principle and the open choice principle are two major forces which govern the language, but the idiom principle seems to be the more

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

influential of the two, that is to say, there seem to be more constraints on the speaker / writer at any given moment of time than free choice options.

Erman and Warren's (2000) influential article also addresses the two principles formulated by Sinclair (1991); through a manual analysis of 19 extracts of 100-800 words representing both speech and writing, they show that over 55% of the words in their corpus represent 'prefabs'. There are certain differences between spoken and written texts, but these are not as significant as might be expected. Notably, spoken language employs more prefabs, which may be due to the fact that speakers are generally subject to different types of pressure and face the necessity of coping with them rapidly, hence they have recourse to standard and predictable phrases to save encoding and processing time (Altenberg, 1998; Wray, 2002) and increase their fluency (Pawley & Syder, 1983). In the words of Altenberg (1998):

Speakers engaged in spontaneous interaction are in constant need of easily retrieved expressions to convey their intentions and reactions in discourse. At their disposal they have a large stock of recurrent word-combinations that are seldom completely fixed but can be described as 'preferred' ways of saying things -- more or less conventionalized building blocks that are used as convenient routines in language production. These building blocks come in all forms and sizes, from complete utterances to short snatches of words, and they display varying degrees of flexibility (pp. 121-122).

Hoey (2005) describes this phenomenon in terms of lexical priming:

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

“Every lexical choice starts off a series of options and predilections that result in an amazing fluency in any situation in which the speaker has been primed to perform” (p. 163).

Speakers show higher degrees of fluency when they speak of familiar topics in familiar contexts; in less familiar settings the pressure increases, and they “are more likely to produce dysfluencies” (Pawley & Syder, 1983: 207).

However, written language is also characterised by a large number of formulae, although not as much as speech (Erman & Warren, 2000). A writer is exposed to less pressure than a speaker who has to make decisions in real time, often without prior rehearsal, and is deprived of the possibility of deleting an inappropriate utterance and replacing it with a more adequate one, but both speech and writing are governed predominantly by the idiom principle which restricts the available lexical choices at each moment. It has been argued that due to these restrictions native speakers sound fluent and native-like, and that excessive reliance on the realisation of the creative potential of syntactic rules (Sinclair’s open choice principle) would result in unnatural and non-native-like speech, whereas the use of institutionalised ‘sentence stems’ is what makes their speech fluent and idiomatic (Pawley & Syder, 1983; Stubbs, 2001). Corpus studies confirm this view by yielding large numbers of “conventional combinations” (Stubbs, 2001: 73), rather than idiosyncratic and creative uses.

Idiomaticity is sometimes understood as semantic opacity, semantic or syntactic irregularity, or non-compositionality: if the meaning of a phrase cannot be deduced

from the individual meanings of its components, i. e., it is non-compositional, then it is regarded as idiomatic. Indeed, it was interpreted in that way in the early stages of research in this area, when phraseological studies were limited largely to the study of idioms. However, as Erman and Warren (2000) point out, semantic or syntactic irregularity makes the phrase more conspicuous and salient, but it is not the only indicator of idiomaticity. Consider the widely-cited examples of *strong tea*, *make a decision*, or *take a step*, all of which, are perfectly analysable and transparent expressions, at the same time show certain degrees of idiomaticity in that their constituents prefer the company of each other over synonymic items (we do not say **powerful tea*, **perform a decision* or **obtain a step*, though in languages other than English such combinations might well be possible). Erman and Warren (2000) suggest that idiomaticity has been treated as a marginal phenomenon because of this notion of non-compositionality, although corpus studies increasingly support the view that a sequence does not have to be opaque or non-compositional to be idiomatic (Hoey, 2005; Sinclair, 1991).

Formulaic sequences, besides adding idiomaticity and fluency to the utterance, have also been shown to contribute to accuracy in the speech of second language (L2) users due to the fact that these sequences are stored and reproduced holistically (Hunston, 2002; Jiang & Nekrasova, 2007; Nattinger & DeCarrico, 1992; Schmitt & Underwood, 2004; Underwood, Schmitt & Galpin, 2004; Wray, 2002), therefore, a single mental effort is required for producing a whole phrase instead of making separate efforts to produce single words. This view relates to the idea that human

memory is capable of retaining 7 ± 2 items at a time (Miller, 1956), hence producing 7 separate words would probably exhaust capacity, while producing a chunk of 7 words would occupy only one 'slot' in the short-term memory capacity (Hunston, 2002; Nattinger & DeCarrico, 1992).

Wray demonstrates that these formulaic sequences allow the speaker to organize his or her speech, and save processing time for the listener by enabling her to process it efficiently and more rapidly. This finding is also supported by the study carried out by Tremblay, Derwing, Libben and Westbury (2011) who in a series of experiments demonstrated that formulaic sequences require less time for processing than other similar structures which are not as frequent. Whether this is due to their being stored holistically or to the decrease in time required for computing them is unclear, but formulaic sequences are reliably shown to provide a processing advantage, as is the case with collocations in general (Nation, 2001). Jiang and Nekrasova's (2007) study confirms that native speakers of English require less time to make grammaticality judgements of a phrase if it is formulaic; furthermore, their non-native speaker subjects also demonstrated that effect.

Thus, formulaic sequences function as shortcuts for speech production and processing, time-buyers, and / or devices for manipulation of information (Nation, 2001; Wray & Perkins, 2000; Wray, 2002). Language learners, be they children acquiring the first language or adult or child learners acquiring a second, pass through four phases in the process of the acquisition of formulaic sequences. The first phase

involves, according to Wray and Perkins (2000), the mostly holistic approach towards the language, with the use of unanalysed chunks as the basic unit of input and output. During the second stage, the learner begins to analyse the chunks which she encounters in everyday conversations or other sources of input, and during the third phase, these chunks begin to form part of the user's conscious competence. The fourth phase (which remains unattained for many L2 and foreign language learners) is the formation of a lexicon or linguistic repertoire which consists mainly of prefabricated units, analysed and learned previously, but perceived as single units at this stage, shortcuts for language comprehension and production. The user in this phase analyses only what is novel, while the greater part of the language is received and produced in unanalysed chunks which save effort and fulfil a number of important pragmatic and discursive functions both for the speaker and for the hearer. A more detailed description of these functions can be found in Table 1.2 in the next section.

The study of collocations has confirmed the assumption that meaning is not encoded in single words but in phrases (Firth, 1962; Hoey, 2005; Stubbs, 2001; Sinclair, 2008). Research on collocations demonstrates that “meaning is use” (Stubbs, 2001), and where there is no choice, there is no meaning (Sinclair, 2008), while Hoey’s theory of lexical priming reveals complex and unexpected relationships between seemingly unrelated lexical items. Hoey (2005) maintains that every word is primed to occur in company of certain collocates, in particular semantic sets and in association with certain pragmatic functions. Hoey explains the concept of semantic

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

prosody labelled after Firth’s phonetic prosody: each word “spreads over” (2005: 22) its neighbours in a way that some of its semantic features are contained in the neighbouring words, just as each sound in a word is not pronounced in isolation but is shaped by its neighbouring sounds and in its turn shapes them (Louw, 2000). Priming is always personal since it reflects the cumulative experience of the language user’s encounters with a particular lexical item. Thus, different individuals often have different primings for one and the same word, although there are many universal primings shared by most language users, especially natives. Hoey goes on to note that:

even when writers are straining at the limits of what a language is capable of expressing, they make use of more of their primings than they reject. A sentence that overrode all its reader’s primings would only be a sentence by virtue of starting with a capital letter and ending with a full stop; it would not correspond to anything recognisable as an instance of language in use (p. 177).

Both Hoey (2005) and Sinclair (2008) defend the view that different grammatical forms of the same lemma have different meanings (uses) to an extent that if they did not share the same spelling we might be inclined to regard them as different words.

Sinclair (2008) concludes that

a. many, if not most, meanings require the presence of more than one word for their normal realization;

b. patterns of co-selection among words, which are much stronger than any description has yet allowed for, have a direct connection with meaning” (p. 133)

Words combine with one another to create meaning, rather than comprising one or more fixed meanings within isolated lexical units (Sinclair, 2008; Stubbs, 2001). In the words of John Sinclair:

we have to concede that the normal primary carrier of meaning is the phrase and not the word; the word is the limiting case of the phrase, and has no other status in the description of meaning. (Sinclair, 2008, p. 409).

Sinclair writes about the three strata of meaning, the topmost of them being the contextual settings (genre conventions, assignment of fixed points of reference, such as personal pronouns and deixis in general), the second stratum comprising the phrase, or collocational frameworks, and the third, comprising lexis and grammar. The phrase, thus, becomes central in the description of meaning.

Cosme and Gilquin (2008) show that even minor word classes, such as prepositions, reveal consistent tendencies for favouring certain combinations and constructing meaning through those combinations. Their contrastive analysis of the English preposition *with* and its intuitive counterpart in French, *avec*, reveals important differences in the use of these seemingly equivalent prepositions across the two languages. Their study shows that direct translation of *with* as *avec* occurs only in 30% of all the instances, while *avec* is translated as *with* in 60% of occurrences. The authors argue that this may be due to two reasons: the polysemy of *with* resulting in

different meanings covered by other prepositions (or even other linguistic means) in French, and the greater number of bound uses of *with*, as opposed to free uses. Furthermore, the authors point out that these bound uses are predominantly characteristic of nominal phrases (e.g., *with ethnic tensions thus exacerbated*).

Thus, prepositions appear to have two types of meanings: the core meaning, being used mostly in free combinations and having typically direct equivalents in another language, and the metaphoric meaning, used more often in bound combinations and with no direct equivalents in other languages. The metaphoric meanings, hence, need to be translated using other linguistic resources and are likely to cause problems to language learners. These metaphoric meanings should be included in curricula and be explicitly taught to learners, the authors conclude.

This section attempted to discuss the growing interest in phraseology based on large corpora over the past decades and the key aspects of formulaic sequences described in a range of studies, such as the prevalence of prefabricated chunks in native speech and writing, the idiom principle, the benefits of using formulaic sequences for achieving idiomaticity, native-likeness and gaining processing advantage. It also briefly addressed the theory of lexical priming and the construction of meaning through patterns of co-selection rather than through isolated words. As different uses and meanings began to emerge through corpus analysis, researchers representing different linguistic trends and approaches became increasingly concerned with the study of different types of word combinations under a variety of different names. The

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

following section will address the terminology and classification of formulaic sequences represented in the work of different authors.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

2.2. TERMINOLOGY AND DEFINITIONS OF FORMULAIC LANGUAGE

In the previous section, it was shown that the growing understanding of the ubiquitous nature of prefabricated sequences has gradually led to increasing interest towards these lexico-syntactic units among scholars working in a range of linguistic traditions and approaches. The differences in focus and approach have resulted in a wide variety of terms and classifications describing different types of formulaic language. Among these are “lexical phrases” (Nattinger & DeCarrico, 1992), “sentence stems” (Pawley & Syder, 1983), “phraseologisms” (Gries, 2008), “recurrent word combinations” (Ädel & Erman, 2012; Altenberg, 1998; Conrad & Biber, 2004; Granger & Lefer, 2016), “phrasal lexemes” (Moon, 1998a, b) “phraseological units” (Cowie, 1998), “prefabricated expressions / language / patterns”, or prefabs (AlHassan & Wood, 2015; Coxhead & Byrd, 2010; Granger, 1998; Wray, 2002), “set expressions” (Gläser, 1998; Howarth), “idioms” (Wulff, 2010) and “restricted collocations” (Howarth, 1996, 1998; Kuiper, 2004; Siyanova & Schmitt, 2004).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Many studies point to the consistent disagreement in the field as to how to define, classify and deal with formulaic language:

“Phraseology is a field bedevilled by the proliferation of terms and by conflicting uses of the same term” (Cowie, 1998, p. 210)

This confusion in terminology is mentioned by many other researchers (Ädel & Erman, 2012; Granger & Paquot, 2008; Moon, 1998; Wray, 2002). It may partly be due to the Soviet origins of phraseology and its being taken up by Western linguists as a complex but largely theoretical and abstract discipline which, despite its valuable theoretical framework and developed body of abstract concepts, did not meet the demands of Western linguistic research, particularly the branch of Applied Linguistics concerned primarily with the application of research results to the foreign and second language classroom and to other real life contexts. Be that as it may, researchers in the field have developed varying terminologies which sometimes overlap by using different terms for the same linguistic phenomenon, and sometimes use the same word to denote entirely different types of sequences (see Table 1.2). Wray (2002) has identified over 50 distinct terms referring to formulaic sequences in a wide range of studies. Furthermore, any attempt to organize the existing terminology or suggest a new one for different items results in an increase in the number of classifications and terminologies thus causing further confusion.

In a bid to avoid arcane debate, most studies prefer to adopt a terminology which is most suitable for their own purposes without discussing the matter any further (Granger & Paquot, 2008). An additional and potentially more serious problem is

that linguists working in the field of phraseology often fail to specify the criteria according to which they extract or identify phraseological units, thus exposing themselves to the accusation of vagueness (Gries, 2008).

The most general terms used for referring to these large lexical units are “formulaic expressions”, “formulaic language”, “formulaic sequences” (Wray, 2002), or simply “formulae” (Simpson-Vlach & Ellis, 2010). These cover a wide range of phrases, including idioms (*kick the bucket*), routine expressions (*of course*), politeness formulae (*thank you very much*), noun phrases, verb phrases, adverbial phrases, as well as all kinds of lexico-syntactic constructions (*I don't know if you*), structurally complete or incomplete, completely fixed or relatively free collocations, alongside a number of lexico-grammatical patterns with fixed and variable slots (*in this paper / study / chapter we / I discuss / address*). As can be seen from the above description, formulaic sequences come in all shapes and sizes and are extremely diverse and difficult to pin down (Cowie, 1998). Different definitions underscore such characteristic features of formulaic sequences as their prefabricated nature (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Wray, 2002), frequency (Biber et al., 1999; Cortés, 2004, 2006, 2008, 2013; Salazar, 2011, 2014), idiomaticity or semantic opacity (Mel'cuk, 1998; Moon, 1998), fixedness or boundness (Howarth, 1996), continuity and discontinuity (Biber, 2009), etc. One of the most extended definitions is that of Wray (2002):

“a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored

and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.”

(p. 9)

This is also one of the most general and inclusive definitions that has been encountered in literature, as well as one of the most influential. However, Read and Nation (2004) have the following observation to make:

[identifying such sequences] is a challenging goal because the means of storage and retrieval of the same sequence can differ from one individual to another, and can differ from one time to another for the same individual depending on a wide range of factors such as changes in proficiency, changes in processing demands, and changes in communicative purpose (p. 25).

Erman and Warren (2000) also observe that the identification of prefabs is a challenging task for two main reasons: firstly, what is formulaic for one speaker may be non-formulaic for another, even if the speakers are L1 users of the same language and members of the same discourse community; secondly, it is easy to overlook many prefabs because of their high frequency and apparently non-compositional and non-salient character. Furthermore, formulae differ from morphemes or syntactic units in that they are probabilistic and greatly depend on the context, corpus and criteria set by the researcher.

In other words, defining a sequence as ‘formulaic’ becomes a highly subjective and context-dependent activity without any formal criteria (note the “is or appears to be prefabricated”). It is not clear how one is to decide whether this or that sequence

“appears to be” prefabricated, or even who is in a position to make such a decision. Another dimension to this challenge is added if we consider some types of formulaic sequences which transcend the boundaries of grammatical structures (Biber et al., 1999; Wray & Perkins, 2000), and in conversation they may even transcend turn boundaries (consider the ritualized sequences such as *Thank you very much. – You are welcome*).

As can be seen, the difficulty in distinguishing formulaic sequences from non-formulaic language has led to a vast number of definitions, each based on criteria favoured by this or that author, and each suggesting a different focus and a different term for formulaic sequences. Some of the most extended terms and definitions or descriptions of formulaic sequences are covered in Table 1.1.

Table 1.1. Definitions and descriptions of formulaic sequences.

Study	Term	Definition / description
Pawley & Syder, 1983	sentence stems	a unit of clause length or longer whose grammatical form and lexical content is wholly or largely fixed; its fixed elements form a standard label for a culturally recognized concept, a term in the language
Yorio, 1989	conventionalised language	(1) They are thought to be learned as wholes, largely unanalyzed. (2) They cannot always be analysed like other lexical items or syntactic strings. They are often exceptional or constructed syntactically, semantically or situationally. (3) They make language performance appear “native-like”, hence the notion of “idiomaticity”

Study	Term	Definition / description
Nattinger & DeCarrico, 1992	lexical phrase	chunks of language of varying length, conventionalized structures that occur more frequently and have more idiomatically determined meaning than language that is put together each time
Howarth, 1996	collocation	a general term covering the whole continuum from free word-combinations to idioms
Altenberg, 1998	recurrent word combinations	sequences which occur more frequently than expected by chance
Gläser, 1998	phraseological unit	a lexicalized, reproducible billexemic or polylexemic word group in common use, which has relative syntactic and semantic stability, may be idiomatized, may carry connotations, and may have an emphatic or intensifying function in a text
Mel'cuk, 1998	phraseme, set phrase	a phrase which is not free
Moon, 1998a	phrasal lexeme	the whole range of fixed and semi-fixed complex items which dictionaries in the Anglo-American tradition classify and treat as 'phrases' or 'idioms' (according to their metalinguistic terminology)
Moon, 1998b	fixed expression	a general term used to cover several kinds of phrasal lexeme, phraseological unit, or multi-word lexical item: that is, holistic units of two or more words
	idiom	refers loosely to semi-transparent and opaque metaphorical expressions such as <i>spill the beans</i> and <i>burn one's candle at both ends</i>
Granger, 1998	formula	phrases that have a pragmatic rather than a syntactic function
	collocation	the linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than its "synonyms" because of constraints which are not on the level of syntax or conceptual meaning but on that of usage

Study	Term	Definition / description
Biber et al., 1999	lexical bundles	multi-word sequences occurring with much higher frequency than expected by chance in a given text type
Erman & Warren, 2000	prefab	a combination of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization.
Nation, 2001	collocation	any generally accepted grouping of words into phrases or clauses
Stubbs, 2001	collocation	a lexical relation between two or more words which have a tendency to co-occur within a few words of each other in running text
Wray, 2002	formulaic sequence	a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use
Granger & Paquot, 2008	Phraseme, phraseological unit	the vast variety of word-combinations and collocations addressed in phraseological research of different trends
Gries, 2008	phraseologism	the co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or sentence and whose frequency of cooccurrence is larger than expected by chance
Tognini-Bonelli, 1996	collocation	the most immediate patterning, simple word co-occurrence
	colligation	co-occurrence at the level of word and sentence classes
	lexical set	the coming together of grammatical classes and lexical items

Most studies focus on one or more main features to identify formulaic sequences, such as their prefabricated nature, their being stored and reproduced holistically,

semantic unity, size, (non-)compositionality, irregularity, and frequency, among others.

There is strong evidence that formulaic sequences are stored and processed holistically, as unanalysed chunks or lexicalised units (Ellis, 2002; Ellis, Simpson-Vlach & Maynard, 2008; Granger & Meunier, 2008a, 2008b; Jiang & Nekrasova, 2007; Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Wray, 2002). Jiang and Nekrasova (2007) conducted two experiments among native and advanced non-native speakers of English, the first aimed at measuring processing times for prefabricated units as compared to non-prefabricated chunks and ungrammatical utterances. The second experiment was aimed at revealing whether the shape of the words affected the processing time, for which purpose the authors used a number of formulaic sequences, non-formulaic sequences, and ungrammatical sequences, all represented in upper case. It was assumed that this modification would counterbalance the effects of the shape of words. Both the first and the second experiments confirmed the holistic hypothesis for native speakers, as well as for the L2 users. The experiments showed a clear processing advantage for formulaic sequences which also seems to affect L2 users. The authors propose two explanations for this effect: on the one hand, the formulaic sequences may have been learned directly as such, that is, without analysing them into fragments, together with their translation into the learners' L1. On the other hand, they may have been analysed first and then 'glued together' to become lexicalised wholes. In any case, the processing advantage for both groups (L1 and L2) was confirmed.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Wray (2002) highlights the heterogeneous character of the formulaic sequences and their different forms and functions in discourse. Among these she cites “situation manipulators” or “hearer manipulators”, namely sequences which get people “to do, think or feel things” (Wray, 2002: 88). These are typically such as politeness formulae.

Wray argues that the difficulty of identifying and classifying formulaic sequences may be due to the fact that formulaic sequences are essentially non-linguistic phenomena and one should not attempt to define them through linguistic categories. Rather, she suggests looking at formulaic language from the point of view of function. Wray demonstrates that there are several functions formulaic strings may perform, all of which ultimately can be said to serve the single goal of promoting the speaker's interests. The functions listed by Wray are as follows:

- having easy access to information (via mnemonics, etc.);
- expressing information fluently;
- having physical and emotional needs satisfactorily and promptly met;
- being provided with information when required;
- being perceived as important as an individual;
- being perceived as a full member of whichever groups are deemed desirable.” (pp. 95-96).

Wray & Perkins (2000) list another set of functions performed by formulaic sequences. These are shown in Table 1.2.

Table 1.2. Functions of formulaic sequences (adapted from Wray & Perkins, 2010)

Function	Effects	Type	Examples
Manipulation of others	Satisfying physical, emotional and cognitive needs	Commands Requests Politeness markers Bargains, etc.	Keep off the grass; hand it over
Asserting separate identity	Being taken seriously Separating from the crowd	Story-telling Turn claimers and holders, etc. Personal turns of phrase.	You're never going to believe this, but. . . Yes, but the thing is... Thank you very much (<i>in response to invitation to speak</i>); The first thing that you have to realize, of course, in addressing this issue is. . I wanna tell you a story; You know what I mean, Harry
Asserting group identity	Overall membership Place in hierarchy (affirming and adjusting)	'In' phrases Group chants Institutionalized forms of words, etc Ritual Threats Quotation Forms of address Hedges, etc.	Praise the Lord!; as the actress said to the bishop We are the champions Happy birthday; dearly beloved, we are gathered here today. . . Our Father, which art in Heaven. . . I wouldn't do that if I were you "I wouldn't want to belong to any club that would have me as a member" Your Highness Well I'm not sure (<i>as a polite denial or refusal</i>)
Processing short-cuts	Increased production speed and / or fluency	Standard phrases (with or without gaps)	Put the kettle on, will you?; I have known __ for __ years in my capacity as __

Function	Effects	Type	Examples
		Standard ideational labels with agreed meanings	Personal computer; bullet point; the current economic climate
Time-buyers	Vehicles for fluency, rhythm and emphasis Planning time without losing the turn	Standard phrases with simple meanings Fillers Turn-holders Discourse shape markers Repetitions of preceding input	Make a decision; draw a conclusion; a sea change; at the end of the day (<i>in the sense of 'really'</i>); one way and another If the truth be told; if you want my opinion; if you like And another thing; and let me just say.... There are three points I want to make. Firstly. . . Secondly. . . Thirdly/Lastly. . . (A: What's the capital of Peru?) B: What's the capital of Peru? (Lima isn't it?)
Manipulation of information	Gaining and retaining access to information otherwise unlikely to be remembered	Mnemonics Lengthy texts one is required to learn Rehearsal	Thirty days hath September...; Richard of York gave battle in vain Shall I compare thee to a summer's day? <i>Rehearsing a telephone number while looking for a pen</i>

As can be seen from Table 1.2, many functions identified by Wray (2002) and Wray and Perkins (2000) are similar, and some even overlap: *manipulation of others* may correspond to *having physical and emotional needs satisfactorily and promptly met*, *asserting group identity* and *asserting separate identity* seem to be forms of *being*

perceived as a full member of whichever groups are deemed desirable and being perceived as important as an individual, whereas manipulation of information can be regarded as a way of having easy access to information, etc. However, there is no full correspondence between these two sets of functions.

With regard to academic discourse, perhaps the most important of these functions is that of asserting group identity: novice researchers should be able to comply with the conventions of the genres they seek to instantiate and the discourse community they aspire to become part of in order to achieve the status of membership and at a later stage expertise in the given community (Bhatia, 1993; Swales, 1990).

Thus, formulaic sequences perform a number of important functions that transcend the boundaries of discourse. Taking into account the diversity of functions and types of formulaic sequences, Wray (2002) develops an integral model for acquisition of formulaic language for different types of learners. This she calls a Heteromorphic Distributed Lexicon. Wray's model integrates five lexicons (grammatical, referential, interactional, memorised, and reflective), each of which involves three types of elements: morphemes, polymorphemic words, and formulaic sequences. The distribution of the element types in lexicons is not homogeneous: some have more of one type, others are composed principally of the other type, but all the three types are present in all five lexicons and perform different functions. It is noteworthy that Wray considers formulaic any sequence of linguistic elements which need not be analysed, although it may be perfectly analysable into smaller segments, only by virtue of being

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

perceived as a whole. Wray's model is a complex and dynamic system which sheds light on the nature and functions of formulaic language and allows a more specific focus of teaching formulaic sequences (teaching for formulaicity, teaching for idiomaticity, teaching for linguistic knowledge). She herself acknowledges that her model is, nevertheless, far from straightforward and is still open to debate (Wray, 2002).

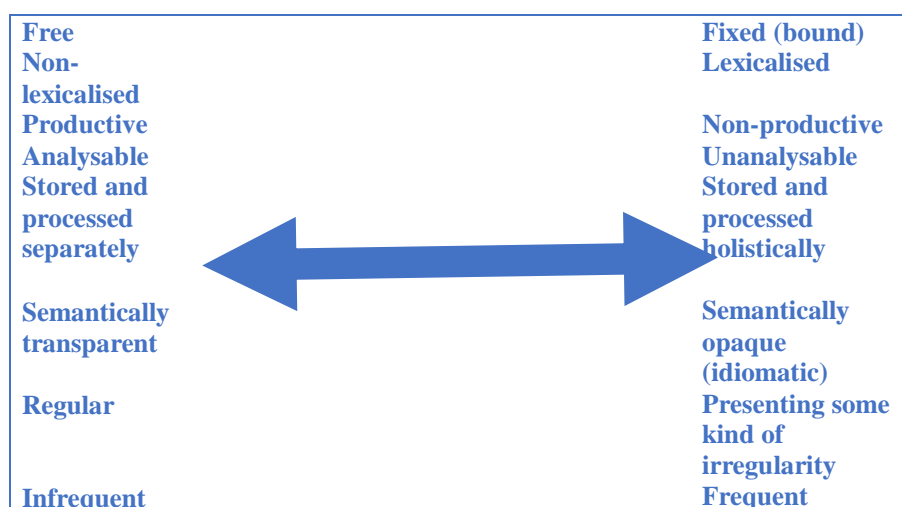
Prioritising two or more features, such as idiomaticity, irregularity, frequency, and so on, is the basis for the classification of formulaic sequences in many studies. These will be discussed in the remainder of this section.

An important point must be made before proceeding to the issue of classification of formulaic sequences. Most researchers agree that there are no clear-cut boundaries between free and bound word combinations, as well as between the types of word combinations. In fact, *fuzzy* is one of the most recurrent adjectives found in publications on this topic (Ädel, 2014; Altenberg, 1998; Cowie, 1998; Handl, 2008; Howarth, 1996; 1998; Meunier & Granger, 2008; Salazar, Stubbs, 2001; 2014; Wray, 2002; Wulff, 2010), along with *pervasive* and *ubiquitous*.

Drawing boundaries between the formulaic and novel or creative utterances is extremely challenging, even having set solid criteria for their identification. Various authors deal with the fuzzy area between idioms and free collocations differently, according to their purposes and the perspective of their study. Most classifications refer to a continuum or cline (Cowie, 1998; Biber, 2009; Ellis, 2002; Erman &

Warren, 2000; Fontenelle, 1998; Handl, 2008; Howarth, 1998; Meunier & Granger, 2008a, b; Salazar, 2014; Stubbs, 2001) across which different types of formulae are distributed according to the extent to which they represent the selected feature (fixedness, idiomaticity, frequency, etc.). Figure 1.1 is an attempt to illustrate this approach.

Figure 1.1. Parameters of classification of formulaic sequences along continua.



In his analysis of a number of phraseological and collocational dictionaries and their presentation and classification of phraseological units, Cowie (1998) argues that the main problem of classification is “the fuzzy zone between the fixed and the free” (p. 227) which keeps presenting difficulties when compiling such dictionaries. Collocational and phraseological dictionaries have been critiqued in a number of studies which pointed out the lack of clarity in selection and organisation criteria and

the focus on infrequent phrases and non-authentic examples (Howarth, 1996; Sinclair, 1991).

Phraseological dictionaries are few and arbitrarily organized in terms of what should be included and what should not. The traditional monolingual or bilingual learner dictionaries present even greater problems for learners (Siepmann, 2008; Sinclair, 1991), such as being static (rigidly organized and typically serving only one purpose), centralized (dependent on the entry headwords) and passive (the user needs to identify a gap in his / her knowledge in order to consult a dictionary and will not consult it if no gap is felt to exist), as described by Wible (2008).

Granger and Paquot (2008) argue that different studies select and prioritise one of the following five features to identify and classify formulaic strings: “(1) internal structure (e.g. verb + noun or verb + preposition); (2) extent: phrase- vs. sentence-level; (3) degree of semantic (non-) compositionality; (4) degree of syntactic flexibility and collocability; (5) discourse function” (p. 35).

A definition and classification of formulae, therefore, will necessarily depend on the data involved in a particular study, on the purposes of the study in question, the method and the framework used in the study. (Wray, 2013). As Wray and Perkins (2000) put it, “[i]t is all a question of emphasis” (p. 13).

Formulaic sequences, according to Wray (2002), always present some kind of irregularity:

If there is a standard view of what formulaic language is (and the range of descriptions reviewed in this book must cast severe doubt on that), at its heart will be something about word strings which ‘break the rules’. They can break phonological rules, by displaying fewer stresses than expected and by being articulated faster and less clearly. They can break syntactic rules, by resisting pluralization, passivization and so on, and by containing constituents which do not take on their normal grammatical function. They can break lexical rules, by containing items which are archaic or have no independent existence. And they can break semantic rules, by combining to mean something other than they ought to, and by being more idiomatic than an equivalent nonformulaic combination (p. 261).

The majority of studies of formulaic sequences focus on a combination of two or more features to identify and classify formulae. However, certain studies only rely on one parameter.

Mel’cuk (1998), for example, constructs the meanings and the structure of formulaic sequences (which he calls “phrasemes” or “set phrases”) on the basis of lexical functions identified for each phraseme. His classification of word-combinations includes free phrases and phrasemes or set phrases, the former defined as unrestricted and regularly constructed and the latter as fixed and more or less unalterable, in line with the above-mentioned ‘free versus bound’ continuum. The fuzzy space between free and fixed phrases, according to Mel’cuk, is occupied by collocations.

Phrasemes, furthermore, are classified into pragmatic and semantic set phrases, the latter being subdivided into idioms (totally fixed), quasi-idioms (semi-fixed), and collocations (relatively loose but presenting important semantic and cultural features in a given language or culture). The classification is based on the lexical functions of each unit defined through the property of fixedness. Although Mel'cuk's classification might seem both very general and very formal, as well as somewhat unrelated to the present study, it is useful for defining the boundaries of phraseological units, an issue which, according to a number of researchers, constitutes a challenge (Erman & Warren, 2000; Lindstromberg, Eyckmans, & Connabeer, 2016).

Nattinger and Carrico (1992) use four parameters to classify formulaic sequences, which they call 'lexical phrases': length and grammatical status, canonical or non-canonical shape, variability or fixedness, continuity or discontinuity. Table 1.3 illustrates their classification.

The authors propose a language teaching programme based on presenting these unanalysed chunks extracted from a corpus of authentic university lectures and classified according to the above-mentioned four parameters for pedagogic purposes. This approach allows second and foreign language learners to acquire the foreign language in a way similar to first language acquisition.

Table 1.3. Nattinger & DeCarrico's (1992) classification of lexical phrases according to four dimensions (adapted from p. 45).

	Grammatical level	Canonical / Non-canonical	Variable / Fixed	Continuous / Discontinuous	Examples
Polywords	word level	both	fixed	Continuous	<i>for the most part, by the way, by and large, as it were</i>
Institutionalised expression	sentence level	canonical	fixed	Continuous	<i>How do you do? Give me a break. long time no see</i>
Phrasal constraints	word level	both	somewhat variable	mostly continuous	<i>a _____ ago, to _____ this up: to tie this up, to wrap this up, as I was _____ saying / mentioning</i>
Sentence builders	sentence level	canonical	highly variable	often discontinuous	<i>I think that, not only X, but also, Y, I am a great believer in X</i>

Nattinger and DeCarrico convincingly argue that the students' knowledge of formulaic sequences results in dramatic improvement in all four basic language skills: reading, writing, listening and speaking, as it increases their level of comprehension and improves language processing skills and fluency. With regard to the latter, the authors make the following interesting observation:

Fluency is not so much the ability to coin appropriate phrase for novel situations as it is the ability to retrieve from a wide assortment of lexical phrase variants for recurring situations. It is discouraging and tiresome for language learners (and their audience!) to have to use the same phrases repeatedly for the same functions (p. 168).

This seminal publication, alongside Pawley and Syder's (1983) study and John Sinclair's fundamental research, has inspired much of the phraseological research discussed in this chapter.

The phraseology in learner writing has been studied by many linguists, one of the first of whom is Howarth (1996), whose classification of word-combinations comprises five types distributed along a continuum of boundedness, from completely free to almost totally fixed units, or idioms (Fig. 1.2). The middle part of the continuum is occupied by restricted collocations. Howarth thus follows the classification of phraseological units proposed by the Soviet linguists (Vinogradov (1947) and others): free combinations, restricted collocations, and idioms, the latter being further subdivided into figurative idioms and pure idioms.

Figure 1.2. Classification of word-combinations according to their degree of fixedness (adapted from Howarth, 1996).



Howarth analyses the use of word-combinations in native and non-native learner academic writing. The study examines verb + noun collocations taken from two corpora, one produced by native writers, the other consisting of essays written by a

group of proficient international students of social sciences taking the author's EFL course and coming from seven different countries with different cultures and sociolinguistic backgrounds. The social sciences were chosen due to the author's conviction that the students of natural sciences, such as physics or biology, seem to have fewer problems with lexical and collocational features. This view is supported by Kuteeva and Negretti (2016) who suggest that acquiring genre knowledge poses more challenges for L2 writers in the humanities and social sciences than it does for those working in the natural sciences.

Howarth (1996), along with many other researchers, defines phraseological units on the basis of pure idioms, i.e., completely opaque, unmotivated and fossilised expressions, as opposed to 'free combinations'. Idioms seem an arbitrary starting point, given their low frequency and marginal character in the language in general, but this starting point is not infrequent in phraseological research. The categories for defining phraseological units proposed by Howarth are as follows: syntactic patterns, institutionalisation, semantic transparency, commutability (collocational preferences), semantic unity, and motivation.

Of these, that of commutability seems to be of particular importance. According to this parameter, there are five levels of collocations, ranging from almost free (level 1: there is a wide range of nouns that collocate with the given verb) to very restricted (level 5: the verb in this particular sense collocates with a very limited number of nouns). The results suggest that neither the free collocations nor the ones that are

closer to the restricted side of the continuum present major problems for learners. By contrast, the central area of the continuum, comprising the collocations which can be labelled as moderately restricted, is a greater challenge and should be addressed in learner's dictionaries and teaching materials.

Howarth notes that the level of linguistic proficiency of the students is not an indicator of their phraseological competence. Moreover, the more experienced and advanced the learner, the more deviant the phraseological behaviour she shows approaching the capacities of the expert English speaker who is ready to take risks and experiment with language by creating novel, non-standard collocations. The author explains that the most problematic areas of learner phraseology are the restricted collocations which are described as word-combinations whose constituents combine with a limited number of words. This phenomenon, similarly, is most patent in expert writers who have mastered the purely idiomatic part of the language and fully understand the characteristics of idioms, and, on the other hand, are fairly well-acquainted with the grammatical and lexical production of free word-combinations, while the middle part of the continuum, the restricted collocations, are a problematic issue for them. The main problem, Howarth maintains, is that EFL teachers, both native and non-native, are aware of this difficulty but not of the phraseological mechanisms of the language, thus being unable to offer any viable solution to the learners' needs in this respect.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

It is noteworthy that Howarth (1996) identifies several deviations, such as deconstruction and blending of restricted collocations, in the native speaker corpus. Although very small in number and thus negligible, these deviations suggest that phraseological knowledge is not a ‘default feature’ of native speakers, it is a learnable and teachable skill which requires both conscious attention and unconscious exposure to native-like collocations. Howarth attributes this to “a slip of the pen” (p. 93) due to the proximity of other words which would collocate perfectly with those that have been erroneously collocated. The author does not seem to refer to deliberate decomposition and non-standard usage of idiomatic expressions which may be used to achieve certain stylistic effects in the text, but to genuinely unintentional miscollocations. Hoey’s (2005) ground-breaking study on lexical priming addresses the issue from a slightly different focus, asserting that every user constructs his or her own language based on previous encounters with instances of speech and writing, and thus no two identical experiences are to be found. The primings of each speaker, then, native or non-native, will vary according to their experience, background and context, although there exist a number of “harmonizing mechanisms” (p. 181) for unifying the experiences of individuals in each culture. These are especially common for native speakers, while L2 users, as a result of their restricted contact with the target language, are more likely to show idiosyncratic primings affected by their L1 and not common for the majority of language users.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Overall, the L1 speakers in Howarth's (1996) study were shown to use about 50% more restricted collocations than L2 users, which correlates with the findings of other researchers who often emphasise the underuse of formulaic sequences by learners.

As far as non-native speakers are concerned, their performance seems to be weakest with respect to the middle of the continuum, where the semi-restricted collocations are located. The most interesting types of error, or deviations, as Howarth terms them, seem to be the overlapping of two collocations (*achieve tasks*) and blends (*take a hardline criterion against*). Both of these types are supposed to be based on collocations which exist in the language and are known to the learners, although not so well known as to allow for their correct use. This finding implies that the learners who make these collocational mistakes are actually more competent than those who rely on free collocations, since it is assumed that they are familiar with these collocations, while still unable to use them correctly.

Howarth's study concludes with a critique of collocation dictionaries and teaching materials which do not seem to follow any solid criteria in their selection of collocations. The author states that these resources are heterogeneous and present inadequate content which may mislead the learners and is no of particular use to them.

Granger (1998b) explores two types of word combinations, collocations and formulae, in a corpus of native and non-native (French) speaking learners' texts. Her results, perhaps predictably, show that the majority of learners tend to underuse

collocations which are recurrent in native data, and overuse several collocations that are relatively rare in native speech and relatively simple: consider the discussion of *very* as “the all-round amplifier *par excellence*” (Granger, 1998: 151). Another interesting finding is that learners are likely to (over)use expressions which correlate with certain expressions or have equivalents in their L1, thus converting the process of text production into a process of ‘calquing’. The collocations in Granger’s study are described as phrases whose components are more likely to co-occur as compared to their lexical synonyms, thus co-occurrence is a matter of usage rather than synonymy. Formulae in Granger’s study, on the other hand, are frames or patterns which display a high frequency in English, for example, “pronoun + verb (*think, claim, maintain*) + that”. These formulae would thus be roughly equivalent to Biber’s (2009) frames, while Nattinger and DeCarrico’ (1992) would call them “sentence builders”. Granger’s study is interesting for its pedagogical implications and the comparison of two learner subcorpora. An important characteristic of this study is the homogeneous character of the L2 corpus (all the L2 learner texts are in French) and its equivalence to the L1 subcorpus also covering learner texts of the same genre, in line with Connor and Moreno’s (2005) principles of *tertia comparationis*. The author concludes by underscoring the crucial difference between L1 and L2 learning, namely that a child uses two strategies to acquire a language, the analytic and the gestalt strategy, by acquiring chunks of prefabricated sequences (analytic) and then deconstructing them into meaningful parts (gestalt), and vice versa. In contrast, adult learners fail to develop analytic language based on the gestalt strategy. This

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

implication is important for understanding second-language acquisition mechanisms and designing teaching materials for adults or second-language learners in general. The author calls for more detailed analysis of prefabs both in English and in learner languages through larger corpora while also challenging researchers to arrive at better descriptions of learner use of prefabs in order to identify the transfer-related problems and other issues which may be of relevance in the teaching of prefabs.

A number of studies have focussed on formulaic sequences in conversation (Altenberg, 1998; Biber & Barbieri, 2007; Pawley & Syder, 1983). Altenberg's (1998) classification of recurrent word combinations is grammatical and includes the following types: dependent and independent full clauses, single and multiple clause constituents and incomplete phrases, some of which are further divided into various subcategories inspired by the generative approach. Altenberg notes that the speakers are subject to "various types of phraseological pressure" (p. 121) which leads them to use prefabricated expressions in order to fulfil their communicative, textual, interpersonal and other purposes. He also concludes that his study in fact confirms the "fuzzy character of phraseology", (p. 121) rather than clarifying it.

Moon (1998a) discusses the frequencies and forms of 'phrasal lexemes' which show the following frequency patterns: pure or classical idioms (*bite the dust*), are very infrequent; functional, rather than fully lexical expressions, or "frozen collocations" (such as *of course*, *at least*), tend to be the most frequent; whereas the more colourful, stylistically marked and metaphorical expressions (*call a spade a spade*) are often

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

exploited or manipulated in context. In her 18-million word corpus, only 18% of which are non-fiction texts (she provides no further specification regarding the nature of these texts), the most frequent grammatical category of “phrasal lexemes” is the predicate – a finding which may be said to indirectly correlate with that of Biber and Gray (2010): expert academic writing tends to rely on phrasal lexical bundles (nominal groups in Moon, 1998a), while other registers and non-expert academic writers employ clausal formulaic sequences (predicates, adverbials and clausal adverbials in Moon, 1998a) more extensively. Biber and Gray (2010) point out that this reliance on phrasal formulae is a distinctive characteristic feature of academic genres, whereas, for instance, fiction and conversation show a tendency to be more clausal. Other studies also support this view (Parkinson & Musgrave, 2014). Stubbs (2001) observes that “[c]omplex noun phrases are formed so that they can be used to refer to complex phenomena” (p. 117), and academic discourse is generally regarded as a complex register (Coxhead & Byrd, 2010), although its complexity is different from the elaborate constructions of other registers. Academic genres have developed syntactic patterns and discursive strategies which allow writers to express ideas in a concise and economical way, among these the use of nominal phrases instead of verbal groups and clauses (Biber, & Gray, 2010; Cortés, 2006; Salazar, 2014).

In another book-length study of Fixed Expressions and Idioms (FEI), Moon (1998b) specifies three principal factors for defining FEIs: institutionalisation, lexicogrammatical fixedness, and non-compositionality, and classifies them

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

according to discourse functions. The discourse functions identified by Moon are described in Table 1.4 below.

Table 1.4. Discourse functions of FEIs (adapted from Moon, 1998b: 221-233).

Function	Description	Example
Informational	vehicles for conveying new information and contribute to a discourse propositionally	<i>by default, clear ONE'S throat, down tools, face to face</i>
Evaluative	communicate the evaluations of the speaker/ writer, rather than simply furthering the narrative	<i>do the trick, down to earth, drag ONE'S feet, drag ONE's heels, get off to a flying start</i>
Situational	typically found in spoken discourse as they are responses to or occasioned by the extralinguistic context: they may also be illocutionary speech acts	<i>excuse me, go for it, good luck, good morning, I beg your pardon</i>
Modalizing	indicating modality (hyperpropositional), typically epistemic or deontic in nature, although some FEIs may be regarded as conative or volitive	<i>as we know it, at all, at any price, believe you me, full well</i>
Organizational	organize texts by signalling logical connections between propositions, deixis, prefaces, summaries, opinions, etc.	<i>all in all, as yet, be that as it may, by the same token, by the way</i>

Many FEIs simultaneously perform more than one function, but the discussion looks at primary functions only. Thus, while informational FEIs are shown to be the most frequent in Moon's (1998b) corpus, many of them simultaneously perform other functions. Moon's results show that metaphors and proverbs tend to occur more frequently in journalism than in other registers, whereas the frequency of pure idioms in spoken interaction is rather low. The study also addresses some issues of stylistics of phraseology linking the evaluative function of FEIs with metaphorical and simile

content and concluding that “the use of institutionalized metaphors is stylistic, bound up with evaluation, and centred on the interaction” (p. 225).

Another study addressing the stylistics of phraseology is that of Gläser (1998), who discusses the role of formulaic sequences in creating stylistic impact from a genre perspective by analysing different genres for idioms and collocations. Idioms, according to the author, form the centre of the ‘phrasicon’ (Gläser’s term for denoting the phraseological stock of a language), and non-idioms (restricted non-idiomatic collocations) constitute its periphery. Gläser identifies the phraseological units using three different tests: the substitution test, the paraphrase test, and the deletion test, and concludes that the role of these units in style is crucial, since they make the text more colourful, vivid, stylistically marked and memorisable. The author calls for further research in the field of phraseo-stylistics arguing that the importance of phraseological units in style is as yet little understood and worthy of study.

Gries (2008) discusses six parameters to be taken into account when studying phraseology: the nature of the elements involved in a phraseologism, the number of these elements, the frequency (the number of times the element occurs), the distance between the elements, lexical and syntactic flexibility of the phraseologism, semantic unity and semantic non-compositionality of the phraseologism. Of these, the first three might seem more objective and measurable: we can accurately describe any sequence in terms of the nature of the elements that compose it (words, phrases, clauses), its size (number of elements composing it), and the distance between these

elements (whether they are adjacent or separated by other lexical items). The last two are relative and can only be measured against a continuum, as is the case with so many other concepts in phraseology. In addition, the frequency parameter (exactly how often a given sequence occurs), although measurable, is not absolute either: it must necessarily be measured across a corpus of different types of text (the British National Corpus or BNC in this case), i.e., in relation to other linguistic elements occurring in a range of texts, and the type and length of the text will have considerable impact on this parameter. Gries admits that this corpus does not attest all the possible occurrences (which in any case no corpus could do). Nevertheless, due to its considerable size, the BNC (Natcorp.ox.ac.uk) can be regarded as representative of the English language, since it is one of the largest freely available corpora compiled to this date and comprises all the main genres and registers.

This notwithstanding, frequency is a popular criterion for identifying formulaic sequences. Strings that occur frequently become institutionalised and conventionalised, and thus obtain the status of formulaic sequences (Altenberg, 1998). The frequency approach is used in a wide variety of studies, especially those concerned with pedagogical applications of phraseology (Adolphs & Durow, 2004; Biber, 2005; 2009; Biber & Barbieri, 2007; Biber, Conrad & Cortés, 2004; Biber, Conrad, Reppen, Byrd, Helt, Clark, Cortés, Csomay, Urzua, 2004; Bishop, 2004; Chen & Baker, 2010; Conrad & Biber, 2004; Cortés, 2002, 2004, 2006, 2008, 2013; Coxhead, 2008; Crossley & Salsbury, 2011; Dörnyei, Durow, & Zahran, 2004; Durrant, 2009; Ellis, Simpson-Vlach & Maynard, 2008; Eriksson, 2012; Friginal, 2013; Handl,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

2008; Hatami, 2015; Hunston, 2002; Hyland, 2008a, b; Hyland, 2012; Jones & Haywood, 2004; Kennedy, 2008; Lindstromberg, Eyckmans, & Connabeer, 2016; Nattinger, & DeCarrico, 1992; Neely & Cortés, 2009; Nesi & Basturkmen, 2006; Pan, Reppen & Biber, 2016; Paquot, 2013; Qin, 2014; Salazar, 2011, 2014), which will be discussed in the next section.

This section focused on the principles and criteria of definition and classification of formulaic sequences and discussed several fundamental studies in this field. The vast variety of terms, definitions and classifications is a source of misunderstanding and fuzziness in phraseological research. In addition, most of the criteria of classification and study of formulaic sequences are relative and depend on the focus of the study and the data involved. This is especially true of the parameter of frequency which nevertheless is a common criterion in the study of formulae. The present study is also based on the criterion of frequency, as are all studies addressing lexical bundles, the formulaic sequences which are extremely frequent in a given type of text. The next section will provide an overview of research on lexical bundles, their key features and functions, their use in expert and novice, native and non-native academic writing, and their role in the construction of discourse.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

2.3. LEXICAL BUNDLES: THE MOST FREQUENT SEQUENCES IN A LANGUAGE

The previous section discussed the wide range of terms used to describe formulaic sequences, and the criteria used for classifying them, such as semantic opacity, functions, size, grammatical structure, frequency and so on. This section will describe a frequency-based corpus-driven approach used for studying a particular type of formulaic sequences called lexical bundles (Biber et al., 1999). It will address the key properties of lexical bundles, their types, functions, and role in academic discourse.

As discussed in the previous section, before the advent of digitalised corpora the study of formulaic sequences was based on intuition and focussed mainly on semantically opaque and infrequent items which later turned out to be of little value to learners due to their low frequency and exceptional character: are a poor source for studying productive syntactic and grammatical patterns (Gläser, 1998). The access to large digitalised corpora and the emergence of corpus linguistics revealed other types of formulaic sequences: non-salient but productive and valuable for language users. During the past decades, many studies have focused on such multi-word sequences analysing their different aspects, such as frequency, use, functions,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

distribution across registers, etc. This was made possible due to the development of corpus-driven methodologies.

Corpus-driven methodologies use the corpus as a source of patterns which emerge directly, without pre-selecting any linguistic units for study. They are inductive in the sense that the patterns emerge from the corpus while the researcher analyses it. Corpus-based studies, by contrast, first pre-select the elements which are going to be searched in the corpus using linguistic theory as a basis, and then analyse the corpus for these elements (Biber, 2009; Sinclair, 2004). Biber (2009) emphasises that neither methodology is better or in any way superior to the other, and that the vast majority of studies carried out in this field are based on hybrid methodologies combining corpus-based and corpus-driven approaches.

One type of formulaic sequences discovered through a corpus-driven approach are lexical bundles, also termed clusters (Hyland, 2008 a, b) and N-grams (Anthony, 2010), which can be described as one of the few clearly-defined concepts in phraseology.

These multi-word sequences emerged as a result of analyses of massive collections of texts. What was revealed by these analyses is the extremely frequent occurrence of certain multi-word strings in a corpus of texts representing a particular genre or register. They were first identified by Biber et al. (1999) in their ground-breaking *Longman Grammar of Spoken and Written English*.

The definition of lexical bundles is based on the sole criterion of their frequency (typically at least 20 occurrences per million words, often 40 per million words), without considering other characteristics, such as semantic transparency, structural completeness, functions, etc. Biber et al. (1999) describe lexical bundles as extended collocations “that show a statistical tendency to co-occur” (p. 989) in a given register. Corpus analysis software is used to identify three-, four-, and five-word lexical bundles, specifying the minimum frequency at which they should occur. An additional criterion of distribution or dispersion is also applied: a sequence must appear in a range of texts (typically at least in five different texts or at least in 10% of the texts in the corpus) in order to exclude personal idiosyncrasies of one or two authors who tend to use the same sequence multiple times due to their personal preferences. These personal preference patterns are called ‘phraseological teddy bears’ after Hasselgren, ’s (1994) famous label, “lexical teddy bears”, for lexical preferences of advanced EFL students. The application of conservative frequency cut-off points and distribution thresholds helps to eliminate all the subjective, intuitive and biased attitudes that may interfere with objective study.

The fact that lexical bundles emerge from a corpus on the basis of their high frequency as the sole criterion for identifying them does not imply that they are random co-occurrences. Although some of them (especially the shorter three-word bundles) are sequences of function words only, such as *and of the*, most lexical bundles have been shown to be important building blocks in discourse (Biber, 2006, 2009; Cortés, 2004, 2006, 2008; Hyland, 2008a, b; Salazar, 2011, 2014).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Biber et al. (1999) contrast the use of lexical bundles in conversation and academic writing. The authors draw a distinction between different types of multi-word expressions that function as structural or semantic units: they divide formulae into idiomatic expressions (invariable expressions whose meaning cannot be retrieved from the meanings of their components) and collocations which are described as “statistical associations rather than relatively fixed expressions” (p. 988). Lexical bundles belong in the category of extended collocations and are shown to be much more common in conversation than in academic writing, both in terms of types (each one of the different multi-word strings identified as a lexical bundle) and tokens (the number of times it occurs in a corpus). The difference between idiomatic expressions and lexical bundles is obvious: the meaning of a lexical bundle can be easily derived from the individual meanings of its constituents, as opposed to idioms whose meaning is opaque and cannot be retrieved from the meanings of the individual words that compose it. In addition, lexical bundles show a much higher frequency of occurrence than idioms (less than 5 per million words for idioms versus over 20 per million words for lexical bundles). Further, lexical bundles cannot be substituted by a single word to express their meaning, as is the case with most idioms. Another key feature of lexical bundles is their non-salience: these extremely frequent multi-word sequences are notoriously inconspicuous and often remain unnoticed by readers and listeners.

Biber et al. (1999) study 3-, 4-, and 5-word lexical bundles and show that shorter lexical bundles are more frequent: thus, there are about 10 times as many 3-word

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

bundles as there are 4-word lexical bundles, and 10 times as many 4-word bundles as 5-word bundles. Many 3- and 4-word lexical bundles are parts of longer 4- and 5-word bundles respectively. Lexical bundles both in conversation and in academic prose tend to occur in series (two or more lexical bundles in a string) and frequently bridge two clauses, for example, *in the case of, the base of the* (Biber & Barbieri, 2007).

The following key features of lexical bundles have been identified in different studies: they are extremely common by definition (Biber et al., 1999), they are non-idiomatic (Ädel & Erman, 2012, Biber & Barbieri, 2007), they are non-salient by nature (Biber et al., 1999), they are grammatically incomplete and typically bridge two elements in a clause (Biber et al., 1999; Biber & Barbieri, 2007), they perform systematic discourse functions and act as building blocks for constructing discourse (Biber et al., 1999; Biber, 2009; Cortés, 2004, 2006, 2008; Hyland, 2008a, b; Salazar, 2011, 2014). Further, lexical bundles are field- and genre-specific: different sets of lexical bundles have been identified in expert writing of different genres and academic disciplines (Biber & Barbieri, 2007; Conrad & Biber, 2003, 2004; Coxhead & Byrd, 2007; Durrant, 2009; Durrant & Schmitt, 2009; Hyland, 2008a, b; Oakey, 2002; Peacock, 2012).

As mentioned above, the most outstanding property of lexical bundles is their frequency. Although Biber et al. (1999) established the cut-off point at 10 times per million words, this practice is currently not in use, and the cut-off point is normally

much more conservative, i.e., the minimum frequency must be much higher than 10 occurrences per one million words (Cortés, 2013).

Although extremely frequent, most lexical bundles do not represent structurally complete units. Only 15% of the bundles in conversation have been found to be complete structural units, while in academic prose their proportion is only 5% (Biber et al., 1999). They are structurally incomplete and typically bridge two elements of an utterance, e. g., *do you want me to, I don't know if you* (Biber, Conrad & Cortés, 2003, 2004; Biber, 2006).

In addition to fixed continuous sequences, another broad category of multi-word formulaic sequences related to lexical bundles is described in a radical corpus-driven study carried out by Biber (2009). These sequences are not continuous fixed expressions, but show very high frequency of occurrence similar to that of lexical bundles. Many of them can be represented as discontinuous frameworks consisting of fixed and variable slots. Thus, the fixed slots in a framework remain the same throughout all the occurrences of the framework, while the variable slot is filled with different function or content words. For example, in the framework *the * of the*, the variable slot (the asterisk in the example) can be filled with *end, basis, case, cause, presence, absence*, etc., while the rest of the words remain unchanged. Biber (2009) develops pattern types to describe each framework with both fixed and variable slots. For instance, the sequence *on the other hand* is described through the pattern 1234

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

where all the slots are relatively fixed, and the sequence *in the case of* corresponds to the pattern 12*4 where the slots 1, 2, and 4 are fixed and the slot 3 is variable.

Biber's study adopts a register perspective to analyse lexical bundles, both fixed and variable, in two different types of discourse: conversation and academic writing. His results show that there are several dramatic differences between the two registers. Thus, conversational multi-word sequences typically constitute fixed sentences, whereas academic writing relies more on formulaic frames. Conversational bundles are composed of sequences in which both function and content words can be variable, while sequences in academic writing tend to present frameworks of invariable function words with slots for variable content words. Supporting previous research (Biber & Barbieri, 2007; Cortés, 2004, 2006, 2008), a tendency emerges which proves academic writing to be phrasal in contrast to conversation which is mainly clausal. Biber's study, in contrast to many other corpus-driven studies, adopts frequency as a basic criterion for defining and analysing lexical bundles and lexicogrammatical patterns. He contends that very few studies take into account the frequency of multi-word sequences, except for the distinction between 0 and 1 and the distinction between 1 and 2. The distinction between 0 and 1 is important in such studies as it indicates occurrence versus non-occurrence, and the distinction between 1 and 2 is a criterion that indicates repetition. Otherwise, however, frequency does not seem to be of concern to researchers who perform corpus-driven studies, though it is, as Biber argues, a crucial factor for identifying formulaic language.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Although such formulaic patterns are likely to be productive and valuable for language learners and users, their fixed counterparts, the continuous lexical bundles are a much more popular object of study in phraseology. A large amount of research on lexical bundles uses the structural and functional taxonomy proposed by Biber, Conrad and Cortés (2003), discussed below.

Biber, Conrad and Cortés (2003, 2004) analyse the grammatical structure of lexical bundles in different registers and classify them into the following broad categories: noun phrases, verb phrases and prepositional phrases, each divided further into subcategories. Table 1.5 shows these categories and an example from each subcategory, adapted from Biber, Conrad and Cortés, 2004, and Salazar, 2011.

Table 1.5. Structural taxonomy of lexical bundles (adapted from Biber, Conrad & Cortés, 2004, and Salazar, 2011)

TYPE	EXAMPLE
Noun phrases	
noun phrase with of-phrase fragment	<i>the end of the</i>
noun phrase with other post-modifier fragment	<i>the way in which</i>
other noun phrase	<i>the ability to</i>
Prepositional phrases	
prepositional phrase with embedded of-phrase fragment	<i>as a result of</i>
other prepositional phrase fragment	<i>as in the case</i>
Verb structures	
anticipatory it + verb phrase/adjective phrase	<i>it is possible to</i>
passive verb + prepositional phrase fragment	<i>is shown in figure</i>
copula <i>be</i> + noun phrase/adjective phrase	<i>is one of the</i>
(verb phrase +) <i>that</i> -clause fragment	<i>should be noted that</i>
(verb/adjective +) <i>to</i> -clause fragment	<i>are likely to be</i>
adverbial clause fragment	<i>as we have seen</i>
pronoun/noun phrase + <i>be</i>	<i>this is not the</i>
Other expressions	<i>as well as in</i>

Further research on lexical bundles (Ädel & Erman, 2012; Biber, 2006; Biber & Barbieri, 2007; Chen & Baker, 2010; Cortés, 2006; 2013; Salazar, 2011; 2014, etc.) builds mostly on this classification to explore the characteristic features of these multi-word sequences and their role in discourse, as well as on the functional classification proposed in the same study (Biber, Conrad & Cortés, 2003).

Conrad and Biber (2004) develop their work on the previous studies of lexical bundles (Biber et al., 1999; Biber, Conrad & Cortés, 2003, 2004) to contrast lexical bundles in conversation and academic prose. They suggest six main characteristics or criteria according to which formulaic sequences can be defined or identified: fixedness, idiomaticity, frequency, length, completeness in syntax, semantics, or pragmatics, and intuitive recognition by native speakers. They further discard idiomaticity arguing that idiomatic expressions are extremely rare in natural speech and writing, and focus on more frequent and more relevant building blocks of discourse which perform important pragmatic functions, even though they are not structurally or semantically complete. The authors, however, point out that none of these criteria is absolute and may vary depending on the purposes of the study and the researcher.

Although the most salient characteristic feature of lexical bundles is their frequency, they have been shown to perform important discursive functions, acting as prefabricated building blocks of discourse (Biber, 2006; Biber, 2009). Wray (2002) also points out the discourse marking functions of formulaic sequences (although

Wray’s “formulaic sequences” are not equivalent to lexical bundles: as has been shown in Section 1.2, lexical bundles can be regarded as a variety or subtype of formulaic sequences as described by Wray) and their apparent property of signalling the relationship between two elements of discourse, as well as the speaker’s / writer’s attitude towards the co-text. Despite their non-salient nature and structural incompleteness, lexical bundles:

are not an accidental by-product of corpus frequency analysis. Rather, these word sequences turn out to be consistently functional, indicating that high frequency is a reflection of pre-fabricated or formulaic status. Although they go largely unnoticed by speakers, hearers, and analysts, lexical bundles are pervasive in spoken and written texts, where they serve basic discourse functions related to the expression of stance, discourse organization, and referential framing (Biber & Barbieri, 2007, p. 265).

Thus, Biber, Conrad and Cortés (2003, 2004) first proposed a classification of lexical bundles into referential, discourse-organising and stance bundles, with further subdivisions. Referential bundles are those which refer to the text itself, discourse organisers help to guide the reader through different sections of the text, while stance bundles show the stance or attitude of the author towards the idea expressed in the text. Conrad and Biber (2004) introduce a new category in the classification of lexical bundles suggested earlier: in addition to the three well-known categories - stance expressions, discourse organisers, and referential expressions - they also include a category of "special conversational functions" (p. 66), sequences like *thank you very*

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

much. These are relatively few in type (the authors quote only three of them) and fall outside the scope of this study as they occur exclusively in conversational registers.

Figure 1.3 illustrates this classification and its subcategories.

Figure 1.3. The functional taxonomy of lexical bundles (adapted from Conrad & Biber, 2004: 65-66).

CATEGORY	TYPE	SUBTYPE	EXAMPLES
STANCE EXPRESSIONS	Epistemic stance		
		Personal	<i>you know what I</i>
		Impersonal	<i>*the fact that the</i>
	Attitudinal / Modality Stance		
		<i>Desire</i>	<i>you want to go</i>
		<i>Obligation / Directive</i>	
		Personal	<i>going to have to</i>
		Impersonal	<i>*it is necessary to</i>
		<i>Intention / Prediction</i>	
		Personal	<i>I was going to</i>
	Impersonal	<i>*going to have a</i>	
	Impersonal	<i>*it is possible to</i>	
DISCOURSE ORGANISERS	Topic introduction / Focus		
			<i>have a look at</i>
	Topic elaboration / Clarification		
			<i>*on the other hand</i>
REFERENTIAL EXPRESSIONS	Identification / Focus		
			<i>*one of the most</i>
	Imprecision		
			<i>or something like that</i>
	Specification of attributes		
		<i>Quantity</i>	<i>*per cent of the</i>
		<i>Tangible framing attributes</i>	<i>in the form of</i>
		<i>Intangible framing attributes</i>	<i>*on the basis of</i>
Time / Place / Text Reference			
	<i>Time reference</i>	<i>*at the same time</i>	
	<i>Multi-functional reference</i>	<i>**the end of the</i>	

SPECIAL CONVERSATIONAL FUNCTIONS	Politeness	<i>thank you very much</i>
	Simple inquiry	<i>what are you doing</i>
	Reporting	<i>I said to him</i>

The bundles marked by a single asterisk (*) are common in academic prose.
The bundle marked by a double asterisk (**) is common in both academic prose and conversation.

The functional classification, however, raises certain issues, given the fact that there are differing and divergent views on their function and the role ascribed to each unit in discourse. In the words of Oakey (2002), “[a]pplying functional labels to linguistic units is a highly subjective activity” (p. 115).

Later Hyland (2008a) modified this taxonomy to propose the following categories of bundles: research-oriented, text-oriented, and participant-oriented, a detailed description of which can be found in Table 1.6.

Table 1.6. Functional classification of lexical bundles (adapted from Hyland, 2008a: 10-11).

Research-oriented	help writers to structure their activities and experiences of the real world		
	Location	indicating time/place	<i>at the beginning of, at the same time, in the present study</i>
	Procedure	indicating 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, the surface of the</i>
	Topic	related to the field of research	<i>in the Hong Kong, the currency board system</i>
Text-oriented	concerned with the organisation of the text and its meaning as a message or argument		
	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 figure</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	focused on the writer or reader of the text		
	Stance features	convey the writer's attitudes and evaluations	<i>are likely to be, may be due to, it is possible that</i>
	Engagement features	address readers directly	<i>it should be noted that, as can be seen</i>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Although very extended and well-received by linguists, this taxonomy has also elicited criticism, among them this lengthy critique from Ädel and Erman, (2012):

“We have several reservations about this classification, despite the fact that it is presented as unproblematic in Chen and Baker (2010). Here, the classification is introduced as fully established, even though it was called “preliminary” when first presented in Biber et al. (2004, p. 383). The main problem is that no clear criteria are given for how to decide which (sub)category a given bundle should belong to. While some of the subcategories are somewhat well-defined by previous research or are intuitively clear (e.g., topic introduction, quantifying), others are vague (e.g., identification/focusing, framing). In fact, this vagueness has led to several inconsistencies in previous research. For example, Focusing is labelled Discourse organising in Chen and Baker, but Referential in Biber et al. (2004) and Simpson-Vlach and Ellis (2010). Framing is labelled Referential in Chen and Baker and Biber et al. (2004), but Discourse organising in Cortés (2004). An additional problem is the multifunctionality of many lexical bundles. When classifying a given type, as emphasized in Biber et al. (2004), it is therefore necessary to consider the extended context to determine what the predominant function is.” (Ädel and Erman, 2015, p. 89).

Pan, Reppen and Biber (2016) adopt Hyland’s classification for their research, except for the last category of participant-oriented bundles, which they prefer to leave with its original label, ‘stance-oriented’. The reason for this, the authors argue, is that the

term ‘participant-oriented’ is somewhat inappropriate for academic discourse, whereas ‘stance’ covers all the meanings they wish to ascribe to this category.

Table 1.7 below shows the two taxonomies and the resulting hybrid functional taxonomy based on that of Biber et al. (2004) and Hyland (2008a), with examples.

Table 1.7. Functional taxonomies of lexical bundles according to Biber, Conrad & Cortés (2004), Hyland (2008a), and Pan, Reppen & Biber (2016).

Biber, Conrad & Cortés (2004)	Referential	Discourse organisers	Stance bundles
Hyland (2008a)	Research-oriented	Text-oriented	Participant/oriented
Pan, Reppen & Biber (2016)	Research-oriented	Text-oriented	Stance-oriented
Example	<i>at the beginning of, the use of the, the structure of the</i>	<i>as a result of, in the present study, with respect to the</i>	<i>as can be seen, may be due to the</i>

Simpson-Vlach and Ellis (2010) made several extensions and modifications to the functional taxonomy suggested by Biber, Conrad and Cortés (2004) in order to adapt it to their results. Table 1.8 describes the taxonomy proposed in Simpson-Vlach and Ellis (2010) with some examples. As can be seen from Table 1.8, Simpson-Vlach and Ellis changed several subcategory labels, for instance, the ‘Imprecision’ subcategory has been eliminated, and a similar subcategory of ‘Vagueness markers’ has been introduced and ‘Topic Elaboration’ has been further split into ‘Causal’ and ‘Non-Causal’ subtypes.

Table 1.8. Classification of lexical bundles (adapted from Simpson-Vlach & Ellis, 2010) based on Biber, Conrad & Cortés (2004).

Category	Subcategory	Example
REFERENTIAL EXPRESSIONS	Specification of attributes	
	a) Intangible framing attributes	<i>a) based on the, the notion of</i>
	b) Tangible framing attributes	<i>b) the level of</i>
	c) Quantity specification	<i>c) there are three</i>
	Identification and focus	<i>such as the, an example of</i>
	Contrast and comparison	<i>as opposed to</i>
STANCE EXPRESSIONS	Deictics and locatives	<i>at this point</i>
	Vagueness markers	<i>and so on</i>
	Hedges	<i>to some extent</i>
	Epistemic stance	<i>assume that the</i>
	Obligations and directive	<i>need not be</i>
	Ability and possibility	<i>can be used to</i>
DISCOURSE ORGANISERS	Evaluation	<i>it is obvious that</i>
	Intention, volition, prediction	<i>we do not</i>
	Metadiscourse & textual reference	<i>in the next section</i>
	Topic introduction and focus	<i>what are the</i>
	Topic elaboration	
a) non-causal	<i>a) in more detail</i>	
b) cause and effect	<i>b) for the purposes of</i>	
Discourse markers	<i>as well as</i>	

Salazar (2011) analyses each structural category and assigns corresponding meanings or purposes to every one of them. A first glance at the list of lexical bundles extracted by Salazar and divided according to the structural categories suggests a certain degree of correlation between form and function:

Through the structural analysis of the verbal target bundles, it was possible to establish certain usage patterns that demonstrate the importance of both active and passive, personal and impersonal expressions in scientific writing. Passive, impersonal constructions are employed in the objective discussion of experimental methods and justification of claims, so as to build a sound and universally acceptable foundation for the author's subsequent assertions. Active, personal structures, on the other hand, are used by scientists to explain their aims, findings, accomplishments and conclusions, as a way to underscore their original contribution to the field of research. The judicious use of personal and impersonal expressions reflects the dual role of the scientist as conductor of research and claim maker, and plays an essential role in the construction of an effective research article. These choices of voice and tense constitute a subtle yet important rhetorical function of which non-native or novice writers should be made aware (Salazar, 2011, p. 90).

The author notes that many structural categories are used to express specific meanings. It is rather obvious, for instance, that the anticipatory *it* structure is typically used to denote possibility, likelihood, importance, or certainty and therefore performs an epistemic (stance) or text-oriented function. More correlations between form and function can be found upon careful examination. Table 1.9 below is an attempt to outline the correlations between the structure of a lexical bundle and its function, although it should be noted that this table, adapted from Salazar's (2011) findings, is merely an example of randomly chosen lexical bundles and does not in any way represent a systematic attempt to connect form and function. Thorough

investigation is required to shed more light on the possible relationship of the structure of lexical bundles and their functions.

Table 1.9. The structure and function of lexical bundles (based on Salazar, 2014).

Structure	Meaning	Function	Example
Noun structures			
	Quality	Referential (intangible framing attributes)	<i>the nature of</i>
	Degree	Research-oriented (degree)	<i>the extent of</i>
	Existence	Research-oriented (description)	<i>the presence of</i>
	Events	Research-oriented (procedure)	<i>the beginning of</i>
	Actions	Research-oriented (procedure)	<i>the addition of</i>
	Measurements	Research-oriented (quantification)	<i>an equal volume</i>
	Quantities	Research-oriented (quantification)	<i>a small number</i>
	Proportions	Research-oriented (quantification)	<i>the percentage of</i>
	Location	Research-oriented (location)	<i>the region of</i>
	Groupings	Research-oriented (grouping)	<i>a wide range of</i>
	Membership	Research-oriented (grouping)	<i>a component of</i>
Verb structures			
Passive + prepositional phrase fragment	Locative or logical relations (present tense verb)	Test-oriented (citation) Text-oriented (framing)	<i>are described in</i> <i>is based on</i>
	Activity, procedure (past tense verbs)	Research-oriented (procedure)	<i>was added to</i> <i>was removed by</i>
Other passive fragment	Referring to previous research	Text-oriented (inferential)	<i>has been demonstrated</i>
Verb phrase with personal pronoun <i>we</i>	Objectives	Text-oriented (objective)	<i>we asked whether</i>
	Observations	Participant-oriented (stance)	<i>we found that</i>
	Achievements		<i>we have identified</i>

	Conclusions	Participant-oriented (stance) Participant-oriented (stance)	<i>we conclude that</i>
Other verbal fragment		Text-oriented (causative) Participant-oriented (engagement)	<i>does not affect</i> <i>see materials and methods</i>
Prepositional-phrase fragments			
Prepositional phrase + of	Abstract, logical relationships between propositional elements Method Process Amount Measurement	Text-oriented (framing) Text-oriented (framing) Text-oriented (causative) Research-oriented (procedure) Research-oriented (procedure) Research-oriented (quantification) Research-oriented (procedure)	<i>on the basis of</i> <i>in the context of</i> <i>as a consequence of</i> <i>by the method of</i> <i>in the formation of</i> <i>in a number of</i> <i>at a density of</i>
Other prepositional phrase (fragment)	Referring to the study Place, extremity, orientation Figurative meanings	Text-oriented (structure) Research-oriented (location) Text-oriented (additive) Text-oriented (framing)	<i>in the present study</i> <i>in the region, on the left</i> <i>on the other hand</i> <i>in concert with</i>
Other structures			
Verb/adjective + to- clause fragment	Previous knowledge Ability Likelihood Methodological aims	Text-oriented (inferential) Research-oriented (description) Text-oriented (inferential) Text-oriented (objective)	<i>were found to be</i> <i>is able to</i> <i>are likely to be</i> <i>to determine whether</i>
Verb/noun + that- clause fragment	Present facts or findings Inferences	Text-oriented (inferential) Text-oriented (inferential)	<i>the hypothesis that</i> <i>these results suggest that</i>

Adverbial clause fragment	Directing the reader to parts of text or other texts	Text-oriented (structuring)	<i>as described previously</i>
	Assertion	Text-oriented (inferential)	<i>as judged by</i>
	Comparison	Text-oriented (comparative)	<i>as compared with</i>
Copula <i>be</i> + adjective phrase	Causative relationships	Text-oriented (causative)	<i>may be due to</i>
	Comparative relationships	Text-oriented (comparative)	<i>are similar to</i>
	Assessment	Participant-oriented (stance)	<i>is difficult to measure</i>
Anticipatory <i>it</i> + verb/adjectival phrase	Possibility	Text-oriented (inferential)	<i>it is possible that</i>
	Likelihood	Text-oriented (inferential)	<i>it is likely that</i>
	Importance	Participant-oriented (stance or engagement)	<i>it is important to</i>
	Certainty	Text-oriented (inferential)	<i>it is clear that</i>
Other adjectival phrase	Comparative relations	Text-oriented (comparative)	<i>significantly different from</i>
		Text-oriented (inferential)	<i>closely related to</i>
Other expressions		Text-oriented (comparative)	<i>this is consistent with</i>
		Text-oriented (objective)	<i>in order to</i>
		Text-oriented (additive)	<i>as well as</i>

Biber, Conrad and Cortés (2004) also address the correlation between form and function in the four university genres in their study. They state that the majority of stance bundles represent dependent clause fragments, referential bundles tend to be composed of noun phrase or prepositional phrase fragments, prediction / intention

stance bundles are all verb phrase fragments, while discourse organisers are the only category to make use of all the three structural patterns.

A number of important caveats should be mentioned here. The first is the multifunctionality of lexical bundles, a phenomenon widely discussed in the literature (Ädel & Erman, 2012; Biber, 2006, Biber & Barbieri, 2007; Cortés, 2004, 2006, 2008; Salazar, 2011; 2014). The most frequent lexical bundles are also the ones which perform different functions depending on the context (and on the discipline and genre), and it is not always easy to assign a single function to them (Oakey, 2002). Thus, Oakey identifies the following functions for the lexical bundle *it is/has been (often) asserted / believed / noted that X* in academic discourse: topic priming (introducing the topic), support (the most common function), Straw Man (introducing an argument which the writer will evaluate negatively), and self-reference. The solution provided by most researchers is to apply the most frequent function and consider it as primary, thus assigning a secondary role to the other function or functions. However, this approach does not provide a valid solution for correlating the structure and function. Oakey (2002) notes that “putting linguistic items in tidy functional categories can obscure the fact that they can have multiple functions” (p. 126). Further research is needed to determine the extent to which the structure of a lexical bundle correlates with its function. The second caveat is that certain categories of lexical bundles are not clearly determined in terms of structure. Thus, in verb structures, alongside the categories of ‘Verb + *that*-clause’ or ‘passive verb phrase’, there is a category called ‘other verb structures’ (Salazar, 2014) which comprises all

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

the bundles that do not fit into the other categories. This, strictly speaking, is not a structural category as there is no structural formula to describe it, in contrast to the other perfectly structured categories. These ‘other structure’ categories, obviously, are unlikely to have strong correlations in terms of function. In addition, lexical bundles are ‘open class’ linguistic items: there is potentially a very large or even infinite number of them, which means that there may be any number of exceptions to the most elaborate systematic correlation patterns (Coxhead & Byrd, 2007). Finally, the fact that lexical bundles are discipline-sensitive may influence the form-function relationship in different disciplines (Biber, 2006; Hyland, 2008b). That notwithstanding, several strong correlations can be identified at a glance, for example, the passive constructions being used for describing results or the copula *be* + adjectival phrase expressing causative or comparative relationships, or author’s evaluative assessment, as in *this may be due to, are similar to, or it is difficult to measure*.

Salazar (2014: 54) observes “certain semantic and structural relationships among the lexical bundles” after grouping them by keyword instead of organising by frequency.

Biber, Conrad and Cortés (2004) note that:

“we see a complex interaction between structural form, discourse function, and the typical purposes and situational characteristics of registers”. (p. 398).

From the discussion above it can be seen that the structure of formulaic sequences is linked to a certain extent to their meaning and function in discourse. However, there

can be no full correlation between structure and function as the same lexical bundles often perform different functions across different genres and disciplines (Biber, 2006, Hyland, 2008b). The remainder of this section will focus on the studies of lexical bundles in academic discourse, exploring their functions and distribution in different academic genres.

Despite their relatively low frequency in academic discourse, lexical bundles have been studied extensively due to the important discourse functions they perform in this register. The appropriate use of lexical bundles has been shown to be an indicator of the proficiency and expertise of the speaker or writer (Biber, 2009; Biber & Barbieri, 2007; Chen & Baker, 2010; Cortés, 2002, 2004; Erman, Lewis & Fant, 2013; Pan, Reppen & Biber, 2016). As was shown above, lexical bundles are neither idiomatic nor complete semantic or structural units, but it cannot be said that they are mere sequences of accidentally neighbouring words. On the contrary, they are considered to be important building blocks of discourse which fulfil crucial functions (Conrad & Biber, 2004, 2005; Biber et al., 1999; Biber & Gray, 2010; Chen & Baker, 2010) and have been reported to “[help] to shape meanings in specific contexts and contributing to our sense of coherence in a text” (Hyland, 2008a, p. 4).

Lexical bundles in academic discourse have been studied from a variety of angles and perspectives: comparing expert and novice writing (Cortés, 2002; Chen & Baker, 2010; Hyland, 2008a; 2008b), native and non-native writers (Adolphs & Durow, 2004; Rica, 2010; Staples et al., 2013), different genres and registers (Hyland, 2008b;

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Oakey, 2002), different disciplines (Ädel & Erman, 2012; Sánchez, 2014), etc. Many of these studies have simultaneously addressed more than one parameter. Chen and Baker, for instance, conduct a comparative study of lexical bundles in L1 and L2, expert and novice academic writing in a range of disciplines. Similarly, Hyland's (2008b) study analyses lexical bundles in PhD and Master's theses and research articles in four different fields.

A considerable portion of studies on lexical bundles in academic discourse has addressed four-word bundles (Ädel & Erman, 2012; Chen & Baker, 2010; Cortés, 2004, 2006, 2008, 2013; Biber, 2005; Farvardin, Afghari & Koosha, 2012; Hyland, 2008a). The reason for this choice is that these strings are more frequent than larger, 5- or 6-word bundles and hence yield more objective results. They are also more complete than three-word bundles, most of which are subsumed in larger four-word sequences: for example, *as a result of* contains the shorter lexical bundle *as a result* (Cortés, 2013). However, in very specific contexts much longer bundles can be identified. Thus, Cortés (2013) analyses research article (RA) introductions for lexical bundles of different length and finds clusters of up to nine words which, although relatively rare in the text, still meet the cut-off point of 20 words per million and thus can safely be classified as lexical bundles. These clusters are shown to perform rhetorical functions of moves and steps according to Swales (1990) CRAS model. Longer lexical bundles, Cortés argues, are more informative due to the larger number of words they contain; besides, due to their low frequency, they are more move- or step-bound, i.e., they occur only in one move or step.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Biber and Barbieri (2007) conducted a study of functional and structural characteristics of lexical bundles in a wide range of university genres: classroom teaching, classroom management, office hours, study groups, service encounters, course management, institutional writing, textbooks, and other academic texts. Their findings confirm Biber et al.'s (1999) conclusion regarding the relatively infrequent occurrence of lexical bundles in academic prose, although they still meet the very conservative cut-off point of 40 per million words, and many of their items occur as frequently as 200 times per million words and more. Another finding is that classroom management and service encounters are the two genres that use the widest variety of lexical bundles, a much wider variety than even classroom teaching which had also been shown to contain a large number of lexical bundles (Biber et al., 1999). The authors point out that “the extent to which a speaker or writer relies on lexical bundles is strongly influenced by their communicative purposes, in addition to general spoken / written differences” (Biber & Barbieri, 2007, p. 273), thus conditioning the use and variety of lexical bundles by the mode (spoken or written) and the communicative purpose. Further research has also shown that in addition to mode, genre, and communicative purposes, the use and frequency of lexical bundles is also conditioned by the discipline and field: different sets of lexical bundles can be observed in different disciplines, and their density also varies across fields.

Biber and Barbieri (2007) show that stance bundles are extremely frequent in spoken university genres, obligation / directive bundles being the most common in this category, while written non-academic university genres (course management,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

institutional writing) are the ones with the most lexical bundles. In contrast, written university genres, namely academic prose and institutional writing, make an extensive use of referential bundles, even though lexical bundles are relatively rare in these genres. Biber (2006) finds that classroom teaching is the genre that uses the most lexical bundles, and the most common bundles in this genre are discourse organising and referential bundles. According to Biber, Conrad and Cortés, (2004), classroom teaching employs twice as many lexical bundles as conversation, and about three times more than textbooks. This is due to the fact that classroom teaching relies both on conversational and academic prose features, resulting in elevated levels of formulaicity. In this respect classroom teaching is similar to conversation, which also employs a large number of lexical bundles (Biber, 2006). Academic prose is shown to use fewer lexical bundles than other registers, but the most frequent bundles in academic prose are those belonging to the referential type. Thus, lexical bundles are to some extent present in all registers and all genres, although certain genres and registers are more abundant in lexical bundles than others (Biber, 2006).

Conrad and Biber (2004) look at lexical bundles of different lengths and analyse three-, four, and five-word lexical bundles in a corpus of British English conversations, and British and American English academic prose. They show that there is almost no overlapping in the functions of lexical bundles in the two registers: lexical bundles which are common both in conversation and academic prose fulfil entirely different functions in these two registers. The only two common bundles which also perform the same function are *the end of the* and *at the end of*. It is not

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

unreasonable to suppose that these two bundles are parts of the larger lexical bundle *at the end of the*, thus reducing the number of shared functions to only one instance. The variability of lexical bundles across genres and registers has also received much attention in recent research. For example, Biber, 2006, discovered that the genre of classroom teaching employs four times as many lexical bundles as occur in textbooks, presumably due to the nature of classroom discourse which combines conversational features with those typical of written discourse (*it has to do with* versus *I don't know if you*). This finding was confirmed in a subsequent study (Biber, 2009). Moreover, writers and readers rely on lexical bundles to identify certain characteristics of a text. The 'natural' use of lexical bundles signals the writer's proficiency and familiarity with the conventions of a given genre, his or her belongingness to a given speech community. If, on the contrary, a writer (or speaker) does not meet the discourse conventions of a given community, they can be labelled as non-expert. Hyland (2008a) shows that lexical bundles play a significant role in differentiating between written discourse in one discipline from another: the use of lexical bundles in soft sciences is considerably different from hard sciences.

Hyland's (2008a) study focusses on 4-word lexical bundles in a 3.5, million-word corpus of academic texts (published research articles, MA and PhD dissertations) from four different disciplines: Applied Linguistics, Business Studies, Electrical Engineering, and Biology. The MA and PhD dissertations in his corpus are all produced by L2 writers whose first language is Cantonese thus making the L2 corpus homogeneous and representative. Hyland's findings show that there is considerable

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

variation in terms of the use of lexical bundles across disciplines, with electrical engineering being the discipline which makes the most use of them, and biology the least. The causes of this disparity are not clear, but the author relates them to the genre conventions of each discipline and their ways of presenting and discussing problems: while Biology and Electrical Engineering are similar in their use of visuals, they also present striking differences:

[Biology] is an altogether more discursive and descriptive discipline, with a less active and applied agenda. It is also a discipline more concerned with naming and coding than the other fields, with a more specialised readership, speaking to a relatively narrow group of scientist end users with specific interests in findings which inform their own research.” (Hyland, 2008a, p.10)

The structure of prevailing lexical bundles also varies across disciplines, with social sciences relying more heavily on bundles beginning with a prepositional phrase, such as *on the basis of* or *in terms of the*, and natural sciences making more use of passive + prepositional phrase constructions, for example, *is shown in Fig. 4*. These passive constructions, according to Hyland, alongside the anticipatory *it* pattern, also rather frequent in natural sciences texts, reflect the tendency to disguise authorial interpretations and make the text as objective as possible. Hard sciences also employ more research-oriented bundles which serve the same function:

But the significantly greater use of research-oriented bundles in the hard knowledge fields also expresses something of a scientific ideology which emphasises the empirical over the interpretive,

minimising the presence of researchers and contributing to the “strong” claims of the sciences. Highlighting research rather than its presentation places greater burden on research practices and the methods, procedures and equipment used, and this allows scientists to emphasise demonstrable generalisations rather than interpreting individuals. New knowledge, then, is accepted on the basis of empirical demonstration and experimental results designed to test hypotheses related to gaps in knowledge. The rhetorical conventions of the field, including the preferred patterns of 4-word bundles, help contribute to this epistemological framework.” (Hyland, 2008a p.12)

Social sciences, in contrast, tend to use more interpretative discursive elements, such as text-oriented bundles, “producing discourses which often recast knowledge as sympathetic understanding, promoting tolerance in readers through an ethical rather than cognitive progression” (Hyland, 2008a, p. 16).

The author concludes that “4-word bundles should be regarded as a basic linguistic construct and that their distributions can help characterise disciplinary discourses” (p. 20).

The role of lexical bundles or clusters in defining different genres within a single register is further explored in Hyland (2008b) through a study of clusters in research articles, PhD and master's dissertations. The findings of this study show that Master's students use a greater variety and more instances of clusters than the PhD students and published authors. The most frequent strings used by the three groups overlap only partially, with published authors using a very different set of bundles as

compared to those used by the two novice writer groups. The research articles are the genre with the lowest number and smallest variety of clusters of all three. This is a surprising finding, as many researchers (Biber & Barbieri; Cortés, 2004; 2006; 2008; 2013; Salazar, 2011; 2014) have argued that expert writers tend to use more lexical bundles both in type and in token, and have underscored the phrasal character of expert writing as opposed to novice writing which is typically clausal. It can be partly explained by the fact that some of the bundles in Master's student writing in Hyland's (2008b) corpus were repeated as many as 70 times in a single text, which would appear to be an instance of heavy overuse of a certain cluster by a single author. This overuse of several bundles might have boosted their overall number in the corpus and thus influenced the results which could otherwise have been different. In a corpus of 825.000 words, as the one Hyland's study relies on, 70 clusters in one single text may be statistically significant.

Hyland's (2008b) study shows that different genres employ different sets of lexical bundles and thus can be characterised through these clusters. For instance, research-oriented bundles are particularly abundant in Master's student writing, which, as the author suggests, may be due to the necessity of the students to demonstrate their ability to carry out research using the appropriate methods and supporting it with previous studies. Hyland argues that this difference in bundle usage does not necessarily testify to a deficiency in students' writing skills. It may be due to the requirements and constraints of the genre. Master's theses, in contrast to PhD dissertations and research articles, are a pedagogic genre with an entirely different

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

set of communicative purposes. Evidence from this study supports the idea that each genre requires its own way of dealing with lexical bundles based on formal constraints, style and communicative purpose. As Hyland puts it:

“while clusters are simply statistical regularities of language use for the analyst, they actually reflect a lived reality for users” (p. 44).

Biber's (2009) study also points to the idea that lexical bundles in speech and writing emerge to fulfil the most important communicative purposes, and therefore are critical in language learning and text production. It is thus important to explore the structural and functional properties of these lexico-grammatical units in each genre in order to construct the discourse according to the expectations of the discourse community in question.

This section covered the different classifications of lexical bundles and discussed the studies addressing their role in academic discourse. The next section will dwell on the studies focussing on the use of lexical bundles in novice academic writing and their acquisition by EFL, English for Academic Purposes (EAP) and L2 learners, and will attempt to explain the role of formulaic language in the integration of novice non-native researchers into the academic discourse community.

2.4. LEXICAL BUNDLES IN THE EAP CLASSROOM

Section 1.3 provided an account of several important characteristic features of lexical bundles, such as their crucial role in building different types of discourse, their discipline-specific and genre-specific character, and the important communicative and pragmatic functions that they perform. This section will discuss the studies addressing the pedagogical dimension of phraseology for L1 and L2 learners, and attempt to explain its importance in second language acquisition, foreign language teaching, and EAP.

The implications of phraseological research for language acquisition were clear already at the early stages of the development of this area of research. The Introduction to Cowie's (1998) *Phraseology* concludes with remarks about the crucial role of word combinations in second-language learning and adult language production, an issue followed up by a wide range of studies which will be discussed below.

A vast body of research supports the urge to include formulaic sequences in EFL, ESL (English as a Second Language), and EAP curricula (Ädel & Ermann, 2012; Allen, 2016; Biber & Gray, 2010; Chen & Baker, 2010; Durrant & Schmitt, 2009; Ellis, 2002; Eriksson, 2012; Granger & Lefer, 2016; Handl, 2008; Huang, 2015;

Hyland, 2008a, b, 2012; Jones & Haywood, 2004; Sánchez, 2013; Pan, Reppen & Biber, 2016; Parkinson & Musgrave, 2014; Salazar, 2011; Simpson-Vlach & Ellis, 2010; Staples, Egbert, Biber, and McClair, 2013; Vincent, 2013). The primary goal of EAP is to help novice researchers acquire skills and knowledge which will help them to produce academic texts (written or spoken) which will be acceptable and, ideally, prototypical for their target genres. This skill is crucial for gaining acceptance into the academic discourse community (Swales, 1990) and publishing their research (Cargill & O'Connor, 2009).

Formulaic language is considered to be one of the indicators of membership in a specific discourse community (Allen, 2011; Erman Lewis & Fant, 2013; Yorio, 1989). In order to be accepted into a certain discourse community, one needs to acquire the discourse practices used in that community. It has been shown that achieving idiomaticity is the most challenging task for otherwise proficient English speakers who seek this integration. Idiomaticity or proficiency related to collocations is considered to be difficult to acquire by non-native speakers. Furthermore, there exists a correlation between idiomaticity and grammatical proficiency (Yorio, 1989).

The writing and publishing challenges faced by L2 writers have long been an object of study from a variety of angles and perspectives. Manuscript preparation and submission is particularly difficult for novice L2 writers, who lack the experience and the skills of accomplished authors. Nevertheless, many non-first-language users of English feel the urge to publish in English due to the growing readership of

English-language publications, encouragement (or sanctions in the case of not publishing) from their institutions, and the desire to communicate research internationally (Burgess, 2017). Researchers in certain disciplines, particularly social sciences, often rely on translators or editors to improve their texts and prepare them for publication in English, but this does not address the needs of L2 authors adequately, as a whole range of issues remains unaddressed (Cargill & Burgess, 2017). Although Hyland (2016) questions the disadvantaged position of L2 writers with respect to native speakers, the findings of many studies consistently point to the fact that non-native speakers are faced with considerably more hurdles than researchers publishing in their first language (Cargill & Burgess 2017). L2 researchers have been shown to experience the most difficulties with the Discussion sections of research articles, and this hurdle persists until they gain very high levels of proficiency in academic English (Moreno, Rey-Rocha, Burgess, López Navarro, and Sachdev, 2012). L2 researchers are often met with the frustrating recommendation from journal editors to have their paper revised by a native speaker (Lillis & Curry, 2015).

An element in these difficulties may be the use of fewer lexical bundles than is expected in the given type of discourse and discipline, or the overuse the same lexical bundles as a result of not being conscious of the discourse conventions of the given register, or failing to reproduce set phrases from model texts. The integration of novice researchers into a discourse community is a challenging and daunting process involving efforts to acquire knowledge of the genre conventions and the skills to

employ this knowledge. Among the specific linguistic skills novice authors must acquire are knowledge of phraseology and lexical bundles (Kazemi, Kohandani, & Farzaneh, 2014).

Writing research for publication requires many linguistic skills, such as adapting to the requirements of the particular journal (Cargill & O'Connor, 2009) and the acquisition of the genre conventions of the discourse community in question (Bhatia, 1993, Swales, 1990). Publication instructors recommend thinking of the English language of a particular discipline rather than the English language in general (Cargill & O'Connor, 2009). This involves the lexis, rhetorical conventions, and lexicogrammatical features specific to the given field of research. These conventions and features may well be as unfamiliar to novice L1 users of English as they are to those who use English as an additional language of research communication. Competence in spoken and written varieties of English does not guarantee successful integration in the academic discourse community: according to Hyland (2012), writing for publication requires specialised skills which need to be developed through conscious and focussed work. Therefore, novice L1 researchers will also benefit from learning the phraseology of their field and discipline. In fact, studies concerned with the generation of academic collocation lists (Ackerman & Chen, 2013; Salazar, 2011; Simpson-Vlach & Ellis, 2010) do not exclude the L1 speakers from the list of potential beneficiaries of their studies. Moreover, although published academic texts are regarded as a genre produced by experts irrespective of their L1, important differences have been shown to exist between the writing of L1 and L2 research

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

article authors, particularly in their use of lexical bundles (Pérez-Llantada, 2014). That said, many of the experimental courses aimed at teaching lexical bundles are also directed to native speakers (Cortés, 2006; Eriksson, 2012; Simpson-Vlach & Ellis, 2010).

Most of the studies addressing the use of lexical bundles by L2 writers point out overuse, underuse and misuse of the formulaic sequences. "Misuse" generally refers to a more literal interpretation of a lexical bundle, often resulting in a different pragmatic or semantic function as compared to L1 usage. Kazemi, Kohandani and Farzaneh (2014) investigate these features of lexical bundles in research articles in terms of their impact on the assessment of the article and its quality. Their conclusion is that reviewers are likely to give high scores to articles with moderate use of lexical bundles. In addition, the most negative impact is found not in case of underuse but in case of overuse of lexical bundles. This study underscores the fact that lexical bundles should be part of university curricula and should be taught explicitly to undergraduate students in order to help them develop the necessary academic writing and speaking skills.

The question as to whether L2 writers use more lexical bundles than L1 writers has been debated in various studies reporting contradictory results. The majority of researchers seem to agree that L2 writers overuse some lexical bundles and underuse others, which results in elevated numbers of lexical bundle tokens and reduced number of types in L2 corpora, misuse of certain bundles, as well as significantly

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

different sets of lexical bundles in L1 and L2 corpora (for example, Cortés, 2002, 2004, 2006; De Cock et al, 1998; Huang, 2015; Milton, 1998; Meunier, 1998; Neely & Cortés, 2009; Salazar, 2011). Native speakers are generally assumed to be more proficient in the use of formulaic sequences, as well as in their processing (Wray, 2002; Pawley & Syder, 1983), though it is difficult to find psycholinguistic evidence to support this assumption (Underwood, Schmitt & Galpin, 2004).

This notwithstanding, several studies report significantly more frequent use of lexical bundles in non-native corpora and others insist on considerably less frequent use of the same units by L2 speakers or writers. Thus, contrary to Hyland's (2008b), findings, Ädel and Erman, (2012), Cortés (2004), and Rica (2010) report that non-native speakers tend to use fewer lexical bundles in their writing, both in number and in variety. Chen and Baker (2010) find that the frequency and variety of lexical bundles increase along with the proficiency of the writers. Their study examines several corpora of L2 student writing at undergraduate level to contrast the frequency and functions of lexical bundles with those found in native user corpora. The results point to an existence of transfer of lexical bundles from the writers' L1 to English. L2 users are reported as using predominantly the lexical bundles whose equivalents can be found in their native languages, which are similar in structure, usage and functions. This study supports the view that it is necessary to provide formalised instruction to undergraduate students regarding lexical bundles and their use to make their writing more fluent, coherent and native-like.

Chen & Baker's (2010) article is one of the most representative contrastive studies of lexical bundles. They analyse expert and novice, L1 and L2 writing to compare their patterns of usage and functions producing results that show striking differences between expert and novice writers, both native and non-native. Expert writing relies on phrasal bundles, whereas both L1 and L2 student writing primarily make use of clausal structures. However, L1 students also displayed certain similarities with expert authors, such as the use of hedging devices, while L2 writers use certain idiomatic expressions and connectors, as well as over-generalisation which are exclusive to L2 writing.

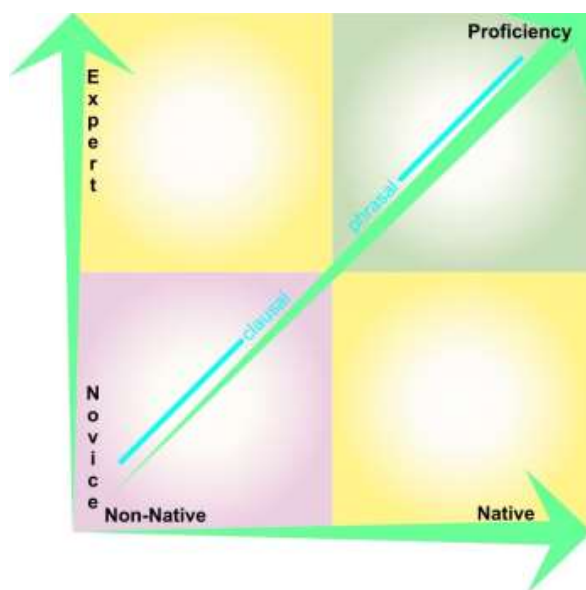
L2 researchers and learners experience great levels of stress because of the urge to write in English (Flowerdew, 2008), but native speakers, as was argued above, do not seem to be especially privileged when it comes to academic writing. Most novice researchers, be they native speakers or non-native speakers, experience considerable difficulties with the adequate use of lexical bundles. Indeed Hyland's (2016) controversial paper focusses on the difficulties faced by native speaker researchers when it comes to pragmatic meanings, supra-lexical units, and rhetorical structure of academic texts.

Pan, Reppen and Biber (2016) propose two dimensions for determining the linguistic competences of researchers: nativeness and expertise. These authors argue that the native versus non-native or expert versus novice dichotomy is not valid for determining the linguistic proficiency of a researcher. Rather, a two-dimensional

model should be used taking into account both the level of expertise and the first language of the researcher. These authors show that both non-native and novice researchers experience a similar process of evolution in the acquisition of academic register proficiency, progressing gradually from clausal language to phrasal forms of expression and incorporating more noun phrase-based lexical bundles into their writing. The latter are regarded as a characteristic feature of expert and mature writing (Biber & Gray, 2010).

Figure 1.4 is an attempt to represent this evolution with regard to the two dichotomies of native versus non-native speaker and expert versus novice writers.

Figure 1.4. The relationship between expertise, nativeness and proficiency in academic writing, according to Pan, Reppen & Biber (2016).



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The matrix implies that there is no direct correlation between nativeness and academic proficiency, although novice L1 writers appear to have a slight advantage over L2 users. There seems to exist a pre-established view of novice native writers as more proficient than novice non-native writers, although this dichotomy should not be regarded as extending to higher levels of expertise: experienced academics are assumed to be experts in academic writing regardless of their first language (Biber, Gray & Poonpon, 2011; Pan, Reppen & Biber, 2016). Many L2 speakers have achieved this status, while novice L1 researchers face challenges similar to those faced by L2 users regarding the integration into their target discourse communities (Hyland, 2016).

Parkinson and Musgrave (2014) confirm that proficiency is a result of expertise rather than nativeness (Biber, Gray & Poonpon, 2011; Pan, Reppen & Biber, 2016). The study outlines important differences in the use of noun phrases by EAP and MA students in their writing. EAP students, less proficient in English than MA students, tend to use more attributive adjectives as noun modifiers, while MA students rely on other noun modifiers such as nouns, possessive nouns, participial adjectives, prepositional phrases, and appositive noun phrases, all of these, according to Biber, Gray and Poonpon (2011), being indicators of a higher level of language development. Parkinson and Musgrave draw pedagogical implications which lead to a focus on noun phrase modifiers which are characteristic of later stages of development in order to improve the writing of EAP students.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Another contrastive study is that of Biber and Gray (2010), who examine a corpus of written university genres (research articles, university textbooks, department websites, course syllabi, etc.) and compare them with conversation to reveal the differences between them. Their findings challenge the assumptions that academic writing is complex, that it contains elaborated structures, and that the relations in it are explicit in comparison to conversation, the structure of which is presumed to be simple, less elaborated, and where shared context allows for less explicitation. Academic writing, the authors note, has so far been presented as ‘decontextualised’, impersonal and utterly elaborated. Their results show that none of the aforementioned assumptions holds true upon close examination. Academic writing turns out to be less ‘elaborated’ in terms of *that*-clauses. In fact, the authors show that many sentences in academic writing present a simple verbal structure with very few, if any, dependent clauses. Instead, this register shows a preference for phrasal structures, such as prepositional phrases used as nominal modifiers, structures which until recently have not been considered complex or elaborated, in contrast to dependent *that*- and *WH*-clauses which have been seen as indicators of elaborate writing. The authors mention the length of the sentences in academic writing which is generally the result of phrasal structures, prepositional phrases, noun phrases that occur in sequence, as well as appositive phrases. All these features, the authors argue, are in fact signs of elaborated writing and require a high level of proficiency to produce, although they are not typically perceived as such (for example, in Brown & Yule, 1983).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The authors go on to challenge the perception of academic writing as being characterised by explicit relations in contrast to the supposed implicitness of spoken discourse. Their results show that academic writing cannot be considered explicit, as the phrases in it have no grammatical cues to their meaning. The relations in academic writing, Biber and Gray argue, are concise and compressed, but not at all explicit. In fact, they proceed to note, speech is much more explicit than writing, even though it relies on the shared context to create non-verbal explicitness. This concision, according to the authors, is a time-saver for expert academics who can scan the text quickly and find the sections they are most interested in, making their reading effective and fast, but the concise and compressed nature of the writing is a challenge for novice readers who need to make sense out of the condensed language of academic writing. This quality is not unique to research articles: the authors found similar characteristics in all the academic genres they analysed. Thus, according to Biber and Gray, academic writing, though not ‘simple’ at all, is different in its complexity from conversation. The authors conclude that students and novice researchers need to acquire the style of academic writing to be able to find their way through complex, elaborated and implicit phrasal structures, and this need should be addressed in university courses aimed at increasing the learners’ fluency in academic writing.

The characteristics of expert and novice discourse are discussed by Ädel and Erman (2012) who analyse lexical bundles in two corpora produced by undergraduate students of linguistics. One subcorpus is composed of the undergraduate writing of

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Swedish students and the other consists of writing by British students. They perform a qualitative and quantitative analysis of the frequency and functions of lexical bundles contained in each subcorpus and compare their results with those obtained by Chen and Baker (2010). Although their corpus is significantly larger than that of Chen and Baker, their results compare well. Ädel and Erman (2012) report a wider range of lexical bundles used by native speakers, both in terms of type and token. 22% of the lexical bundles were shared by both groups, which, according to the authors, is a considerable amount. Ädel and Erman (2012) also identify four features that distinguish L1 discourse from L2: greater use of unattended *this*, existential *there*, hedging devices (also observed by Chen and Baker, 2010), and passive lexical bundles. All these have been shown to be indicators of expert discourse, alongside stance bundles which were also found to be more abundant in native discourse.

In a number of studies, L2 speakers were found to rely too heavily on a limited number of sequences and to underuse or misuse other strings favoured by native speakers (Cortés, 2004; Huang, 2015; Milton, 1998). For example, Huang (2015) analyses the use of lexical bundles in Chinese junior and senior student essays (first and second and third and fourth grades respectively) in terms of their accuracy. By accuracy Huang means the correct, appropriate, or native-like use of the lexical bundle. This involves several levels: orthography, bundle composition and correct use in context. The results show that even though senior students tend to use lexical bundles more frequently, their level of accuracy is no higher than that of the junior students. Both types and tokens of lexical bundles are more frequent in senior

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

students, but there are instances of misuse of these multi-word strings. The author concludes by suggesting that more implicit and formalised teaching is needed to cover L2 learners' needs concerning the adequate use of lexical bundles.

Staples, Egbert, Biber, and McClair (2013) look at the frequency, functions, distribution and fixedness of lexical bundles in TOEFL iBT (internet-based test) written assignments. Their corpus consists of written tasks submitted by 480 participants divided into 3 groups according to their proficiency. They use Biber, Conrad and Cortés' (2004) taxonomy for classifying lexical bundles into stance expressions, discourse organisers, and referential expressions, and Biber's (2009) idea of fixed and variable lexical bundles. This framework allows for the study of lexical bundles from a number of different angles and provides important pedagogical insights.

Firstly, Staples et al.'s study (2013) reveals a correlation between the level of proficiency of non-native speakers and the tendency to use lexical bundles: TOEFL iBT test takers with lower proficiency used more lexical bundles in their written responses than the test takers with higher proficiency, although the former relied more heavily on lexical bundles contained in the prompts, whereas the higher-proficiency participants showed a tendency to use other lexical bundles rather than repeat those that were already in the task.

This finding supports the idea that learners first acquire unanalysed chunks of language at early stages of language acquisition and afterwards, as their proficiency

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

grows, tend to analyse them and construct new chunks based on the learned models.

The less frequent use of lexical bundles not contained in the tasks may also indicate an attempt to avoid plagiarism and to produce original discourse.

The participants with the highest level of proficiency were the ones who used the lowest number of bundles overall, although their bundles were ‘original’, i.e., different from those contained in the prompts, whereas the less proficient test takers relied heavily on the bundles from the prompts, thus boosting the number of bundle types in their writing. When these prompt bundles were removed, the overall distribution of the bundles across the three proficiency levels turned out to be rather similar. However,

Compared to the highest level of test takers, level 2 used significantly more bundles overall, and this difference remained once the prompt-based bundles were removed. There seems, then, to be a pattern of use of more prompt-based bundles at the lowest level, meaning, as mentioned above, that these participants are relying heavily on repeating unanalyzed language from the immediate environment. In level 2, the participants developed skills to draw on bundles not found in the prompts, and seemed to rely more heavily than the highest level group on these bundles. Finally, the highest TOEFL scorers used fewer bundles overall, and while they were adept at developing their own recurrent sequences, relied on them less than test takers in level 2. (Staples et al., 2013, p. 220)

As for the functions of lexical bundles in TOEFL test written assignments, the authors show that their use of lexical bundles on all levels of proficiency resembles that of conversation rather than academic writing. Test participants use very few referential bundles which have been shown to be the most frequent type of bundles in academic writing (Conrad & Biber, 2004; Chen & Baker, 2010). Besides that, the fixedness analysis suggests that the test takers are adept at filling the variable slots with function words, a feature that has also been proved to be characteristic of conversation rather than academic writing (Biber, 2009). Finally, the use of continuous sequence bundles by all three groups of participants 75% of the time indicates a close resemblance to conversation patterns (Biber, 2009).

Staples et al. (2013) suggest that new approaches need to be developed to focus on teaching “bundle frames rather than fixed sequences” (p. 224) arguing that even the more proficient test takers display a lack of such skills as referring to abstract entities, common in academic writing. Staples et al. (2013) underscore the need for direct instruction of such features:

“Frequently used referential bundles could be explicitly taught and students could be given opportunities to practice using such bundles in the context of academic writing assignments” (p. 224).

The studies emphasising the importance of including formulaic language in university curricula are numerous (Allen, 2011; Cortés, 2013; Coxhead, 2015; Erman et al., 2013; Handl, 2008; Huang, 2015; Neely & Cortés, 2009; Nesi & Basturkmen, 2006; Salazar, 2011, 2014; Staples et al., 2013, etc.). This is due to the point

expressed repeatedly in the current study that these formulaic sequences have proved to be of considerable importance in terms of integrating into the academic discourse community and will be useful both for L1 and L2 researchers.

Explicit teaching becomes especially important for EFL learners who, unlike native speakers or second language learners, have not been exposed to academic or, for that matter, any other variety of English for the same length of time as native speakers. Studies show that the processing advantage described by Dörnyei, Durow, and Zahran (2004), Wray (2002), Wulff (2010), and other researchers is realised in different ways in native and non-native speakers. Ellis, Simpson-Vlach and Maynard (2008) use a series of psycholinguistic experiments to show that language processing (reading and understanding) is considerably facilitated by the presence of lexical bundles, or prefabricated units, but it occurs in slightly different ways in L1 speakers as compared to L2 speakers. For native speakers, the processing speed correlates with the MI (Mutual Information) score, whereas L2 speakers are more sensitive to the frequency of lexical bundles to process a text. MI score is a parameter used for calculating the probability of a collocation: it shows how likely two or more words are to co-occur by measuring the frequency of occurrence of each of the words and comparing them to the frequency of the collocation. The mutual information score is used in a number of studies and has been shown to be a reliable parameter for determining the strength of association between collocates (Ackerman & Chen, 2013; Biber, 2009; Durrant & Schmitt, 2009; Ellis, Simpson-Vlach & Maynard, 2008; Salazar, 2011, 2014; Simpson-Vlach & Ellis, 2010; Siyanova & Schmitt,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

2008). However, the MI score was originally conceived for two-word collocations and may not render accurate results for longer strings, as it does not take into account the order of the items in a string (Biber, 2009; Hyland, 2012).

Other parameters have been used to determine the strength and pedagogical relevance of collocations. For example, Siyanova and Schmitt (2008) report three studies examining adjective + noun collocation. L2 English learners' argumentative essays were contrasted against the writing of native English speaker students to extract all adjective + noun collocations and check their frequency through the BNC (Natcorp.ox.ac.uk). The results were surprising; both L1 and L2 writers used collocations half of which were very common in the BNC, a quarter of the collocations were not to be found in the BNC, while another 25% were present, but not very frequent. The second study involved native and non-native participants who were asked to judge 52 collocations in terms of their "commonness". Again, the intuitions of native speakers turned out to work quite well, though the non-native speakers also performed fairly well on this task. However, they failed to distinguish between extremely frequent and medium-frequent collocations, and showed poor performance on the judgement of uncommon collocations. In other words, if a collocation was very common, they were likely to judge it as such, but when the collocation was not very frequent or not frequent at all, they were likely to fail.

The third study measured the reaction times for judging the frequency of the same 62 collocations by two groups of participants: L1 and L2. The native participants were

both more accurate and quicker than learners on judging the frequency of the collocations.

Siyanova and Schmitt (2008) point out that while the accepted standard is to turn to corpus evidence rather than native speaker intuitions, their findings indicate that intuition may be a reliable criterion, as was the case with their study. This finding contradicts Sinclair's (1991) and Wray's (2002) view, who maintain that native speaker intuitions are often misleading and wrong (Siyanova & Schmitt, 2008).

To investigate the issue of processing advantage for L1 and L2 speakers, Schmitt and Underwood (2004) conduct an experiment studying eye movements of both L1 and L2 users while reading texts containing formulaic sequences. Their results do not seem to reveal any significant advantage for L1 speakers in terms of formula processing, which is counterintuitive and seems to contradict the personal experience of many EFL and ESP teachers.

Different attempts have been made to teach lexical bundles to students, with varying results. Schmitt, Dörnyei, Adolphs and Durow (2004) conduct an extensive study aimed at determining the extent to which various factors influence the acquisition of lexical bundles characteristic of academic discourse in international students. The bundles are selected on the basis of their frequency, presence in the learning materials of the students and their value for the students. The authors extract bundles from a corpus, add bundles suggested by other authors (Hyland, 2008b) and bundles based on words from Coxhead's (2010) academic wordlist. Intuition also plays a part in the

selection of the target bundles: the authors evaluate each candidate bundle in terms of its relevance for the students and seek advice from the instructors of the EAP course for which the study is conducted. The students are exposed to the target bundles at least once throughout the course, after which their progress is measured through a variety of tests. The overall progress of students is considerable, however, the surprising finding of this study is that neither the aptitude measure nor the attitude / motivation measure had any significant effect on the acquisition of lexical bundles. The authors suggest that other individual differences may affect this process.

Investigating this issue further, Dörnyei, Durow nad Zahran (2004) pose the following question: what learner characteristics and what environment features are most likely to contribute to the acquisition of social skills and the formulaic sequences related to them? They conduct a multiple case study exploring the cases of the most successful and the least successful formulaic language learners. Three factors appear to be the most influential in achieving proficiency in formulaic sequences: language aptitude, motivation and sociocultural adaptation, of which sociocultural adaptation is the most important. The absence or low level of sociocultural adaptation can only be compensated by very high levels of both language aptitude and motivation. By contrast, a high level of sociocultural adaptation can “override below-average initial learner characteristics” (p. 105).

Erman et al (2013) combine two different methods of studying formulaic sequences (which they call “MWSs – Multi-Word Sequences”) to analyse the spoken

production of English and Spanish advanced L2 speakers and corresponding native speaker control groups. The two methods they use are the lexical bundle method (corpus-driven, automatised, involving very little manual work) and the comprehensive method (identifying the multi-word sequences manually and checking them against corpora). Their results show a remarkably larger number of formulae in the two L1 groups as compared to the L2 groups. Although the authors admit that their study is small-scale and their corpus is not large, the results are significant and are likely to have important implications for second language teaching and for the study of formulaic sequences in general.

The importance of formulaic sequences for achieving native-like fluency is crucial, yet at the same time Erman et al. (2013) observe that even prolonged exposure to an L2 environment is not comparable to the life-long input enjoyed by native speakers.

Crossley and Salssbury (2011) argue that the phraseological proficiency of learners increases depending on the time they spend studying English. Their study is set in a major American university hosting a number of L2 learners preparing for the TOEFL examination. Four out of their six subjects show a considerable improvement in the use and variety of two-word clusters (bigrams) throughout the year. The authors, however, do not put much emphasis on the L1 environment, concentrating rather on the formal instruction received by the learners. It would be interesting to compare the results of this longitudinal study with the results obtained by students studying

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

English in their home countries, i.e., without the advantage of everyday exposure to an authentic L2 environment.

The role of sociocultural adaptation in the acquisition of formulaic sequences is investigated by Adolphs and Dorow (2004) in a longitudinal case study of two MA students from the sample of Dörnyei, Durow nad Zahran’s (2004) study, one of them somewhat integrated into the local community and the other quite (self-) isolated from it. Their results suggest a strong relationship between sociocultural integration and the acquisition of formulaic sequences, though the authors admit that the dimensions of the study do not allow for generalisations.

It is important to note that the studies cited above refer to the acquisition of idiomaticity, or native-like fluency, in terms of everyday interaction, not in academic discourse. Therefore, there is no contradiction between Pan, Reppen and Biber’s (2016) view of novice L1 speakers being almost as disadvantaged as novice L2 speakers and the idea of sociocultural adaptation.

Formulaic sequences are generally defined in literature as “prefabricated” or “unanalysed” chunks (Wray, 2002). What would happen if the learners were provided with whole sequences and encouraged to learn them as wholes, without applying any syntactic or semantic analysis to them? This is the question that Wray’s (2004) study attempts to answer, based on an adult learner preparing to participate in a Welsh TV programme. The learner is given unanalysed sequences to memorise for her presentation, but eventually introduces several errors typical of early stages of

learning a foreign language which involve attempts to analyse, among others. However, she is able to learn and deliver a competent presentation on cooking after an intensive but very short preparation (4 days), which is understood and well-received by native speakers. This approach seems quite unconventional, but it may offer good perspectives for second language acquisition.

Another creative way of teaching formulaic sequences is examined in Bishop's (2004) study, focussed on increasing the salience of formulaic sequences by highlighting them in a text in order to draw the readers' attention to them. The results show that readers are much more likely to check the meaning of a formulaic sequence if it is salient in the text, but not necessarily improve their performance on the knowledge of these sequences. Bishop addresses restricted collocations with opaque or semi-opaque meaning, but the teaching technique seems applicable to other types of formulaic sequences as well.

Lindstromberg, Eyckmans, and Connabeer (2016) report an experiment whose aim was to compare two teaching techniques for formulaic sequences. They administered a dictogloss task to two groups of students (native speakers of Dutch), one of which did a traditional dictogloss, while the other group received the task sheet with the target items listed on it. Their results show that the experimental group outperformed the traditional group both in the immediate and the delayed posttests. The authors point out the usefulness of this technique (listing the target items in the dictogloss worksheet) and its pedagogical potential in EAP classes. However, they add that

“an issue which is at least as worthy of researchers’ attention is the extent to which the modified dictogloss leads EAP learners to use the targeted [Formulaic sequences] autonomously, or at least under freer conditions than seen in our study, after a lapse of time greater than one week.” (p.18).

Research on lexical bundles in academic discourse is predominantly contrastive, focussing on their frequency, structure, functions and use in more than one field and / or more than one genre. In addition, several studies compare the use of lexical bundles by authors of different status: L1 and L2, expert and novice.

Nesi and Basturkmen (2006) use a corpus of university lectures from four major disciplinary groupings (Arts and Humanities, Social Sciences, Life Sciences, and Physical Sciences) to extract the lexical bundles and analyse their functions, particularly, their cohesive role in oral academic discourse. The number and distribution of the lexical bundles found in their corpus is virtually identical to that found by Biber, Conrad and Cortés (2004): 34 four-word bundles occurred at least 60 times in a corpus of 1,248,800 words. The authors found that four-word lexical bundles play an important role in discourse construction acting as cohesive devices and facilitating comprehension and processing, hence their pedagogical value for EAP learners.

Another study focusing on the use of lexical bundles in academic speech is that of Neely and Cortés (2009), featuring 5 lexical bundles and examining their functions in instructor and student speech. The findings show that the same lexical bundle can

perform multiple functions, and that concordance lines may suggest recurrent patterns of collocation and usage for certain lexical bundles. The authors highlight the pedagogical implications of their study and propose several lesson plans and activities for introducing lexical bundles in EAP / ESP classrooms.

The communicative and pragmatic functions of lexical bundles are examined in several studies of their functions as related to the rhetorical structure of discourse. One of the most influential studies along these lines is that of Cortés (2013) who described the distribution of lexical bundles in RA Introductions and related them to specific moves in the Swalesian CARS model (Swales, 1990). Le and Harrington (2015) analyse lexical bundles (which they call “clusters”) in Applied Linguistics RA Discussion sections and show how they relate to the communicative functions of each step in the *Commenting on Results Move: Interpreting results, Comparing results and Accounting for results*. Their findings reveal several patterns typical of these Steps, particularly such frequently occurring clusters as *finding(s) suggest(s) that, is/are consistent with* and other lexical bundles which clearly bear the communicative functions of commenting on research findings. Le and Harrington also confirm that hedging is a common strategy for communicating results in academic writing.

The Move/Step model is used by Ädel (2014) to develop a framework for selecting relevant data from a corpus for qualitative and quantitative analysis. She discusses the example of the pattern involving the anaphoric impersonal *it* to contrast native and non-native data, manually selecting cases of overuse, underuse and equal use of

the pattern. Her findings do not show any significant differences between native and non-native novice writers' use of the *it* pattern, with the exception of *it is clear* and *it is evident* slightly overused by L2 writers. The distribution of this pattern across the moves and its habitual collocates tend to be similar in the two corpora.

Oakey (2002) studies one pattern, *it is / has been noticed / observed that...*, which he calls a "lexical phrase" after Nattiinger and DeCarrico (1992), in published academic writing to compare it with Nattiinger and De Carrico's analysis. The findings of the study show that one formulaic pattern may have different functions in different segments of an academic text, as well as different functions of one single instance. Oakey admits that determining the function of a lexical item or a sequence is always subjective, however, his findings are in line with the work of other linguists (for example, Moon, 1998).

Allen (2011) addresses the issues of genre and generic conventions and that of discourse community, noting that lexical bundles are part of the genre conventions of each discipline and each discourse community, hence the notion that novice researchers or learners aspiring to join a discourse community should be instructed to use these sequences according to the expectations of their target audiences. Allen's study focusses on four-word lexical bundles in research papers produced by native speakers of Japanese, all of them students of the University of Tokyo. He compares the bundles in his corpus to those extracted by Hyland (2008b) and two corpora of published research articles: The Professional English Research Consortium (PERC)

Corpus and the British Academic Written English (BAWE) corpus, and finds that novice L2 writers converge with novice L1 researchers and published authors in the use of many bundles, although this, according to the author, should be regarded as a result of careful revision, editing and peer-review. 41% of the bundles in Allen's corpus are noun structures, which is in line with the results obtained by other studies and confirms the phrasal character of academic discourse (Ädel & Erman, 2012; Biber & Gray, 2010; Chen & Baker, 2010; Pan, Reppen & Biber, 2016). However, many bundles also diverge from the native and expert usage. Allen observes a possible underuse of passive constructions and total absence of structuring signals (a subtype of discourse organisers or text-oriented bundles). This is a surprising finding for a corpus of over 700,000 words comprising undergraduate student reports whose authors otherwise seem quite proficient. Allen attributes the divergent practices to transfer from Japanese and "input factors" (Allen, 2011, p. 111).

The influence of the first language on the writing of non-native speakers of English has been explored in a variety of studies (Flowerdew, 2010; Paquot, 2013; Rica, 2010; Tognini-Bonelli, 1996). The findings of these studies typically indicate that L2 writers tend to transfer formulaic sequences from their first language into English (Chen & Baker, 2010; Paquot, 2013; Rica, 2010; Spöttl & McCarthy, 2004; Tognini-Bonelli, 1996).

Flowerdew (2010) examines the use of signalling nouns in L1 and L2 argumentative writing in terms of four types of deviation from first language use: overuse, underuse,

misuse, and learner idiosyncratic forms. His findings reveal several of these deviations in the learner corpus, namely, overuse of certain signalling nouns, underuse of others, a narrower range of signalling nouns employed by learners and patterns similar to their first language, Cantonese. Flowerdew attributes the overuse of certain formulaic patterns by L2 students (*the first advantage of...*, *the second advantage of...*), as well as other deviations from L1 corpus, to first language influence or “teaching and/or text type effect” (Flowerdew, 2010, p.49).

Paquot (2013) finds that the French learners tend to use lexical bundles in a way which is similar to that of native English writers, but they seem to prefer the bundles which have direct cognates in their first language, and / or are more frequent in French. Thus, the sequence *deeply rooted* is used much more frequently by French writers than the synonymous and also very frequent *firmly rooted*. This is likely to be due to the fact that *deeply rooted* has a direct equivalent in French, *profondement enraciné*. Paquot concludes that the transfer effects of the use lexical bundles by French learners can be explained by Hoey's lexical priming theory, particularly in terms of L1 priming (Hoey, 2005).

Contrastive studies of lexical bundles across different languages are somewhat less numerous. One such study is Cortés (2008) addressing the similarities and differences between lexical bundles in history research articles in English and Argentinian Spanish. Although the number of identified lexical bundles in Spanish is almost twice as large as the number of English lexical bundles, the author finds

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

striking similarities between the two corpora. As expected in an academic text, most of the bundles were referential (Biber et al., 1999; Conrad & Biber, 2004; Staples et al., 2013), and many of them have equivalents or quasi-equivalents in both languages. Furthermore, most lexical bundles in both corpora were prepositional phrases and noun phrases, due to which the general character of the texts in both languages was defined as phrasal. This is in line with the findings of previous research on lexical bundles in academic prose (Biber et al., 1999; Biber & Barbieri, 2007) showing that academic prose tends to be phrasal, while conversation is essentially clausal. This corpus-driven study reports a major difference between the two corpora: in Spanish texts, a separate category of “structural only” (Cortés, 2008, p. 47) lexical bundles emerged. These bundles consisted of a preposition + *that* clause fragment whose counterparts in English would be shorter expressions and thus would be left out of the scope of the study since it examines only 4-word lexical bundles (e.g., *en el que la = in which the*). Finally, some bundles in English texts authored by Spanish researchers looked like direct (and often poor) translations from the source language.

Another contrastive study was conducted by Pérez-Llantada (2014). The analysis of an almost 6-million-word corpus shows that L1 English, L2 English and L1 Spanish academic texts diverge considerably in their use of lexical bundles, over one-third of the L1 Spanish bundles being ‘idiosyncratic’ (not occurring in the native corpus). Much of this divergence may be attributed to transfer from Spanish into English, and L1 influence is also supposed to play an important part in the use of the core bundles, i.e., those which occur frequently both in L1 and L2 subcorpora. Pérez-Llantada’s

study adopts a cross-disciplinary perspective covering different disciplines and focusing on genre, although the author stresses the importance of discipline-sensitive approach and calls for further research to shed light on this issue.

Already at the outset of corpus linguistics and modern phraseological research the pedagogical implications of studying formulaic sequences were patent: the pioneering studies in the field (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983) were carried out with explicit pedagogical purposes. The concept of lexical bundles emerged as a result of a corpus-based grammar textbook (Biber et al., 1999) designed to describe linguistic phenomena extracted from authentic data and aimed at providing the learners with actual instances of language in use produced by real people instead of supplying them with a portion of artificially, albeit skilfully, constructed examples.

The influence of corpus-based and corpus-driven research on EAP and SLA has been significant. Such data allow for the development of objective criteria to determine the frequency and relevance of the lexical units to teach rather than relying on teachers' and textbook authors' intuition. However, there is still a dearth of corpus-based materials in EAP and second language teaching environments. As Salazar (2011) observes, classroom teaching still depends greatly on the intuitions of the teachers and materials designers. The efforts of the pioneers of the field did not result in the immediate inclusion of formulaic sequences in language / EAP teaching

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

materials. In fact, phraseology can still be considered as gradually making its way to the language / EAP classroom.

Several studies have contributed to the inclusion of formulaic sequences in academic curricula by generating pedagogically useful lists of formulaic sequences. These are discussed below.

Simpson-Vlach and Ellis (2010) generated a list of pedagogically relevant lexical bundles employing a number of techniques to refine and adjust the list to the needs of modern learners. They extracted recurrent word-combinations of different lengths from several corpora and analysed them for relevance. The authors used a number of criteria to establish the relevance of lexical bundles, finally determining a formula based on two parameters: frequency and MI (mutual information) score. After that they asked twenty experienced EFL instructors and language testers to rate a large number of collocations in terms of their pedagogical usefulness and arrived at the conclusion that the combined frequency and MI score metrics yielded reliable criteria for determining the pedagogical value of the lexical bundles. They developed a concept called "Formula Teaching Worth" (FTW) on the basis of this analysis and used it for the generation of the lexical bundle list. An interesting feature of this study is that it analyses both spoken and written academic discourse, and it looks at lexical bundles of different lengths, from three to five words. The authors select lexical bundles according to the following criteria:

1. frequent recurrent patterns,

2. patterns occurring significantly more frequently in academic discourse than in other types of discourse,

3. patterns which are common for a wide range of academic genres.

The less relevant bundles presenting no difficulties for EFL learners or not constituting any pedagogical value were discarded, among them *and of the*. The authors' intention was to generate an empirically derived, psycholinguistically valid measure of utility, "Formula teaching worth". To develop this measure they considered parameters such as frequency, dispersion, log-likelihood, MI score, finally opting for combining frequency and MI score.

This work is also notable for updating and extending Biber, Conrad and Cortés' (2003, 2004) taxonomy of lexical bundles. Simpson-Vlach and Ellis (2010) compare their results to those obtained by other researchers and comment on the discrepancies, attributing them to the relatively low frequency range (10 per million), which was compensated for by introducing other metrics, the fact that both spoken and written discourse were considered, and the varying length of lexical bundles in their study. Their results disagree with those obtained by Hyland (2008a) who had claimed that there is no homogeneity across disciplines in terms of lexical bundles and their use. This difference, according to Simpson-Vlach and Ellis (2010), can be explained by the fact that Hyland had looked at lexical bundles consisting of four words, whereas most of the lexical bundles which occur across all disciplines consist of three words. This study is significant in many ways: the focus on many different disciplines, the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

study of lexical bundles of varying length, the analysis of both spoken and written discourse, and the explicit goal of creating a pedagogically valuable list of lexical bundles for classroom use.

Another attempt to create a pedagogically valuable list of recurrent word combinations was made by Ackerman and Chen (2013), who also used a hybrid approach combining frequency and human judgment to identify the most pedagogically valuable collocations in a 37-million-word corpus of academic English. Their focus is on collocations consisting of two content words (noun + noun, noun + adjective, verb + noun, adverb + verb, etc.). The resulting Academic Collocations List (ACL) comprises over 2.000 units and appears to be the phraseological equivalent of Coxhead's (2000) Academic Word List. It can be said that these two lists, the AFL devised by Simpson-Vlach & Ellis (2010), and the ACL created by Ackerman and Chen (2013), are complementary, as the former comprises bundles of 3-5 words, and the latter covers 2-word collocations (bi-grams).

Salazar (2011) also aims at generating a pedagogically relevant list of lexical bundles to be taught in natural sciences EAP classrooms. She carries out a corpus-based study focussed on identifying the most relevant lexical bundles and creating a list that would be helpful in language teaching. She also analyses the use of lexical bundles by L2 (Spanish-speaking) authors and compares it to the SciE-Lex corpus outputs in order to determine the strategies of published Spanish-speaking authors related to the use of lexical bundles in their texts. The list of lexical bundles used frequently in

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

academic writing is extracted from a corpus of published research articles, all authored by native English-speaking authors (or authors perceived as such due to their family names and institutional affiliation) in biomedical science. After identifying all the multi-word sequences, the list was scrutinised to determine the most pedagogically valuable lexical bundles in the context of an EAP classroom. Salazar (2011) uses frequency, MI score and the idea of “prototypical bundle” to identify the lexical bundles of utmost pedagogical value. The prototypical bundle is a core lexical bundle with canonical meaning which is central to the paradigm of other bundles sharing the same keyword. Thus, the bundles *an important role, a critical role, an essential role* are variations of the core bundle *a role in*, while the constituents of the pair *were able to* and *are able to* perform different functions: the former is preceded by the pronoun *we* or by the noun *colleague* and relates to the accomplishments of the author(s) or their colleagues, whereas *are able to* collocates with nouns related to research subjects, hence these two similar lexical bundles are classified as two different prototypes (Salazar, 2011). Her results suggest that non-native writers, all of them experts in their fields and authors of published research in English, although demonstrating similarities in use, frequency and functions of lexical bundles as compared to native authors’ writing, tend to underuse several lexical bundles which are common in native texts, as well as overuse other lexical bundles. Salazar attributes this overuse to the phenomenon of “linguistic teddy bears” (Hasselgren, 1994).

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The hypothetical reliance on “linguistic teddy bears” is confirmed by Durrant and Schmitt (2009) who use t-scores and MI scores to compare bi-grams (two-word collocations) in several corpora of native and non-native writers. Their main conclusion is that non-natives do make use of collocations, though they prefer to stick to certain favoured formulae with low MI scores, as opposed to native writers who tend to use strong collocations with high MI scores.

L2 authors, thus, including those who are expert members of their discourse communities, may benefit from learning to adequately use lexical bundles to make their texts more native-like, coherent and fluent. Salazar (2011) underscores the idea that it is unknown whether these authors use other lexical bundles or other linguistic devices to perform the functions covered by the corresponding lexical bundles in native texts, as her study limits itself to identifying a set of target bundles and not analysing the whole corpus for multi-word sequences or pragmatic devices. In addition to the list of teachable lexical bundles, Salazar suggests a range of exercises and activities designed to help learners to acquire lexical bundles, most of them similar to vocabulary-learning exercises.

The acquisition of lexical bundles and formulaic language in general is commonly perceived as acquisition of vocabulary; most of the studies aimed at teaching lexical bundles are based on activities typically used for teaching vocabulary (Alali & Schmitt, 2012; Jones & Haywood, 2004; Cortés, 2006). Although Cortés (2006) observes that the efficiency of certain vocabulary exercises is not attested even for

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

teaching lexis, such exercises (as, for example, ‘fill in the gaps’ exercises) are being used not only for their primary purposes (teaching lexis), but also for teaching other linguistic content, such as lexical bundles, which are in fact lexico-grammatical units. The effectiveness of these methods for teaching is partly confirmed by Alali and Schmitt (2012), with the reservation that their study was aimed at teaching idioms which often function as single words and have one-word semantic equivalents (Martin, 2008), and hence can be treated as single lexical units for the purposes of teaching vocabulary. Lexical bundles, on the contrary, cannot often be replaced by single words due to their structural incompleteness. However, most of the studies focussing on the teaching of lexical bundles have so far used vocabulary acquisition exercises and learning activities (Cortés, 2006; Jones & Haywood, 2004; Salazar, 2011).

One of the most exhaustive studies of teaching lexical bundles, called “formulaic sequences” in the study, is that of Jones and Haywood (2004). The authors designed a training programme for a list of about 70 target bundles to be taught to a mixed-discipline group of 10 students from diverse L1 backgrounds in a pre-sessional EAP course setting. The duration of the course was 10 weeks, during which phraseology sessions were presented for 2 hours per week. The authors designed a large variety of exercises, including paraphrasing and c-tests, adapted several academic texts for the students to peruse, and held extensive discussions of formulaic sequences, their structure, meaning, use, and functions. The authors report that as a result of these activities, the awareness of formulaic sequences among students clearly increased,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

together with their understanding and correct usage of lexical bundles in exercises. However, the students did not show significant improvement in their written production in other contexts, for instance, in their essays. Thus, the students learned to tackle exercises focussing on lexical bundles and were capable of noticing and retrieving, the first two stages of learning vocabulary according to Nation (2001), but they still lacked proficiency in generation (production) of formulaic sequences, the third stage of learning. This is a surprising finding, because the students in Jones and Haywood's (2004) study felt the course to be enriching and insightful, they seem to have understood the importance of phraseology and were willing to improve their writing.

Another study aimed at teaching formulaic sequences is AlHassan and Wood (2015), addressed to a group of 12 Economics students taking an EAP course at a Canadian University. The authors taught a course designed on the basis of several dozen pre-selected formulaic sequences presented to students. The procedures and precise descriptions of the exercises and activities used in the course are not provided in the study, but the results are encouraging: all the participants showed a significant improvement of their writing technique in the immediate and delayed posttests. However, it should be noted that the students performed one and the same task throughout the course (writing an abstract in their discipline). They re-wrote one and the same abstract during both posttests, which can be regarded as a reiteration of the same activity, likely to improve due to consistent repetition over a certain period of time, whereby the use of formulaic sequences would contribute significantly to the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

improvement. However, it is unknown to what extent these or other relevant formulaic sequences would appear in other texts written by the course participants.

In general, learners seem to respond well to courses aimed at teaching formulaic language, but their text production does not appear to improve substantially after training, as in the case of Cortés' (2006) study where the author delivered 5 micro-lessons on lexical bundles to native speaker students of a history writing-intensive class. Only four-word lexical bundles were selected for the study, since, as the author argues, they are more frequent than longer bundles, and most of them also contain three-word bundles. The micro-lessons were designed as traditional EAP lessons with explanation of the structure and functions of lexical bundles followed by typical vocabulary exercises and activities aimed at practicing phraseological skills. They also involved discussions of lexical bundles. However, the final texts produced by the students did not contain many more bundles than the ones written prior to instruction, neither did the variety of types increase with training. As in the case of Jones and Haywood's study (2004), the students seemed motivated and claimed to realise the importance of formulaic language in writing, but did not show any substantial improvement in producing it. Cortés (2006) suggests a number of explanations including insufficient duration of the course, the explicit or implicit requirements and communicative purpose of the genre of student essays (they might feel that their intended audience, namely, professors, did not expect them to use formulaic language as it is used in published academic texts), or the shortcomings of

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

the teaching method and materials which are extremely common in language teaching, though their efficiency has not been confirmed empirically.

However, several attempts to teach formulaic language were more successful. One of them is Peters and Pauwels' (2015) study aimed at teaching academic formulaic sequences to a group of students at a Dutch university. Course duration was ten hours distributed in five sessions. The instructors engaged students in activities of noticing and retrieval (Nation, 2001) of 24 formulaic sequences extracted from the Manchester Academic Phrasebank (Morley, Phrasebank.manchester.ac.uk, n.d.), a brief description of which is provided in Section 5.4. One of the noticing activities consisted in reading "sentences or excerpts from scholarly articles in which they had to underline formulaic sequences that seemed relevant to academic discourse" (Peters & Pauwels, 2015, p. 32). This approach may be pedagogically viable, but it must be borne in mind that formulaic sequences, as a general rule, are defined as strings which are extremely frequent and hence familiar to the language user due to multiple encounters with them. It is very difficult, if not impossible, for a language learner to identify a string of words as a formulaic sequence based on their intuition only. In other words, a learner has no way of knowing that a particular sequence is a formulaic expression unless s/he has encountered it several times in speech and/or writing or has been explicitly informed of the formulaic nature of the string. The results of the study, however, were quite encouraging: the students who had taken the formulaic sequences training marked high scores on the posttest on formulaic sequences and were able to produce final papers of better quality, in accordance with the academic

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

expectations of their instructors, containing a greater number of academic formulaic sequences used according to the genre conventions.

Vincent (2013) proposes a methodology inverse to the conventional corpus-based and corpus-driven approaches: the author suggests extracting collocations from a chosen text based on intuition and searching for them in a corpus to see if these intuition-based phrases are frequent or not. The next stage of this method is to identify the different variables for the phrasal core and to find collocates for every given phrase by using concordance lines. After that the sequences are checked for their pedagogical value and teachability, and a paradigm of variables and collocates is developed for each phrasal core. Thus, a compendium of paradigms is generated which is then brought into classroom together with learning activities and exercises. This approach is interesting in several ways: first, it is more holistic, offering a wider view instead of focussing on one lexical bundle at a time, second, it proposes many insightful and engaging learning activities, third, it enables the students to create their list of lexical bundles based on the texts they are studying, and explore them in terms of collocates, variables and usage. Furthermore, this approach covers both fixed and variable formulaic sequences. However, some drawbacks of this method should be mentioned. First, it is not, strictly speaking, corpus-driven, but rather intuition-based, at least at the first stage. Secondly, the results would largely depend on the selection of the source text from which collocations are to be extracted. Finally, the approach lacks rigour and solidity as its results will always vary and each genre, register, or even text will produce a different set of phrases and collocations for different users.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Most pedagogically-oriented studies of lexical bundles have treated academic discourse as a homogeneous field, without dividing it into disciplines (for example, Allen, 2001; Biber, Conrad & Cortés, 2004; Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortés, 2013; Nesi & Basturkmen, 2006). However, several studies have focussed on discipline- or field-specific bundles. Thus, Farvardin, Afghari and Koosha (2012) analysed a corpus of physics research articles to create a list of lexical bundles common in that science. Eriksson (2012) conducted a series of workshops with biochemistry doctoral students in order to identify and help them acquire the lexical bundles in that discipline. Pan, Reppen and Biber (2016) analysed the structure and functions of lexical bundles in telecommunications. Ädel and Erman's (2012) study addresses the bundles in linguistics. Durrant (2009) and Hyland (2008b) contrast the types of lexical bundles in different disciplines. Their findings show that a discipline-specific approach is likely to render better results in the teaching of lexical bundles. A significant amount of research on lexical bundles in academic writing has focussed on natural sciences, particularly in the fields of biology (Cortés, 2004), biomedicine (Salazar, 2011), physics (Farvardin et al., 2012), biochemistry (Eriksson, 2012), emphasising the importance of a discipline-specific approach in lexical bundle studies and the pedagogical implications of such an approach.

Eriksson (2012) underscores the importance of discipline in the selection of lexical bundles to be taught. He uses a corpus of biochemistry research articles to extract the most frequent multi-word sequences and also contrasts this list with the lists created by Simpson-Vlach and Ellis (2010) and Hyland (2008b) to generate a comprehensive

list of lexical bundles for the field. His teaching methods are based on Nation's (2000) three principles of vocabulary acquisition (noticing, retrieval and generation), and some of the activities designed for the students seem to replicate the well-known vocabulary exercises. However, other activities were designed to involve the students in the identification and selection of the formulaic sequences: the students were instructed to use corpus analysis software using mini-corpora to extract lexical bundles relevant for their field, and were taught to use them in their writing. Eriksson's study offers pedagogical insights and a methodology for the teaching of lexical bundles to post-graduate (doctoral) students. One interesting aspect of this study is that the students took part in the identification of the most valuable lexical bundles before starting to learn and analyse them. Eriksson suggests that many lexical bundles are acquired by the students implicitly through reading, and it would be redundant to teach the already well-known bundles to these students. Eriksson's study is original and insightful, and it would be interesting to follow up the work of the participants of his workshops and analyse their writing to investigate whether their use of lexical bundles has changed.

The idea of using specialised corpora for generating discipline-specific lists of lexical bundles instead of using one-size-fits-all generic lists seems to be gaining popularity in phraseological research. In their influential article on the use of lexical bundles across disciplines, Byrd and Coxhead (2010) pose a number of questions regarding the role of lexical bundles in academic discourse and the challenges faced by students, teachers, materials designers, and researchers when dealing with these

formulaic sequences. One of their suggestions is that the students should work with the lexical bundles which they will need for their reading and further studies. Although lexical bundles are much more frequent than any other sequence in a given corpus, their occurrence rate is still not high enough for students to familiarise themselves with a large number of relevant formulae. The example cited by Byrd and Coxhead is that of the lexical bundle *on the basis of* which occurs 308 times in a 3,6 million word corpus; hence the authors calculate that it will occur twice per 15,625 words. It is rather improbable that an average student would read that much so as to encounter even the most frequent lexical bundles in sufficient quantity. Therefore, the solution should lie in using corpora and focussing the attention of the students specifically on these formulaic sequences, taking into account the discipline they are majoring in and the genres they will most probably use in the future. Byrd and Coxhead (2010) argue that for EAP classes the focus should be as specific as possible, providing the students with lexical bundles which will be useful for them later. They suggest using sequences which are frequent in the students' reading curriculum. Hatami (2015) supports this view stating that the task of teaching all the possible formulaic sequences in a classroom is impracticable, hence a selective approach should be adopted based on the frequency of the formulae and the goals of the learners. Peacock's (2012) study of high frequency collocations in eight major disciplines confirms that each discipline possesses its own set of collocations which can be regarded as genre or discipline conventions. His findings show consistent differences between individual disciplines, even if they belong to the same broad

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

area, for instance, frequent collocations in biology will be different from those in biochemistry, although they both belong to the broad area of natural sciences.

It seems that in arts and humanities students rely heavily on their capacity to produce original texts, while in ‘hard’ sciences they tend to reproduce the formulaic language that they come across in the professional literature of their field. Research has shown (e.g. Cortés, 2004) that native English speaker students of history also have difficulties with reproducing the accepted discourse practices in their field, even after taking a mini-course focused specifically on the correct use of lexical bundles. This circumstance may be due to the fact that arts and humanities students are supposed to meet higher expectations regarding their written expression skills and feel they need to be creative and original in their writing, while also having the implicit conviction that they are not allowed to ‘copy’ (plagiarise?) the text of other authors. Students in the hard sciences, on the other hand, are less concerned with the form and creativity of their writing and try instead to make use of all the available linguistic resources in order to publish their study. Indeed, Salazar’s (2014) study reveals that students in natural sciences responded to a focussed course on the correct use of lexical bundles by improving their results significantly. This difference between natural sciences and humanities and social sciences students is a tentative hypothesis, however, and it is important to take into account that the two studies cannot compare directly because the course duration in Cortés’ study was significantly less than in Salazar’s. Furthermore, these results may be due to a number of other factors, such

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

as the institution, the students' cultural and linguistic backgrounds and the way the courses were designed.

We can conclude that lexical bundles, as important building blocks of discourse, are one of the most relevant types of formulaic language for EAP students, whether native or non-native, for a number of reasons. First, they make speech or writing fluent and native-like, as has been shown throughout the review of the relevant literature presented here. Second, formulaic language is a mechanism for faster text processing, i.e., it represents a processing advantage for readers and listeners (Elis, Simpson-Vlach & Maynard, 2008; Simpson-Vlach & Ellis, 2010; Tremblay et al., 2011; Wray, 2002). Third, it is one of the linguistic features characteristic of a particular genre and hence constitutes a mechanism for the validation of a text as a prototypical instance of the genre (L. Flowerdew, 2002; Paltridge, 1997; Swales, 1990). Finally, formulaic sequences are also a crucial component of the linguistic features of each discipline, and as such are important for establishing membership in the target discourse community.

Formulaic sequences may present difficulties not only for L2 learners and non-native or novice members of a discourse community, but also for translators. The problems faced by these professionals include transfer of phraseological units and phraseological borrowing. In relation with the latter, it may be the case that a translator comes across an expression which is widely known by the speakers of the target language in its original English version, but has no adequate translation in the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

target language. Examples of such problematic items are: *famous last words*, *no comment*, along with well-known English catchphrases and sayings.

Fiedler (2017) distinguishes between three types of phraseological borrowing: direct borrowing (the phrase is borrowed as is from L1 into L2 without translation or any modification), hybrid borrowing (the phrase is partially translated, normally with the core item remaining untranslated and the accompanying function words being translated into L2), and loan translation or calque (when the item, typically a catchphrase, is translated literally, word-for-word from L1 into L2 often violating the grammar and priming principles of L2 and resulting in an unnatural collocation). She observes that most cases of phraseological borrowing are difficult to detect since they represent the latter type, calquing. Although unauthentic collocations are likely to elicit a negative reaction from native speakers, this effect may be swiftly erased through repetition, especially if the phrase becomes very common, thus making the collocation sound familiar and common and making it more difficult to identify the calquing. This is what happens with most loan translations from English into German, as well as many other languages. As a result, cultural and pragmatic concepts of English-speaking countries (notably the United States) get filtered into the cultures of target languages and the speech of their inhabitants.

Translation has always been an arbitrary and risky endeavour. The translator is inevitably faced with many hurdles, one of the most famous being the absence of direct equivalents in the target language. The issue inevitably arouses the eternal

question: what is the unit of translation? The word? The phrase? The meaning? The translator's permanent dilemma is slowly being formulated in linguistic terms. For example, Tognini-Bonelli (1996) proposes a methodology of identifying formal patterns that are correlated with certain meanings in the source language and the target language. She focusses on three types of contextual features, collocation, colligation, and lexical sets by examining their manifestations in a corpus and in concordance lines. This system allows the translator to study the relationships between lexical units in the source language and find their equivalents in the target language to "negotiate equivalent effects within different contexts" (Tognini-Bonelli, 1996, p. 216).

Monti, Barreiro, Elia, Marano, and Napoli (2011) show how the integration of formulaic sequences into translation software can improve the quality of machine translation.

From the above discussion it can be concluded that lexical bundles are an important part of speaking and writing competences, and therefore should be included in ESP / EAP curricula.

This section attempted to provide a discussion of the pedagogical implications of research on lexical bundles for EAP and EFL classrooms. As can be seen from the above discussion, knowledge of lexical bundles is an important part of genre knowledge for novice researchers, both native and non-native. However, despite the large number of studies supporting this idea, phraseology is still a marginal issue in

language classrooms. Although the theoretical dimension of phraseology has moved from periphery of linguistic research to the centre, the practical instruction remains very much on the periphery.

The present study is an attempt to contribute to the knowledge on the practical dimension of lexical bundles by attempting to raise phraseological awareness among novice researchers and help them acquire the lexical bundles that they will need for their work.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

CHAPTER THREE

CORPUS AND METHODOLOGY

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

3.1. CONTEXTUALISING THE PRESENT STUDY

The previous chapter outlined the studies of formulaic language and particularly of lexical bundles. Several key features of lexical bundles were discussed, namely, their presence in all registers and genres, their relatively lower frequency in academic discourse, their crucial role as building blocks of discourse, and their discipline-specific character.

The purpose of this study is to analyse the patterns of usage of lexical bundles in three disciplines, namely psychology, linguistics, and literary studies which seem to be underresearched in terms of lexical bundles, and to address the needs of novice L2 researchers by devising a methodology for student-centred teaching or self-directed learning with regard to lexical bundles.

This chapter will highlight the methodology, corpus, approach and criteria used in the present study. The next section describes the corpus used in this study.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

3.2. METHOD

The present study relies on a set of four subcorpora, two of which were collected for a small-scale preliminary analysis of novice L1 and L2 writing, one represents a collection of expert academic writing (research articles from the three disciplines), and the last one is a corpus of texts produced by the participants of the workshop.

The study involved some preliminary research into the phraseological features of upper-undergraduate and post-graduate student writing of the previous years as compared to an equivalent corpus of novice L1 authors. The aim of this preliminary research was to gain an understanding of the use of lexical bundles by a L2 group of novice researchers comparable to the group of the future participants of the workshop. The purpose of this preliminary research was to reach a better understanding of the needs of the participants in order to be able to address them adequately and make the workshop as useful as possible. This included a preliminary study of the lexical bundles typical of the selected disciplines, and an initial analysis of Final Degree Research Projects written by the students of the Department of English and German Philology of the University of La Laguna during the past two academic years (2014-15 and 2015-16) to identify possible cases of overuse,

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

underuse, and misuse of lexical bundles, as described by Cortés (2004, 2006), Chen & Baker (2010), Pan, Reppen & Biber (2016), and others, discussed in the previous chapter. The Final Degree Research Projects (Trabajo de Fin de Grado, or TFG in Spanish) are the Spanish equivalent of a Bachelor Thesis and are the first piece of research that an undergraduate student typically produces over the course of her studies. In this study they will be referred to by the Spanish acronym, TFG, for convenience, and the corpus of the TFGs from 2015-2016 will be called ‘TFG’15-16’. The primary purpose of a TFG is to teach research skills to undergraduate students, for which reason it can be described as a pedagogic research genre (Hyland, 2008b).

For the preliminary study, two subcorpora were used, the TFG’15-16, and the PhD subcorpus, as described in Section 2.3. The PhD subcorpus was used to generate a list of lexical bundles which could be potentially useful to the participants of the workshop. these bundles were used as “target bundles” (Cortés, 2006) to examine the TFG’15-16 corpus and identify learner strategies of the use of lexical bundles. Although the learner corpus was composed of undergraduate student writing, PhD theses were selected for the native corpus because they potentially represent a ‘target genre’ for TFG writers. The primary purpose of the TFG subcorpus was to identify the general patterns of lexical bundles and their usage, and compare them with the output of the two groups of learners, those who participated in the workshop, and those who did not. Despite the small size of the corpus, several patterns emerged, as will be discussed in the next chapters.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The second stage of the study was based on the other two subcorpora, the Expert corpus comprising research articles from the three disciplines, and the workshop participants' corpus comprising novice L2 texts written after attending the workshop. these are also described in Section 2.3.

The analysis was performed using AntConc (Anthony, 2010), free software designed for corpus analysis.

The first stage of the study included only the Literature Review and Methods sections of the final degree projects, as it was expected that most of the students would have finished that part of their work early enough to submit them for analysis.

The two L1 corpora (the PhD subcorpus and the Expert corpus) were converted into separate .txt files using AntFileConverter (Anthony, 2015) and each file cleaned from titles and subtitles, author biographies, headers, footers, page numbers, figures, graphs, tables, footnotes and other irrelevant textual elements. The L2 corpora were not treated in this way, as they were not going to be searched for emerging patterns but for target bundles extracted from the native corpora. Instead, a .pdf reader was used to search the texts *en masse* for target bundles which had emerged from the native corpora. The software that best suits for this purpose is *Foxit Reader* which allows one to search for a word, phrase, or even pattern in a collection of .pdf documents without having to open each document and search it separately. Foxit Reader was used to search all the .pdf files in the two learner corpora for the target bundles and their immediate contexts.

The criteria for the analysis of the PhD subcorpus are presented below.

The search for three-word lexical bundles in a reduced corpus of the first two chapters had a very conservative frequency cut-off point of 20. In addition, the dispersion (range) criterion was fixed as high as 4, which means that the bundle had to occur at least 20 times in the corpus (20 per 200.000 words = 100 per million words) and at least in four of the 28 texts. While it is not common practice to fix such conservative thresholds in this kind of study, my purpose was not to identify all the lexical bundles that occur in this genre and discipline, but only the most frequent and common ones. For the purposes of this study, I decided to address only the most frequent lexical bundles which would be teachable to a group of students in a limited period of time during an introductory session.

The frequency cut-off point for four-word lexical bundles was set at 10 (50 per million words), and the dispersion threshold was maintained at 4. This search yielded a list of 28 extremely frequent lexical bundles.

For five-word bundles, the frequency cut-off point was reduced to 5, and the dispersion threshold to 3. This search yielded a list of 16 five-word lexical bundles.

The three lists of three-, four- and five-word lexical bundles were then compared to identify shorter bundles that are subsumed in longer ones, such as *the nature of* and *the nature of the*, and remove any irrelevant bundles that had found their way to the top list due to some kind of error. Appendix 2 contains the 3 initial lists of the bundles identified at the first stage of analysis.

At this stage, the lexical bundles representing topic-specific collocations (*teaching and learning*), toponyms (*in the United States*), and other irrelevant bundles were removed from the lists. Several lexical bundles were discarded from the initial list due to the following reasons:

- Inadequate frequency / range ratio: the bundle *higher education and* occurred 29 times in 4 texts, which suggests that it was topic-specific. The bundles *the philosophy of*, and *teaching and learning* were also excluded as topic-specific, and because they were felt to present no difficulty for advanced learners.
- The bundle *the context of the* was removed because in most cases it was part of the larger bundle *in / within the context of (the)*. The remaining three instances of this bundle appearing alone would not have been included in the list by the software if the longer bundles had been counted separately. The same applies to *on the other* as part of *on the other hand*, and *the heart of* and *at the heart of*, of which the longer bundle was included in the list of target bundles.

After cleaning up the list of lexical bundles, their occurrence patterns were checked in the TFG'15-16 subcorpus. The L2 TFG'15-16 corpus contained about 200.000 words, roughly equivalent to the L1 PhD corpus. The preliminary results suggested that many of the bundles were overused, and several others were underused in the L2 corpus. The bundles which showed a frequency similar to that of the native texts were removed from the list of target bundles (see Appendix 2 for all the bundle lists for the introductory session).

The principal corpus of this study is the research articles (Expert) corpus, containing 180 texts, 60 from each discipline selected from representative journals (see Section 2.3 for a detailed description). It was searched for three-, four-, and five-word frequent clusters using AntConc (Anthony, 2010).

Separate searches were conducted to extract three-, four-, and five -word lexical bundles in each discipline, a total of 12 searches. Table 3.1 shows the criteria and cut-off points for each of the searches. One final search was performed through the whole corpus to extract the four- and five-word lexical bundles shared by all the three disciplines. The raw frequency threshold for this generalised search was established at 60 (30 per million words), but the distribution threshold was raised to 18 in order to limit the results to bundles contained at least in 10% of the texts.

Table 3.1. Frequency and range cut-off points for each type of clusters in each discipline.

	3-word			4-word			5-word		
	Freq	Normed freq	Range	Freq.	Normed freq	Range	Freq.	Normed freq	Range
Psychology	50	62,5 p.m.w.	4	16	20 p.m.w.	4	12	15	4
Linguistics	50	62,5 p.m.w.	4	16	20 p.m.w.	4	12	15	4
Literary Studies	31	62 p.m.w.	4	10	20 p.m.w.	4	7	14	4

A lower frequency threshold for literary studies was established due to the smaller size of the corpus: the literary studies research articles are on average 15% shorter than in the other two disciplines, contrary to literary studies PhD theses, which are considerably longer than those of the other two disciplines. For this reason the

number of literary studies texts is 14, as opposed to 15 for the other two disciplines in PhD corpus.

Bundles of different size were searched separately due to the overwhelming abundance of three-word bundles in any register (Biber et al., 1999), and relatively lower frequency of five-word bundles. A combined search would yield inflated numbers in three-word bundles and negligibly low numbers for five-word bundles. The specified criteria seemed the most appropriate for the purposes of this study and reflected the phraseological content of the corpus adequately.

The resulting lists of lexical bundles were examined to remove unrelated text fragments, such as *et al s* (the software treats the apostrophe as a space and considers words containing apostrophe as two separate lexical items), or any other elements that could have escaped my attention while cleaning the files.

The final lists of three-, four-, and five-word lexical bundles were combined in one for each discipline. At this stage, all the bundle fragments subsumed in longer bundles and showing a frequency similar to that of the longer bundle were removed from the list. For instance, *the other hand* and *other hand the* were both removed because they were regarded as parts of *on the other hand the* whose frequency was very similar to those of the two shorter bundles, i.e., the shorter bundles virtually always occurred as parts of *on the other hand the*. By contrast, *as well as* and *as well as the* were kept as separate bundles because their frequencies differed drastically,

and while *as well as* occurred many times as part of *as well as the*, in most cases it occurred separately and was counted as such.

The combined lists of lexical bundles were analysed structurally and functionally, and searched in the learner corpora to compare their frequencies and usage. These can be found in Appendix 3.

The bundles were classified according to Biber, Conrad and Cortés's (2004) functional taxonomy updated by Hyland (2008b). For functional classification, only the broad categories were used, namely, text-oriented (discourse organisers), research-oriented (referential), and participant-oriented (stance) bundles. This was due to the fact that further classification according to subcategories would inevitably be subjective, and would suppose little pedagogical value. This issue is discussed in detail in the following chapter.

Interrater reliability test was used for classification, reaching 88% of agreement. All instances of disagreement were resolved on reference to the concordances and the context.

The next section will provide detailed information on the set of corpora used in this study.

3.3. CORPUS

The PhD subcorpus includes PhD theses written by presumably native speakers of English (judging by the name and awarding institution) in linguistics, literary studies and education. The latter was included because initially several students from the Education department expressed their interest in attending the workshop. Later these students were not able to participate, and Doctoral students from the Faculty of Psychology were invited, but since this corpus was collected only for the introductory seminar, these texts were not replaced by others.

An initial difficulty with this subcorpus was that several theses, notably those on literary studies, did not have explicitly signalled or even implicitly highlighted chapters or sections on literature review and methods. The absence of clearly signalled literature review and methods sections led to the urge of selecting those parts of the dissertations which most closely correlate with those of literature review and methods sections in conventional linguistics theses. In most cases, these were the Introduction and the first chapter. It should be noted, however, that some texts in linguistics and education also presented a somewhat unconventional structure lacking the methods section. In addition, the first chapters of the literary studies

dissertations were almost twice as long as the other chapters, which resulted in splitting them in two roughly proportional parts to obtain two separate files from each thesis. The fact that literary studies dissertations were significantly longer also led to the necessity to include only 4 PhD dissertations of this discipline in order to maintain a more or less equal amount of words for each of the disciplines. This was done in order to maintain the corpus homogeneity and have equal representations of each discipline in the analysis, following the principle of *Tertium Comparationis* (Connor & Moreno, 2015). Each chapter or section was saved in a separate file, thus resulting in 28 .txt files from 14 different theses written between 1999 and 2016 and downloaded from online thesis libraries of different universities located in English-speaking countries. See Table 3.2 for a brief description of the subcorpus, and Appendix 1 for details.

Table 3.2. The PhD dissertations subcorpus.

Subject	No of texts	Word count
Education	5	76.100
Linguistics	5	55.500
Literature	4	72.700
Total	14	204.200

The second subcorpus comprises the TFGs written in the previous years by students of the University of La Laguna and downloaded from the online repository of the same university. Its full description is given in Appendix 1. A total of 20 TFGs were selected which amounted to around 240.000 words, with an average length of 12.000 words, the longest text having 16.441 words, and the shortest 7.400. The number of available TFGs was limited, and almost all of them were selected for this analysis,

with the exception of those which could not be converted into .txt files. This subcorpus was used to search the target bundles identified in the PhD subcorpus and contrast their frequency, functions and usage in the writing of the students of ULL.

To address the research questions regarding the discipline-specific lexical bundles in psychology, linguistics, and literary studies research articles, a third corpus was compiled, composed of 60 papers in each discipline from a range of high-impact journals. The journals selected for this corpus were as follows: *Psychological Review*, *Annual Review of Psychology*, *Psychological Bulletin* (for research articles in psychology), *English for Specific Purposes*, *Journal of Pragmatic*, *System* (for research articles in linguistics), *Contemporary Literature*, *Early Modern Literary Studies*, and *Journal of Modern Literature* (for research articles in literary studies). The journals were selected on the basis of their availability online through the virtual library of ULL, and their impact factor according to Elsevier Scopus (Elsevier.com), as well as the number of available issues and their publication dates. All the articles selected for this corpus were published in 2015-2105, as my intention was to collect a corpus as recent as possible in order to perform a state-of-the-art analysis, avoiding any longitudinal elements in this particular corpus. 20 research articles were selected from each journal to include in the Expert corpus, thus totalling 180 articles, 60 in each discipline, and amounting to around 2,1 million words. The journals were selected so as to represent a wide range of topics in each discipline in order to obtain a large amount of topic-neutral bundles. Although the selected subcorpora are relatively small, they are sharply focussed and can yield significant results

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

concerning the phraseology of each discipline (Biber, 2006). Vaughan & Clancy (2013) argue that small, tailor-made, domain-specific corpora are likely to reveal features which would pass unnoticed in large generalised corpora. This view is also supported by Howarth (1996), who uses a corpus of less than 60.000 words to extract restricted collocations in L2 writing. According to Howarth, a small corpus will reflect the most characteristic features of the discourse community without underscoring the importance of overgeneralised features, and besides, it will be more viable for manual analysis. Cortés (2013) points out that “the frequency of individual lexical bundles becomes higher as the corpus becomes more focused or restricted” (p. 42). The corpus in my study is much larger than those used by Howarth (1996) and Cortés (2013), and it was hoped that it would help to reveal the most characteristic lexical bundles in the three disciplines without falling into the extreme of either unnecessary generalisations or too specific, irreproducible results. Table 3.3 describes the size and composition of this corpus.

Table 3.3. Research Articles corpus.

	Journals			RA's	Word count
Psychology	<i>Annual Review of Psychology</i>	<i>Psychological Bulletin</i>	<i>Psychological Review</i>	60	800.000
Linguistics	<i>English for Specific Purposes</i>	<i>Journal of Pragmatics</i>	<i>System</i>	60	800.000
Literature	<i>Early Modern Literary Studies</i>	<i>Contemporary Literature</i>	<i>Journal of Modern Literature</i>	60	550.000
Total				180	2.150.000

The fourth subcorpus includes the TFGs of the students who participated in the workshop on lexical bundles (the TFG'17 subcorpus). This corpus is significantly smaller than expected, comprising only around 75.000 words. Although over 30 students participated in the workshop, only seven of them were able to provide their final texts for analysis in this study. All the seven participants were undergraduate students of applied linguistics or literary studies. It was not possible to obtain any post-workshop text from the PhD students of psychology, though some of them provided their pre-workshop materials and the initial lexical bundle lists extracted from their own mini-corpora during the workshop. One PhD student from the English Philology Department provided pre-workshop and post-workshop texts which were used for a case study, since the account of this student's language evolution offers several insights into the issue.

The principles of corpus analysis are based on comparability and homogeneity of the corpora involved in the study (Connor & Moreno, 2015). Besides being homogeneous in terms of such variables as L1 / L2, expertise, and discipline, the TFG'17 subcorpus is homogeneous in one more way: all the writers share the same L1 and linguistic background, an important parameter according to Granger (1998a). In addition, they all come from the same educational institution, the University of La Laguna.

Thus, the study relies on four different sets of corpora: the Expert corpus with three subcorpora representing psychology, linguistics, and literary studies research

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

articles, the L1 novice corpus of PhD dissertations (the PhD subcorpus), the L2 novice corpus of the final degree research projects produced by students of previous years who had not received any instruction on lexical bundles (the TFG'15-16 subcorpus), and the L2 novice corpus of final degree research projects produced by students who had participated in the workshop on lexical bundles (the TFG'17 subcorpus). The analysis was carried out along the following variables: three different disciplines (psychology, linguistics, literary studies); three levels of expertise (expert L1, novice L1, and novice L2). The L2 writing component is additionally divided into pre-instruction and post-instruction production, represented respectively by the TFG'15-16 and TFG'17 subcorpora.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

3.4. PARTICIPANTS AND WORKSHOPS

All the participants were speakers of Spanish with varying proficiency in English: the linguistics and literature students were all from the Department of English and German Philology and can be described as advanced learners of English. All of them had passed a third year English language course at C1 ALTE level and were currently taking or had completed an academic speaking course taught at C1+ level. Their exposure to academic writing had occurred mainly through textbook reading rather than professional academic writing, though at the time when the workshop was conducted they were preparing to write their TFGs and reading research articles and book chapters on their respective topics. They were taking a writing course which was designed to help them with the writing of their TFGs. The psychology students involved were all at different stages of Doctoral studies, and were mostly familiar with academic texts in their disciplinary area through extensive reading of academic texts, although their writing skills varied greatly from individual to individual. A total of 32 students participated in the workshop: 21 undergraduate students and 3 PhD students from the Department of Philology, and 8 PhD students from the Faculty of Psychology.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The same workshop was conducted for different groups of participants; in some groups, it was conducted in English, and in others, in Spanish due to the presence of students from the Faculty of Psychology who were not as comfortable with English as the Philology Faculty students.

The linguistics and literature students also took part in a short introductory session which involved explanation of the nature and key features of lexical bundles and several activities designed on the basis of vocabulary learning exercises. The handout of the introductory session, as well as the handout of the workshop containing all the activities, can be found in Appendix 4.

The introductory session had a duration of approximately 40 minutes and was structured as follows:

1. The definition and characteristic features of lexical bundles (frequency, semantic transparency).
2. Lexical bundles as important building blocks of discourse, their genre-sensitive and discipline-sensitive character.
3. Examples of sentences containing lexical bundles extracted from authentic data.

The students were offered two sets of sentences, one of them containing the original text extracted from the PhD corpus, while in the other set the sentences were modified and lexical bundles removed from them. Otherwise these sentences presented no irregularities, and the purpose of this activity was to ask the students' opinion about the acceptability of these sentences in academic discourse.

4. Students were shown two lists of lexical bundles and asked to judge their acceptability in academic discourse. One of the lists comprised the bundles which were found to be overused in learner writing, the other contained the most underused bundles.

5. Students were asked to form pairs to work on the exercises designed on the basis of the preliminary research which had revealed the most frequent bundles in the PhD corpus, and which were underused or overused by learners.

After this introductory session, a workshop was conducted for the same students. The participants received a pre-workshop email explaining briefly the essence of lexical bundles and instructing them to download the corpus analysis software, AntConc (Anthony, 2010), as well as the file converter software, AntFileConverter created by the same developer (Anthony, 2015). The students were also asked to prepare their own mini-corpora by choosing and downloading 5-7 articles or book chapters which they considered representative of their field and topic, and as good models for their writing. The pre-workshop email is also included in Appendix 4.

The workshop consisted of a brief explanation of the steps to follow to extract lexical bundles, and a practical session conducted with the students during which each student converted her .pdf files into .txt format and removed the references and other parts which did not belong to the text itself. The frequency and distribution criteria were specified, and each student generated a list of bundles for personal use. The most interesting bundles were discussed and their concordance lines consulted. The

students were also given a brief description of the other functions of the software that could be useful to them.

Feedback was collected both formally, through an online questionnaire, and informally by asking their opinion of the workshop.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

CHAPTER FOUR

LEXICAL BUNDLES IN PSYCHOLOGY, LITERARY STUDIES, AND LINGUISTICS RESEARCH ARTICLES

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.1. INTRODUCTORY REMARKS

This chapter summarises the results of the analysis of three-, four-, and five-word lexical bundles extracted from the corpus of research articles in the three disciplines.

Although using a piece of software to identify the most frequent clusters may seem a relatively straightforward task, it raises several issues which must be resolved before reporting the results. The first issue is what should be considered as a word. For example, the contracted forms *it's*, *isn't*, *don't* have been assigned different status in different analyses. Even though such forms are virtually absent in academic writing, words containing an apostrophe (*Shakespeare's*) are counted as separate words by AntConc (Anthony, 2010), and the two-word collocations such as *et al.'s* must be removed manually. However, there are other arbitrary cases, such as compound nouns which may have three different orthographic forms: the components may be separated by a space, a hyphen, or written as a single word. Furthermore, items like *meta-analysis* are clearly single words, but the case of *twenty-first* or *African-American* is not so straightforward. There are also some proper nouns which have two or more separate components (*the New York City*), phrasal verbs which are generally recognised as single lexical units, but graphically

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

presented as two or more words, etc. Therefore, clear and solid criteria are needed to define the concept of word for a particular study.

In this study, it was decided to consider as a ‘word’ any lexical unit that is identified as such by the software used for the analysis (AntConc: Anthony, 2010), with the exception of the words containing apostrophes. Hyphenated words were all considered to be separate items: thus, both *non-native* and *meta-analysis* are considered collocations of two different words, but *Shakespeare’s* is considered to be a single word. This approach is somewhat mechanical, as it does not consider the content and meaning, but it is suitable for the purposes of the present study, since these hyphenated words / collocations are not very numerous and have little impact on the general tendencies of usage in each discipline.

A second problem is the overlap of different bundles which potentially form one single bundle. For instance, *as well as* occurs 406 times in the psychology subcorpus, and *as well as the* occurs 68 times. It is clear that 68 instances of *as well as* are parts of the larger bundle *as well as the*, but the remaining 348 instances are not followed by the definite article. This means that it is not viable to simply add up the number of three-, four-, and five-word bundles to calculate the total number of tokens in a given subcorpus. A careful examination is needed to combine the three lists of bundles and remove any bundles which are fragments of larger sequences. Another related issue is the occurrence of two different bundles which form a longer bundle, for example, *in the context* and *the context of* combine to form the bundle *in the*

context of, but *the context of* also occurs separately. Counting all the types and tokens rendered by the software may inflate the actual numbers, but many bundles offer interesting insights and should not be discarded, even if they are repeated as fragments of other bundles. This study uses two different lists of lexical bundles: those retrieved by the software where many bundles are repeated as subsumed in longer ones or forming part of other bundles (overlap), and combined lists containing the consolidated versions of different bundles, for instance, *on the other hand* and *on the other hand the* would appear separately in the first list, but only (typically the longer) one would be included in the second list.

Chen and Baker (2010) provide a useful discussion of this issue. They also address the question of types and tokens, i.e., whether we should consider the number of different lexical bundles to be found in a corpus (types) or each occurrence of each type of lexical bundles (token). In other words, if there are 140 lexical bundles in a corpus, and one of them occurs 121 times, shall we look at the 140 different bundles or consider also each one of 121 occurrences of a particular combination?

This study is concerned with identifying general patterns of both usage and frequency, therefore, at different points both types and tokens will be discussed as necessary.

The most important hurdle encountered during analysis was the functional classification of the bundles. As mentioned in Chapter one, the functional taxonomy was first proposed by Biber, Conrad and Cortés (2003) and taken up, improved,

modified and used by many researchers (Chen & Baker, 2010; Hyland, 2008b; Pan, Reppen & Biber, 2016; Salazar, 2011, 2014; Simpson-Vlach & Ellis, 2010, to name but a few). However, some studies have pointed to certain discrepancies in the application of this functional taxonomy by different researchers (Ädel & Erman, 2012; Nesi & Basturkmen, 2006).

For instance, *at the same time* was classified as research-oriented following Hyland (2008b) and Salazar (2014), but the examples below show that it can also perform a discourse organising function, i.e., operate as a text-oriented bundle.

(1) I would like to suggest that the subordination of emotions to appearance must also occur in order to maintain the idea that Penthea and Ithocles differ from the other characters because they were born at the same time.

[Literary Studies]

(2) [...] the poet is bound by the specifics of the text at the same time that the text is made to serve a purpose beyond its author's original intent.

[Literary Studies]

(3) She is ostensibly referring to her doomed love for Orgilus, but at the same time she is also denying her connection to her twin, because he had specifically identified that organ as proof of their togetherness. [Literary

Studies]

While in (1) and arguably (2) *at the same time* serves a more or less clear referential function, in (3) it is most definitely a discourse organiser indicating contrast. The

multifunctionality of lexical bundles has been discussed, among others, by Salazar (2014), Biber, Conrad and Cortés (2004), and by Coxhead and Byrd (2010). The solution proposed by these researchers is to classify the bundles by their most common function, but the less common functions are not regarded as less important (Coxhead & Byrd, 2010).

Emphasising the most frequent, primary function may be pedagogically useful, though the learners should be warned against relying too heavily on the primary functions and taking the classification dogmatically. However, in some cases, it is difficult to determine the primary function of a bundle, as is the case with *(at) the end (of)*. Depending on the context, it can indicate time, place, process, or perform a text-deictic function (Salazar, 2014). Below are some examples to illustrate the different functions of this bundle in linguistics and literary studies research articles:

- (4) Inspired by Operation Babylift, the militarized evacuation and transnational adoption that sent Vietnamese orphans to the United States toward the end of the Vietnam War, Phan’s collection follows stories of young adult adoptees and refugees in Orange County, California’s Little Saigon and of people who cared for orphaned children during the war. [Literary Studies]
- (5) In order to determine the effect of a number of factors on PV knowledge, a biodata questionnaire was included at the end of the test. [Linguistics]

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

- (6) When Mai returns home at the end of the day, she discovers a white envelope from Wellesley College, placed next to the birthday presents from her foster parents, whose obligation to shelter her has come to an end. [Literary Studies]
- (7) Seven students in Peter's classes volunteered to take part in a semi-structured interview about their ER experiences and their reading motivation at the end of the semester. [Linguistics]
- (8) **Near the end of the narrative**, he reverts to this state of linguistic play, articulating an ecstatic stream of nonsensical syllables: “it’s hot,” “it’s lot,” “it’s not,” “it’s top” (161). [Literary Studies]
- (9) In neither sample was there a discernible pattern for the distribution or particular clusterings of such statements in certain parts of the essay, such as at the end of sub-topics or overall essay conclusions. [Linguistics]
- (10) When, by the end of the novel, Austerlitz’s travels take him to Germany, he is struck by what appears to him as the erasure of the physical traces of its past. [Literary Studies]
- (11) At the end of the long second paragraph discussed above, we hear only that “she was going to go away like the others, to leave her home” (37). [Literary Studies]

Even in this reduced sample, there seem to be instances of multifunctionality of (*at the end (of)*), for example, in (4), depending on whether we think of “the end of the Vietnam War” as a particular moment in time or of a stage in a process (its end), or

as in (5), where “the end of the test” can be viewed as place or time. In the last example, the same bundle seems to perform a text-deictic function, although it could also indicate place.

In order to determine the primary function of a lexical bundle, concordance lines were consulted in many dubious cases, but there remain arbitrary instances. For example, *as a result of* is classified as ‘text-oriented resultative signal’ in Salazar (2011), but as ‘referential intangible’, i.e., research-oriented in Cortés (2008). Below are the concordance lines for this bundle in the linguistics research articles.

Figure 4.1. Concordance lines for *as a result of* in Linguistics research articles.

1 Europe (see Wächter & Maiworm, 2014), possibly	as a result of	the lower levels
2 student lecture comprehension problems in EMI	as a result of	lecturers’ weak pronunciation
3 participate less dynamically in the classroom,	as a result of	their limited oral
4 of impact, if any, professionals perceive	as a result of	their use in
5 pseudo-individualistic . . .” (Sociology Text 8)	“As a result of	this financially selective
6 ‘professional knowledge and skills is progressed	as a result of	EMP instruction. The
7 their disciplinary identity and share	as a result of	collectively carrying out
8 continues to evolve in the language	as a result of	the use people
9 are the initiatives patients take	as a result of	sickness experiences or
10 In a number of instances, sometimes	as a result of	fatigue, as supported
11 comparisons with other research. Fifth,	As a result of	the methodological improvement
12 in subsequent sentences as a topic.	as a result of	our representation of
13 are always power interactions that operate	as a result of	the social structure,
14 a peripheral to a primary role	as a result of	cross-modal exchanges.
15 WTC and fluency. Her fluency declined	as a result of	the reduced WTC,
16 and often the only option’ (p. 7).	As a result of	the widespread use
17 English or something like that. (#7, male)	As a result of	the positive connotations
18 reason: avoidance of native speaker associations	As a result of	the perceived relationship
19 to say what they are thinking.	As a result of	their teachers’ openness
20 episodes have been reported to occur	as a result of	not only morphosyntax, but

The concordance lines above show that *as a result of* is used to indicate a cause-consequence relationship, i.e., it clearly has a discourse organising function. On the other hand, it also deals with actual facts of the real world, which is how research-oriented bundles are described (Hyland, 2008a, b; Salazar, 2014). The attribution of

one label or another ultimately seems to depend on the perspective adopted by the researcher, therefore, it is important to specify clear criteria and follow them throughout analysis. However, the perspective adopted by the researcher will have considerable impact on the results of the study, which implies a great deal of responsibility. In this case, *as a result of* was labelled as text-oriented (discourse organising), but the above-mentioned caveats should be taken into account when interpreting the results of this study.

As can be seen from the examples above, the functional classification of lexical bundles in many cases may depend on the subjective judgment and intuition of the analyst. Moreover, minute classification of lexical bundles into the subcategories would have little pedagogical value, since most of the bundles perform many different functions, and those which have only one function (e.g., *the number of* virtually always indicates quantification) are straightforward enough for advanced users and are not likely to present problems. However, a classification into the major categories would be pedagogically useful, as it would provide an overall picture of the usage patterns and the genre conventions of a particular discipline. For instance, knowing that 9% of all the bundles in psychology are participant-oriented, a novice researcher could adjust her use of participant-oriented bundles in order to avoid sounding too informal or too distant in accordance with the norms of her discipline.

In this study, I attempted to classify the bundles according to the three broad categories, namely text-oriented, research-oriented, and participant-oriented bundles,

but did not use the numerous subcategories proposed by different researchers (Biber, Conrad & Cortés, 2003, 2004; Hyland, 2008a, b; Simpson-Vlach & Ellis, 2010). This approach was based on the above-mentioned considerations of the potential pedagogical value of classification into subcategories, as well as the arbitrariness of many instances encountered in the corpus. The functional classification was performed based on the concordance lines, broader contexts (whole sentences and paragraphs), and the classifications proposed in other studies.

This section outlined the general principles of analysis of lexical bundles and addressed several hurdles related to it. The following sections report the results obtained for lexical bundles in psychology, literary studies, and linguistics research articles.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.2. LEXICAL BUNDLES IN PSYCHOLOGY RESEARCH ARTICLES

4.2.1 The frequency of lexical bundles in psychology research articles

The search for three-, four-, and five-word lexical bundles in psychology research articles rendered the largest number of lexical bundles in the whole corpus: 126 three-word bundles, 115 four-word bundles, and 13 five-word bundles totalling 254 bundles. It is important to bear in mind that many of the three-word bundles are parts of four- or five-word sequences, therefore, the total number of the bundles is inferior to the sum of the figures indicated above. After removing the redundant bundle fragments (cases where a bundle was completely subsumed within another), the list was reduced to 172 lexical bundles of three, four, and five words. The complete list of all the bundles, as well as the combined one, can be found in Appendix 3.

Forty-seven bundles in the psychology subcorpus occur more than 100 times per million words, most of these being three-word bundles. This is not surprising given the extremely high frequency of these bundles in other corpora. Only 3 four-word bundles made it to the top, two of which are longer versions of the corresponding

three-word bundles, *the other hand* and *a function of*. These bundles are subsumed by the four-word bundles *on the other hand* and *as a function of* respectively, and the three-word bundles were deleted from the list as their occurrence rate was identical to that of the four-word bundles. The third bundle is a case of overlap: *the context of* occurs as part of *in the context of* in 89 instances, but *the context of* occurs also separately, which is why both were kept in the list. Table 4.1 shows these bundles and their frequencies.

Table 4.1. The most frequent lexical bundles in psychology research articles.

No	Bundle	Freq.	p.m.w.	No	Bundle	Freq.	p.m.w.
1	as well as	406	507	25	is that the	97	121
2	the number of	259	324	26	the context of	96	120
3	the effects of	215	269	27	is associated with	94	117
4	in terms of	194	242	28	on the other hand	94	117
5	one of the	146	182	29	of the model	91	114
6	in which the	145	181	30	in the context	90	112
7	such as the	145	181	31	as a function of	89	111
8	more likely to	141	176	32	the idea that	89	111
9	a number of	137	171	33	as a result	88	110
10	consistent with the	134	167	34	is consistent with	87	109
11	the development of	124	155	35	the importance of	87	109
12	the fact that	123	154	36	that it is	86	107
13	a function of	121	151	37	a set of	85	106
14	in order to	121	151	38	in addition to	85	106
15	the role of	120	150	39	it is not	85	106
16	the effect of	119	149	40	a variety of	84	105
17	based on the	118	147	41	likely to be	84	105
18	on the other	117	146	42	in the context of	83	104
19	the presence of	115	144	43	individual differences in	83	104
20	with respect to	112	140	44	it is possible	82	102
21	the relationship between	104	130	45	each of the	81	101
22	according to the	101	126	46	in response to	81	101
23	the absence of	101	126	47	to account for	80	100
24	there is a	101	126				

Psychology research articles use the largest number of lexical bundles, both in types and tokens, therefore, many bundles are found only in the psychology subcorpus, as will be discussed later in this chapter.

The lexical bundle *par excellence* is *as well as*, occurring 406 times in a corpus of 800.000 words, i.e., approximately 507 times per million words, or approximately 6 times per ten thousand words. It is present in 52 out of the 60 texts in Psychology as a bridging item between two elements. In this subcorpus *as well as* occurs in virtually any position and colligates with syntactic units of all kinds, connecting nouns, adjectives and even entire clauses:

- (12) Acute as well as chronic stress can increase people’s reliance on habits.
- (13) Through procedural memory, habits could be cognitively represented as distinct from other types of implicit processes as well as from explicit, declarative memories.
- (14) Componential explanations of automaticity specify non/automaticity features such as un/conscious, un/intentional, non/efficient, and fast/slow as well as their interrelations.
- (15) Necessary conditions include both control over the administration of independent variables (e.g., performance or relational-based task) and assessment of dependent variables (e.g., preference for a type of justice

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

principle) as well as over the influence of possible confounding variables that threaten the validity of causal inferences.

(16) Adhering to these principles during the implementation phase bolsters the chances for compliance with the agreements as well as dealing with those actors intent on spoiling them.

(17) But to tackle cross-culturally recurrent elements of religion, as well as plumb the depths of theodiversity, psychology must cast a wider net to capture the full range of human cultural diversity.

This bundle occurs almost twice as often as the next most frequent bundle, *the number of*, with a raw frequency of 259 (324 per million words).

The psychology corpus contains several bundles which are almost meaningless, such as *et al a*, *et al the*, *et al and*, *et al in*, and *et al these*. These were discarded as uninformative, but their extremely high frequency (178, 146, 256, 184, and 61 respectively) leads to the conclusion that they are not altogether irrelevant. Moreover, their occurrence exclusively in the psychology subcorpus suggests that they might be specific only to this discipline. Obviously, these bundles are used in citations and contain punctuation marks and digits which the software does not reflect, but their absence in the other subcorpora is remarkable. Some examples are presented below:

(18) However, finite mixture models, like the ones proposed by Zhang and Luck (2008), Bays et al. (2009, 2011), and others can also be represented using the 2D diffusion model. In finite mixture models the predicted

distribution of report is a mixture of a small number (often two) of von Mises distributions.

(19) Recognition and source memory are enhanced by positive compared with negative feedback (Eppinger et al. 2010, Mather & Schoeke 2011) and by reward anticipation (Spaniol et al. 2014) in older adults as much as in younger adults.

(20) Although some studies have included a ‘mixed-percept’ response option to address this issue (Alpers & Gerdes, 2007; Lerner et al., 2012), the boundary between perception of one image and another in rivalry is often graded and temporally uncertain (Knapen et al., 2011).

From the above examples it is clear that these sequences in fact represent patterns of citation which are typical of psychology research articles and do not occur in the other subcorpora. This finding seems to suggest that psychology uses more citations than the other disciplines, possibly as a strategy of making the text more objective (Hyland, 2008b), or claiming centrality (Swales, 1990). However, it was decided to remove these sequences from the list as they do not represent actual bundles but rather discontinuous patterns (Biber, 2009).

As for the four-word bundles, fifteen of them occur more than 50 times per million words in the subcorpus. They are presented in Table 4.2.

Table 4.2. The most frequent four-word bundles in Psychology research articles.

No	Bundle	Freq.	p.m.w.	No	Bundle	Freq.	p.m.w.
1	on the other hand	94	117	9	the size of the	58	72
2	as a function of	89	111	10	it is possible that	56	70
3	in the context of	83	104	11	are more likely to	50	62
4	on the basis of	69	86	12	in terms of the	46	57
5	as well as the	68	85	13	it is important to	45	56
6	the extent to which	68	85	14	as a result of	41	51
7	in the absence of	63	79	15	at the same time	41	51
8	for a review see	59	74				

Although the vast majority of the four-word bundles are nominal or prepositional, as would be expected in academic discourse, we also find four verb phrase fragments, all of them representing participant-oriented or stance bundles: *for a review see*, *it is possible that*, *are more likely to*, and *it is important to*. Examples of these follow:

- (21) [...] the existing Stroop models predicted a lack of such an effect (e.g., Cohen et al., 1990, Figure 5; Herd et al., 2006, Figure 4; Lovett, 2001, Figure 1; Roelofs, 2003, Figure 13; van Maanen et al., 2009, Figure 1; for one exception see Phaf et al., 1990), the existence of a small but significant reverse effect has been noted in more recent studies (for a review see Blais & Besner, 2006).
- (22) It is possible that the standard two-stage RL decision task does not capture the process that produces outcome-insensitive habits,
- (23) This reciprocal cycle of maladaptive social transactions may help to explain why dispositionally negative individuals experience lower levels

of occupational, financial, and marital success and are more likely to develop internalizing disorders.

- (24) It is important to note that although the RSDH predicts more pronounced CI in larger groups, no study to date has measured the effect in groups larger than four.

It is possible that and *it is important to* are also among the few lexical bundles which may occur in sentence-initial position, although this is not their most frequent position.

Another citation pattern, *et al found that*, had made it to the top showing a frequency of 59 and occurring in over 30% of all the texts in psychology, but it was ultimately removed from the list due to reasons specified above.

Nevertheless, the pervasiveness of ‘*et al*’ patterns is noteworthy, as two of them appeared even among the very limited number of five-word bundles. The search yielded thirteen five-word bundles in total, of which two were of this type: *et al as well as*, occurring 16 times in 10 different texts, and *et al found that the*, occurring 15 times in 10 different texts. These were also removed from the list as not corresponding to the focus of the present study, although it would be interesting to conduct a separate study on citation patterns in different disciplines.

The final list of five-word lexical bundles in psychology is presented in Table 4.3 below.

Almost half of the five-word bundles are verb phrase fragments, and one is a *that*-clause fragment (*consistent with the idea that*), which seems to contradict previous findings (Biber, 2006; Cortés, 2004, 2008) regarding the phrasal nature of academic bundles. However, it should be borne in mind that most of the previous studies addressed four-word bundles which actually are mostly phrasal. Besides, it is not surprising that 5 out of 55 words (11 five-word bundles, $11 \times 5 = 55$) are verbs, and given the overall number of words in these bundles, the proportion of verbs is more than moderate. Furthermore, most of these are the inflected form of *to be*, a ‘weak’ verb characteristic of academic writing (Biber, 2006; Biber & Barnieri, 2007; Biber & Gray, 2010), and the rest are the modal verbs *should* and *can*.

Table 4.3. Five-word lexical bundles in psychology research articles.

Bundle	Freq	p.m.w	Bundle	Freq	p.m.w
as a function of the	22	27	on the other hand the	15	19
it is possible that the	18	22	it is important to note	14	17
on the basis of the	17	21	it should be noted that	14	17
at the end of the	15	19	a risk factor for depression	12	15
consistent with the idea that	15	19	is important to note that	12	15
it can be seen that	15	19			

From the above table it can be seen that two of the five-word bundles show the above-mentioned phenomenon of overlap: *it is important to note* and *is important to note that* occur 12 times together as *it is important to note that*, and *it is important to note* occurs without *that* only twice. A closer examination, however, showed that *that* is

present even in these two cases, although separated from the first words of the sequence by an intercepting *however* in one case and *already at this point* in the other.

Examples are presented below:

(25) We use (as others have used) the terms race and ethnicity somewhat interchangeably in this article, but it is important to note that these concepts are not always interchangeable.

(26) It is important to note, however, that live musical performance does contain significant timing fluctuations and more complex temporal patterns, closer to the free-flowing rhythms of speech

(27) It is important to note already at this point that this categorization (or identification) task entails divided attention:

It is also interesting that many five-word bundles are participant-oriented (stance) bundles. These will be discussed in the following sections.

4.2.2. The structure of lexical bundles in psychology research articles

The analysis of the structures of lexical bundles confirmed their primarily phrasal character: the overwhelming majority of all the bundles were noun phrase or prepositional phrase fragments, an inherent feature of academic writing (Ädel & Erman, 2012; Biber, Conrad & Cortés, 2004; Chen & Baker, 2010). These are described in Table 4.4.

Table 4.4. Structural classification of Lexical bundles in psychology.

Structure	Examples
Adjective + noun phrase	a recent meta-analysis, ventromedial prefrontal cortex
Noun phrase with <i>of</i> -phrase fragment	a number of, a great deal of, a (wide) range of, a set of, a variety of, activation of the, all of the, most of the, number of studies, of the effects of, of the number of, of the stimulus, of the target, of the task, of the two, of this figure, one of the most, our understanding of the, part of the, some of the, the absence of, the absence of a, the activation of the, the amount of, the basis of the, the comparison of, the distribution of, the effect(s) of, the end of the, the impact of, the importance of, the influence of, the level of, the location of the, the magnitude of the, of the model, the nature of the, the presence of a, the processing of, the proportion of, the results of the, the role of the, the set of, the shape of the, the size of the, the strength of the, the type of, the use of, the value of the
Noun phrase with other post-modifier fragments	a risk factor for, an important role in, individual differences in, the association between, the degree to which, the difference between the, the distance between the, the extent to which, the fact that the, the idea that, the relationship between, support for the, the ability to, with the idea that, with respect to the
Prepositional phrase with embedded <i>of</i> -phrase fragment	as a function of (the), as a result of, at the end of, at the level of, in the context of, at the time of, for each of the, in a study of, in a variety of, in terms of the, in the absence of, in the case of, in the face of, in the form of, in the number of, in the presence of, on the basis of the, over the course of, to the development of

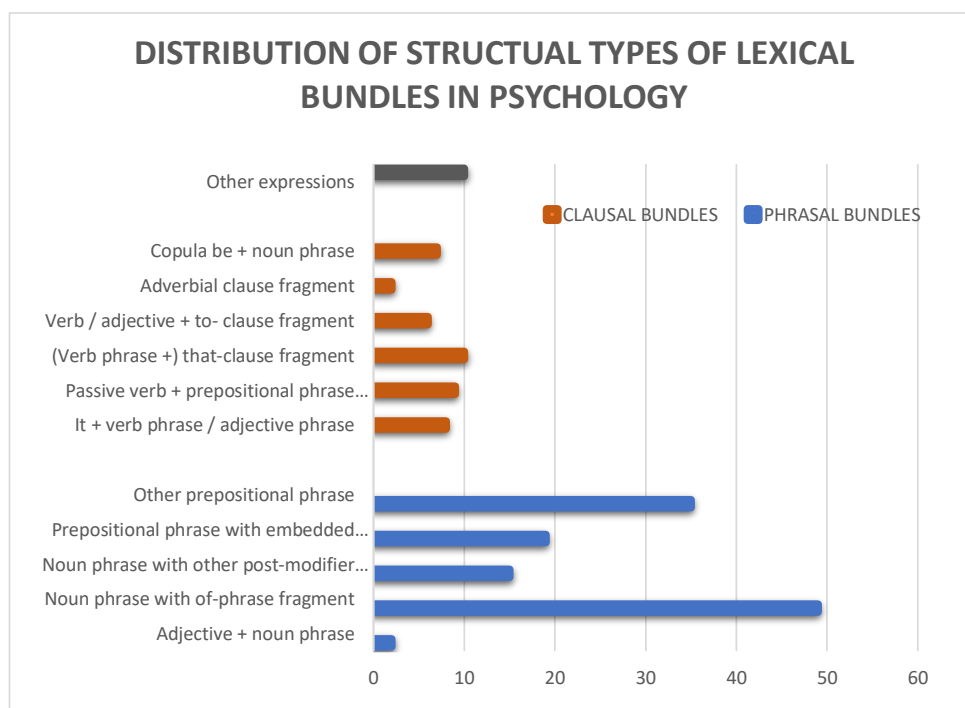
Other prepositional phrase	at the same time, according to the, associated with the, between the two, differences in the, consistent with the idea, for a review see, for example the, from trial to trial, in a way that, in addition to the, in contrast to the, in line with the, in order to, in other words, in response to, in the literature, in the meta-analysis, in the next section, in the present, in the previous section, in the same, in the sense that, in the United States, in this article, in this meta-analysis, in this section we, in which the, of individual differences in, of the variance in, on the one hand, on the other hand, risk factor for depression, to the extent that, to the fact that
<i>It</i> + verb phrase / adjective phrase	it can be seen, it is important to, it is likely that, it is not, it is possible that (the), it is possible to, it may be that, it should be noted
Passive verb + prepositional phrase fragment	can be found in, has been shown to, have been found to, have been shown to, have been used to, included in the, is associated with, is shown in figure, related to the
(Verb phrase +) <i>that</i> -clause fragment	and colleagues found that, is that it is, studies have shown that, suggests that the. that can be, that there is a, that it is, there is evidence that, these findings suggest that, we found that the
Verb / adjective + <i>to</i> - clause fragment	are likely to be, are more likely to, can be used to, is likely to be, is more likely to, more likely to be, were more likely to
Adverbial clause fragment	as shown in figure, as the number of
Copula <i>be</i> + noun phrase	are consistent with the, is based on the, is consistent with the, is possible that the, to be a, that is the, there is no
Other expressions	as long as the, as well as the, as well as to, depends on the, effect sizes were, may not be, play a role in, such as the, to account for the, to respond to

During the structural analysis, a new category emerged, that of adjective / noun + noun phrase. Although in the psychology subcorpus it is represented by only two bundles, the other two disciplines also contained several examples of this category. These will be discussed in the following sections. The two noun phrase bundles in the psychology subcorpus, *a recent meta-analysis* and *ventromedial prefrontal cortex* are both topic-specific, and the same can be said of the nominal bundles found in the literary studies subcorpus. Most studies of lexical bundles discard topic-specific

bundles, but they may present interesting insights, especially in case of small corpora designed for and by learners (see Chapter 5 for a detailed discussion).

Figure 4.2 represents the proportions of different structural types found in the psychology research articles. For this analysis, the combined list of bundles was used where shorter bundles subsumed in longer ones were not included, and several cases of overlap were also combined to form a unified bundle. This made it possible to avoid inflating the numbers and allowed a more accurate picture of the distribution of different bundle types to emerge.

Figure 4.2. The distribution of structural types of lexical bundles in psychology research articles.



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Below are examples of each of these bundle types in context.

- (28) Another factor is that our understanding of the uses and effects of media and communication technologies develop in a variety of disparate disciplines and subdisciplines
- (29) The strength of this inhibition is inversely proportional to the distance between the options in the feature space.
- (30) Several themes are discussed in this section in terms of their relevance to the prenegotiation stage.
- (31) Although some studies have investigated materialism in children and in less economically developed nations, such studies are underrepresented in the literature.
- (32) Rather, it may be that this processing can occur, but is not deployed in situations that are highly aversive such as the original helplessness experiments.
- (33) People who experience stereotype threat have been shown to discount the validity of a high-threat test
- (34) We limited the maximum number of digits constituting a multisymbol number in our simulations to three digits, because there is evidence that the number of digits that can be processed in parallel in working memory seems to be limited.

- (35) Participants in the imagined unpaid prosocial behavior condition were more likely to later keep all the money for themselves relative to a control condition
- (36) Finally, it is possible that the Identity factor also predicts externalizing disorders, but in the opposite direction of the Trait factor (i.e., a positive path for the Reputation factor but a negative path for the Trait factor).

As can be seen from Figure 4.1, the noun phrase and prepositional phrase structures by far outnumber the verb phrase bundles and other expressions, clearly reflecting the phrasal nature of academic writing (Biber & Gray, 2010; Cortés, 2008, Salazar, 2014, etc.). The overall number of ‘phrasal’ bundle types in this subcorpus is 118, as opposed to the 43 verb phrase bundle tokens. The ‘phrasal’ bundles are also among the most frequently occurring, as shown in Appendix 3.

The anticipatory *it*+ verb phrase / adjective phrase category consists of participant-oriented bundles only. This finding seems to support the view that there is some degree of correlation between structural form and function. The functions of lexical bundles in psychology research articles are described below.

4.2.3. The functions of lexical bundles in psychology research articles

As was explained in Chapter one, the functional classification of lexical bundles is somewhat arbitrary (Ädel & Erman, 2012), as besides the fact that many lexical bundles perform different functions, classification is also highly dependent on the subjective perceptions of the researcher and the purposes of the study.

The overall number of the bundles in the combined list is 172, of which 56% are research-oriented bundles, 35% are text-oriented bundles, and 9% are participant-oriented. As will be shown in the following sections, 9% is a relatively high proportion for participant-oriented bundles.

The bundle types and their functional categories are shown in Table 4.5 below.

Table 4.5. Functional classification of lexical bundles in Psychology research articles.

Category	Examples
Text-oriented	according to the, and colleagues found that, as a function of (the), are consistent with the, as a result of, as shown in figure, of this figure, on the basis of, as well as the, as well as to, can be found in, can be used to, consistent with the idea, for example the, has been shown to, have been found to, have been shown to, in addition to the, in contrast to the, in line with the, in other words, in response to, in terms of the, in the case of, in the literature, in the next section, in the present, in the previous section, in the sense that, in this article, in this section we, is consistent with the,

	is shown in figure, is that it is, it can be seen, of the two, on the basis of (the), on the one hand, on the other hand, related to the, such as the, suggests that the, support for the, studies have shown that, that it is, the fact that the, the idea that, the results of the, there is evidence that, these findings suggest that, to account for the, to the fact that, we found that the, with respect to the, with the idea that
Research-oriented	a number of, a great deal of, a recent meta-analysis, a (wide) range of, a set of, a variety of, activation of the, all of the, an important role in, as long as the, as the number of, at the same time, between the two, depends on the, differences in the, effect sizes were, in the context of, a risk factor for, at the end of, at the level of, at the time of, for each of the, from trial to trial, have been used to, in a study of, in a variety of, in a way that, in order to, in the absence of, in the face of, in the form of, in the meta-analysis, in the number of, in the presence of, in the same, in the united states, in this meta-analysis, in which the, is associated with the, is based on the, included in the, individual differences in, it is not, may not be, most of the, of the effects of, number of studies, of individual differences in, of the model, of the number of, of the stimulus, of the target, of the task, of the variance in, one of the most, over the course of, play a role in, part of the, risk factor for depression, some of the, that can be, that there is a, the ability to, the absence of, the absence of a, the amount of, the activation of the, the association between, the basis of the, the comparison of, the degree to which, the difference between the, the distance between the, the distribution of, the effect(s) of, the end of the, the extent to which, the impact of, the importance of, the influence of, the level of, the location of the, the magnitude of the, the nature of the, the presence of a, the processing of, the proportion of, the relationship between, the role of the, the set of, the shape of the, the size of the, the strength of the, the type of, the use of, the value of the, there is no, to be a, to the development of, to the extent that, to respond to, ventromedial prefrontal cortex
Participant-oriented	are likely to be, are more likely to, for a review see, is likely to be, is more likely to, is possible that the, it is important to, it is likely that, it is possible that (the), it is possible to, it may be that, it should be noted, more likely to be, our understanding of the, were more likely to

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

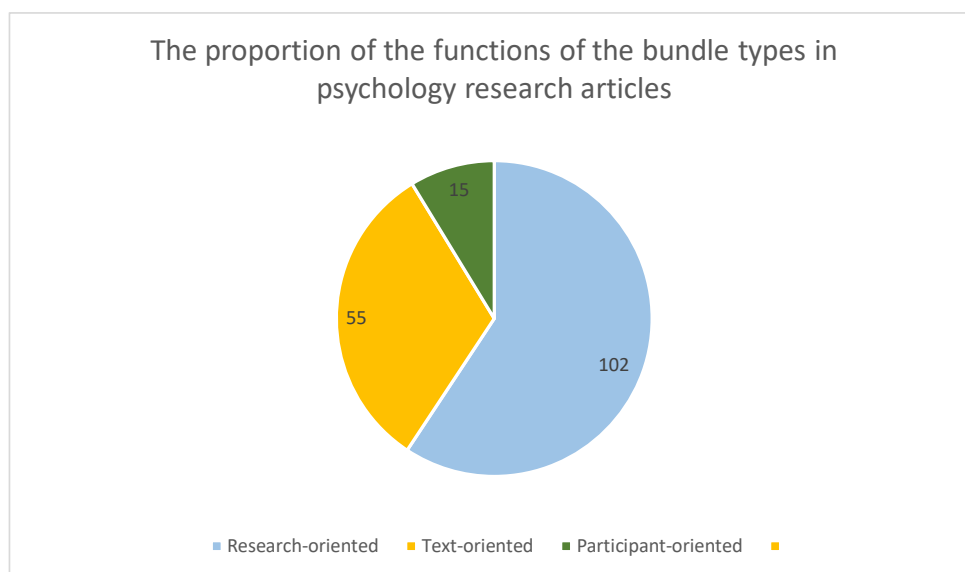
30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The functional analysis of the combined list of lexical bundles in the psychology corpus revealed 55 text-oriented bundles, 102 research-oriented bundles, and 15 participant-oriented bundles, as shown in Figure 4.3.

Figure 4.3. The number of types of lexical bundles for each function in psychology research articles.



Although many bundles display rather straightforward functions deducible from their meanings, many others must be considered in context in order to determine their function. Below are examples of each of the functions in context.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Text-oriented

Text-oriented or referential bundles are used to organise the text and guide the reader through it:

- (37) In addition to the powerful prosocial effects of some forms of rituals, Buddhist contemplative practices of loving kindness may also play a measurable role in well-being, social support, and connection with others
- (38) These strategies fall into three main categories that vary in their commonness as a function of the behavior targeted.
- (39) Longitudinal studies provide the most compelling support for the theory.

Research-oriented

Research-oriented or referential bundles are concerned with activities and experiences of the writer in the real world:

- (40) Effect sizes were coded as nonstressed for subsamples with relatively low stress in comparison to a high stress group in the same study or studies focused on a general population
- (41) Neuroimaging studies of motor sequence learning further confirmed the role of the sensorimotor striatum in chunk formation, along with a frontoparietal network and the mediotemporal lobes
- (42) The results indicated that the activation of the insula and putamen was increased in men, but not in women.

- (43) Participants in this study recorded diaries documenting the extent to which they felt as if they had been transgressed against on a daily basis.

Participant-oriented

Participant-oriented or stance bundles are focussed on the reader or the writer:

- (44) The article closes by demonstrating how the three theories relate to each other and contribute to our understanding of prejudice and its reduction.

As it was mentioned above, the function of many bundles can only be determined in context. Thus, even though *support for the* was classified as text-oriented, in some contexts it functions as a referential (research-oriented) bundle. Concordance lines are especially useful for determining the actual function of each bundle. Figure 4.4 is an illustration of this. The instances of research-oriented function of the bundle are marked by ‘R’, and text-oriented function is marked by ‘T’.

Figure 4.4. Concordance lines for support for the.

<p>R c) cooperation between groups, and (d) authority</p> <p>T Longitudinal studies provide the most compelling</p> <p>T such, this finding in fact provides further</p> <p>T native possibilities. The discovery provides some</p> <p>T to implant such false beliefs, providing strong</p> <p>T the session. Behavioral studies provide further</p> <p>T compensate each other. Some studies provide</p> <p>R there is attitudinal support but not behavioral</p> <p>T review studies that have been used as</p> <p>T information theory’s T) in providing converging</p>	<p>support for the contact. Allport approved of my</p> <p>support for the theory (Binder et al. 2009, Christ</p> <p>support for the scene construction theory, which</p> <p>support for the idea that a counterfactual is</p> <p>support for the mentalistic view that an abstract</p> <p>support for the notion that older adults engage</p> <p>support for the additivity of different types of</p> <p>support for the goal of the intervention will</p> <p>support for the argument that stereotype threat do</p> <p>support for the theoretical construct of the</p>
---	--

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. Su autenticidad puede ser contrastada en la siguiente dirección https://sede.ull.es/validacion/	
Identificador del documento: 973669	Código de verificación: mnf6zN2e
Firmado por: DIANA BALASANYAN UNIVERSIDAD DE LA LAGUNA	Fecha: 30/06/2017 14:44:14
SALLY FRANCES BURGESS . UNIVERSIDAD DE LA LAGUNA	30/06/2017 21:06:24
ERNESTO PEREDA DE PABLO UNIVERSIDAD DE LA LAGUNA	14/07/2017 13:44:17

T to identification of such stimuli, he found
 T in Psychological Bulletin. We are pleased that
 T as well as emotion-related measures. Moreover,
 T in the punishment-only task that provides
 T maze task ($r // .313$) provides even stronger
 T though these findings provide robust differential
 T the low fear model. Reflecting concerns about
 T threat processing in an attempt to boost
 T the beneficial effect of collaboration, providing
 R one occurrence of C. Instead of providing
 R Figure 1. One of these, GEN, reflected direct
 R measured beliefs and attitudes that predicted
 R guy” not only triggered a reversal in
 T for cognition suggests further (albeit indirect)
 T operative). In particular, there is equivocal
 T & Buhrmester, 1992). However, there is growing
 T friend support was significant, demonstrating
 T with nonstressed comparison group would offer
 T SB model; no significant difference would offer
 T support (see O/A in Table 12) demonstrated
 T The current study demonstrated consistent
 T childhood and adolescence. Consistent
 T Overall, these analyses provided continued
 R surprising finding is the general lack of
 R on putative susceptibility markers, this provides
 R with negative child adjustment, this provides
 T with positive child adjustment, this provides
 T in the following sections. However, the
 T to A than B provides somewhat ambiguous
 T threat-related biases. The data provide indirect
 T (PTSD) and problematic alcohol use showed
 T that depression predicts PE would provide
 T allowed us to compare the extent of
 T would predict PE, thereby obtaining stronger
 T increased. This unifying result provides further
 T also expands subjective time – providing further

support for the Euclidian metric. Despite the
 support for the RMH was found across diverse
 support for the RMH spanned male and female
 support for the utility of the RMH. This
 support for the predicted passive avoidance
 support for the RMH versus low fear model,
 support for the overall low fear model,
 support for the low fear model. Even so,
 support for the idea that reexposure to
 support for the right hemisphere’s dominance in
 support for the U.S. entering a war,
 support for the war. For three of these,
 support for the war, but also led to
 support for the idea that entitlement is associate
 support for the importance of social support in
 support for the idea that the importance of
 support for the GB model for each support
 support for the SB model; no significant differences
 support for the GB model (see Table 12 for
 support for the GB theory: The association between
 support for the general benefits (GB) model, but
 support for the general benefits model was also
 support for the GB model, but further exploration
 support for the inclusion of parent/family involve
 support for the differential susceptibility hypoth
 support for the diathesis-stress model. If
 support for the vantage sensitivity model. The
 support for the predictiveness principle offered b
 support for the predictiveness principle over the
 support for the idea that stimulus visibility mode
 support for the sleep dimensional hypothesis.
 support for the scar and complication models.
 support for the scar/complication models versus
 support for the predisposition and pathoplasty
 support for the processing principle: attending
 support for the processing principle. These

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

T Regardless, this literature provides strong **support for the** idea that gesture functions, in
T . . . This principle of learning provides empirical **support for the** efficacy of lifelong access to
T presented. Research studies have also found **support for the** association between agitation
T dynamics and adaptation effects provided strong **support for the** integrated utility-based model,
R the optical variable θ provides informational **support for the** perception of the severity of
T attitude strength. Then we discuss some empirical **support for the** proposition that attitude strength

Figure 4.4 shows several instances of the bundle *support for the* performing a research-oriented function, though the overwhelming majority of the uses of this bundle are text-oriented. This example illustrates the usefulness of concordance lines for advanced learners: their use in the classroom or in the process of writing can offer many insights.

Below is a discussion of the lexical bundles found exclusively in the psychology subcorpus.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.2.4. Lexical bundles exclusive to psychology research articles

A comparative analysis of lexical bundles in the three disciplines revealed 135 bundles exclusive to psychology. An abridged list of lexical bundles found only in psychology research articles is shown in Table 4.6. This list includes the “prototypical bundles” (Salazar, 2014), i.e., those which resulted from the combination of two or more very similar bundles. For example, *is consistent with*, *is consistent with the*, *is consistent with the idea*, *consistent with the*, *consistent with the idea*, *consistent with the idea that* are all varieties of *is consistent with the idea that*, therefore, this bundle was the only one to be included in the list.

The label ‘exclusive’ in this context does not mean that none of the bundles on the list below occurs in the other disciplines. Rather, their rate of occurrence was not high enough in literary studies and linguistics, and they had not appeared in the separate search results of the other disciplines. They did, however, occur in the final search across the three subcorpora. The combined search revealed several bundles which had not appeared in the linguistics and literary studies subcorpora because their frequency in these corpora was lower than specified in the search parameters, but in the combined corpus where the most part of the occurrences corresponded to psychology research articles, these bundles, occurring only a few times in the other disciplines, form a part of the general list. However, every effort has been made to

remove the bundles which occurred less than 4 times in one of the disciplines during the combined search. This issue will be discussed in subsequent sections.

The number of lexical bundles which occur only in the psychology subcorpus is noteworthy. Psychology has the highest number of bundle types of all the three disciplines, and it would be reasonable to expect to find many unique bundles in this discipline. Even with that expectation, we are confronted by a surprisingly high number of unique bundles: 135 bundles out of the total 254 (53% of the bundles of the initial list). Many of these bundles are topic-specific, as was to be expected, though others are neutral sequences which could be intuitively predicted to occur in any discipline. Table 4.6. covers the lexical bundles unique for psychology.

Table 4.6. Lexical bundles unique to psychology research articles.

Bundle	Freq.	Bundle	Freq.
a function of the	37	it may be that	16
a recent meta-analysis	17	it should be noted (that)	21
(a) risk factor for (depression)	27	may not be	51
a set of	85	most of the	57
(in) a variety of	84	number of studies	51
according to the	101	of individual differences in	22
activation of the	67	of the effects of	19
all of the	71	of the model	91
an important role in	16	of the number of	20
and colleagues found that	19	of the stimulus	60
are consistent with the	19	of the target	65
are (more) likely to (be)	20	of the task	51
as a function of (the)	89	of the two	70
as long as the	16	of the variance in	19
as shown in figure	25	our understanding of the	19
as the number of	17	play a role in	16
between the two	56	related to the	74
can be found in	27	studies have shown that	16

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

can be used to	33	suggests that the	56
(is) consistent with the idea that	55	support for the	52
depends on the	58	that can be	62
differences in the	60	that is the	51
effect sizes were	59	that it is	86
for a review see	59	the ability to	64
for each of the	17	the activation of the	30
for example the	60	the amount of	66
from trial to trial	16	the association between	52
has / have been shown to	45	the comparison of	55
have been found to	16	the difference between the	22
have been used to	16	the distance between the	18
in a study of	17	the distribution of	55
in addition to the	17	the effect(s) of	334
in line with the	27	the impact of	57
in response to	81	the influence of	66
in the literature	62	the level of	60
in the meta-analysis	16	the location of the	19
in the next section	19	the magnitude of the	27
in the number of	19	the presence of a	19
in the presence of	22	the processing of	61
in the previous section	17	the proportion of	51
in this article	57	the set of	77
in this meta-analysis	16	the shape of the	17
in this section we	30	the size of the	58
included in the	68	the strength of the	33
individual differences in	83	the type of	65
is associated with	94	the value of the	22
is based on the	16	there is evidence that	18
(it) is important to note that	12	these findings suggest that	17
is possible that the	18	to account for the	37
is shown in figure	19	to respond to	51
it can be seen (that)	17	to the development of	22
it is important to note	14	ventromedial prefrontal cortex	50
it is likely that	16	we found that the	16
it is possible that (the)	56	were more likely to	27
it is possible to	21	with the idea that	24

This section addressed the frequency, structure, and functions of lexical bundles in psychology research articles, showing that psychology uses a large number of types and tokens of all the structural and functional categories, although the overwhelming majority of the bundles in this discipline are noun and prepositional phrases ('phrasal bundles'). Psychology also displays a considerable number of bundles which do not occur, or occur significantly less frequently, in the other two disciplines. The next section will examine the characteristic features of lexical bundles in literary studies research articles.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.3. LEXICAL BUNDLES IN LITERARY STUDIES RESEARCH

ARTICLES

4.3.1. The frequency of lexical bundles in literary studies research articles

Literary studies research articles also use a considerable number of lexical bundles: 96 three-word bundles, 76 four-word bundles, and 11 five-word bundles, amounting to a total of 183 bundle types. The number of the most frequent bundles in literary studies is 36, as opposed to 47 in psychology, though the subcorpus is about 31% smaller than that of the psychology research articles. In the literary studies subcorpus, 36 different bundle types occur with a frequency higher than 100 times per million words. It is interesting to note that although the number of types is roughly equivalent in these two subcorpora (if we consider the difference in corpus size), the individual frequencies are much lower than in psychology.

Thus, the most frequent bundle both in literary studies and in psychology is *as well as*, but even in this case the difference is striking. As was shown in the previous

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

section, *as well as* occurs 406 times in the psychology subcorpus (approximately 507 times per million words), whereas in literary studies its frequency is 196 (392 times per million words). This is a considerable difference¹. The patterns of usage and colligation are predictably similar in both disciplines:

(45) Mapped onto the narrative of the Assumption of the Virgin is the language of love as well as death. [Literary studies]

(46) This includes the particular details of the event as well as its theme and general structure, though it is the episodic details that are most severely affected. [Psychology]

The lexical bundles which occur more than 100 times per million words in literary studies research articles are presented in Table 4.7. It is worth mentioning at this point that some of the most frequent bundles in literary studies are topic- and discipline-specific (*the New York, of the play, in the novel*, to name but a few), as will be discussed below.

As can be seen from the table, only three of these have four words, the rest being three-word bundles. Two three-word bundles, *in the context* and *the context of*, overlap with the four-word bundle *in the context of*, but since their frequencies are not identical, all the three have been included in the list.

¹ Unpaired t-test showed this difference to be extremely statistically significant with two-tailed p value being less than 0.0001.

Table 4.7. The most frequent lexical bundles in literary studies research articles.

No	Bundle	Freq.	p.m.w.	No	Bundle	Freq.	p.m.w.
1	as well as	196	392	19	the New York	62	124
2	in order to	151	302	20	in other words	61	122
3	the end of	110	220	21	the question of	60	120
4	the fact that	90	180	22	to be a	60	120
5	of the poem	89	178	23	in the context	58	118
6	one of the	87	174	24	the same time	58	118
7	a kind of	83	166	25	there is no	58	118
8	it is the	63	166	26	in the context of	57	114
9	it is a	62	164	27	at the same	56	112
10	the work of	80	160	28	in early modern	56	112
11	in terms of	79	158	29	that it is	56	112
12	end of the	75	150	30	there is a	56	112
13	it is not	74	148	31	of the new	55	110
14	part of the	74	148	32	the use of	54	108
15	in which the	73	146	33	at the same time	52	104
16	the context of	71	142	34	in the novel	51	102
17	of the play	69	138	35	of his own	51	102
18	the end of the	65	130	36	[X]'s use of	51	102

Many of these bundles are parts of longer bundles which did not make it to the top because of their lower frequency.

The last bundle in the list is particularly interesting: it could be regarded as a pattern [proper noun] 's *use of the*. However, in this case it was decided to include it in this list as it seems to be a borderline case of what could be considered a pattern and represents certain pedagogical value. Some examples of this bundle in context follow.

(47) The combination of language and visual symbol in Peirce’s notations demonstrates the visual quality of his thought and harks back to Dickinson’s use of lineation and dashes.

(48) For all that Crooke’s use of ‘scarce’ suggests that he has repeatedly observed that mixed-sex twins have a greater tendency to die, however, he does not offer a remedy for it.

Of the 76 four-word bundles in literary studies research articles, only seven occur more than 50 times per million words: *the end of the, in the context of, at the same time, at the end of, as well as the, on the other hand, and the ways in which*, all of them ‘phrasal’ bundles typical of academic writing.

Surprisingly, five-word bundles are more abundant than those found in psychology, if we take into account that the psychology subcorpus is larger than the literary studies subcorpus (800.000 words versus 550.000 words). There are eleven five-word bundles, as shown in Table 4.8.

Table 4.8. Five-word lexical bundles in literary studies research articles.

Bundle	Freq.	p.m.w.	Bundle	Freq.	p.m.w.
at the end of the	32	64	in the case of the	8	16
in the context of the	20	40	in the wake of the	8	16
what it means to be	11	22	of the ways in which	8	16
in the same way that	9	18	of what it means to	8	16
over the course of the	9	18	the end of the poem	8	16
at the same time the	8	16			

Of these, two bundles can be considered a case of overlap: *what it means to be* and *of what it means to* are parts of the larger bundle *of what it means to be*. Below are some examples of the use of this bundle in context:

- (49) Gilb's refusal of abstraction and the confusion of materiality in his fiction preclude a clear explanation of what it means to be a person in the world if personhood is rooted in the body rather than in ideas of what the body signifies.
- (50) Creeley's assessment of Oppen's relationship to poetry raises the ethical question of what it means to write poetry in an age of increasing calamities.
- (51) The way *Anil's Ghost* represents it, what it means to be a doctor is exactly as we might expect, based on the regulative ideals that universally organize professional practice within biomedicine

It can be seen from these lists of lexical bundles that the clear majority of these sequences are 'phrasal' bundles, as in the psychology research articles. Their structure will be discussed in the following subsection.

4.3.2. The structure of lexical bundles in literary studies research articles

The structural classification of lexical bundles in literary studies research articles is presented in Table 4.9.

Table 4.9. Structural classification of lexical bundles in literary studies research articles.

Structure	Examples
Adjective / noun + noun phrase	the black arts movement, the early modern, the New York Times, the New York, the [years]s and [years]s, the twenty first century
Noun phrase with <i>of</i> -phrase fragment	a great deal of, a kind of, a number of, a sense of the, a series of, of African American, of his own, of the body, of the book, of the city, of the human, of the new, of the novel, of the New York, of the past, of the play, of the poem, of the same, of the story, of the text, of the twentieth century, of the ways in which, of the world, of what it means to, one of the most, part of the, [X]'s use of the, the absence of, the context of the, the course of the, the death of, the figure of the, the history of the, the image of the, the kind of, the language of the, the meaning of the, the name of the, the possibility of, the power of, the publication of the, the question of the, the rest of the, the role of the, the story of, the structure of the, the use of the, the work of, version of the
Noun phrase with other post-modifier fragments	an attempt to, attention to the, the extent to which, the fact that the, the idea that, the long poem, the relationship between the, the same time the, the way in which, the ways in which
Prepositional phrase with embedded <i>of</i> -phrase fragment	as a form of, as a kind of, as a means of, as a result of, as a way of, as part of the, at the center of, at the end of the, at the heart of, in a state of, in the absence of, in the case of the, in the context of the, in the face of, in the form of, in the image of, in the light of, in the process of, in the wake

	of the, in the words of, on the part of, over the course of, over the course of the
Other prepositional phrase	to the fact that, at the same time, at the same time the, at the time, in a letter to, in a way, in a way that, in contrast to the, in early modern England, in order to, in other words, in relation to the, in terms of the, in the novel, in the play, in the poem, in the same, in the same way, in the same way that, in the world, in this way the, in which the, on the one hand, on the other hand
<i>It</i> + verb phrase / adjective phrase	-----
Passive verb + prepositional phrase fragment	be seen as, can be read as
Copula <i>be</i> + noun phrase	but it is, is not a, is not the, it is a, it is in, it is not, it is the, to be a
(Verb phrase +) <i>that</i>-clause fragment	that he is, that there is no
Verb / adjective + <i>to</i>- clause fragment	it is necessary to
Adverbial clause fragment	as a whole
Other expressions	and in doing so, and in the, as we have seen, as well as, as well as a, as well as its, as well as the, it means to be, out of the, such as the, there is a, there is no, what it means to, what it means to be

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

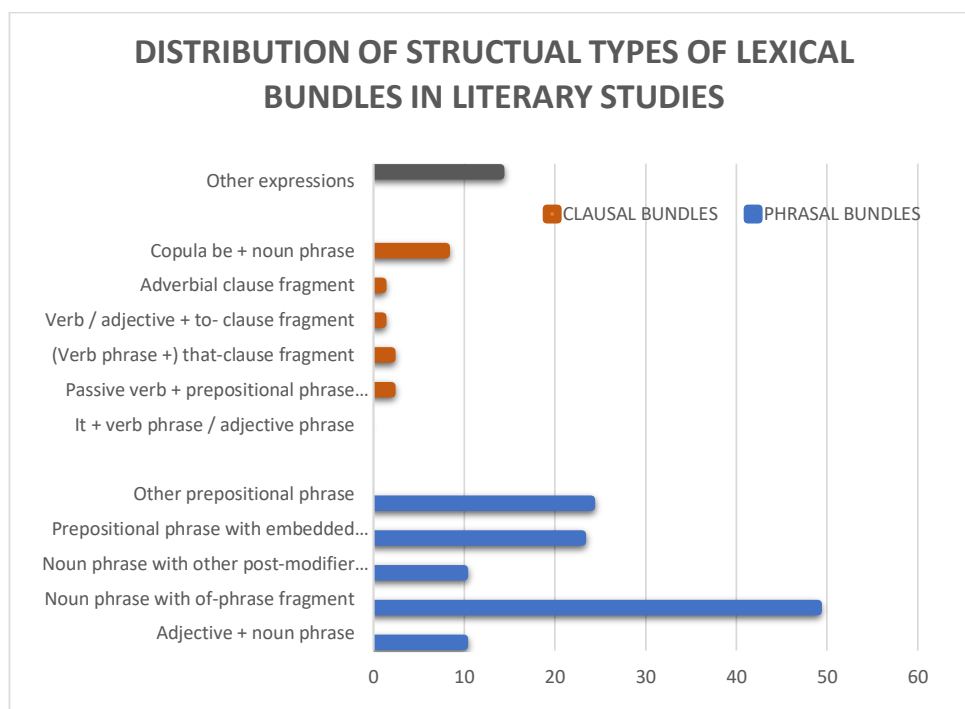
SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Figure 4.5. The distribution of structural types of lexical bundles in Literary Studies research articles.



One finding in this subcorpus is the strikingly high proportion of ‘phrasal’ bundles which outnumber this category in psychology research articles. There are 102 ‘phrasal’ bundle types and only 13 ‘clausal’ types. Figure 4.4 shows the proportion of bundle types of each structure in this subcorpus.

The structural classification of lexical bundles in literary studies reveals that many categories contain very few items, particularly the ‘verb phrase’ categories. One of them presents no items at all, the others contain one or two. The ‘Anticipatory *it* + verb phrase / adjective phrase’ is vacant, although it could be argued that the bundles

it is necessary to fits there. However, in that case, the category ‘Verb / adjective + *to* clause fragment’ would remain vacant, which would mean that at least one structural category contains zero items in this subcorpus.

Below is an example of this bundle in context:

- (52) To make visible Nabokov’s engagement with Myers’s place in this debate, it is necessary to actively resituate Myers’s work in its context.

The limited number of bundles in ‘clausal’ categories indicates the highly phrasal character of academic writing in the field of literary studies.

There are two ‘pattern type bundles’ in this subcorpus: *[X]’s use of the* and *the [year]s and [year]s* which, strictly speaking, are not lexical bundles, but it was decided to keep them since they were retrieved automatically by the software and are valid sequences, as opposed to those removed from the psychology list which contained parenthesis, full stops and other punctuation marks, as in *et al. (2014). The:*

- (53) [...] impaired executive control abilities are responsible for certain language skills, especially those in the realm of discourse participation (Gyóri et al., 2002; Lukács et al., 2014). The fourth cognitive function fulfilled by the precuneus is that of (iv) consciousness. [Psychology]

Clearly, this pattern should not be considered a lexical bundle, although it was retrieved by AntConc (Anthony, 2010) as a cluster. However, the one in the example

below is a well-formed pattern, although it does not present special pedagogical value for advanced learners:

- (54) Many of Pound's readers met the volume with relief that he had returned to a more accessible and personal mode, softening, or at least sidestepping, his prior political and economic stances and exhibiting some degree of contrition about views that had made him unpopular in the 1940s and 1950s

Literary studies would appear to employ a large number of discourse organisers (text-oriented bundles), which will be discussed below.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.3.3. The functions of lexical bundles in literary studies research articles

As mentioned in the previous sections, the functional classification was carried out only across the three broad categories, namely, text-oriented, research-oriented, and participant-oriented bundles. Even this generalised classification is sometimes arbitrary and depends on the context, as was argued at the beginning of this chapter.

Table 4.10 below is a tentative functional classification of lexical bundles in literary studies.

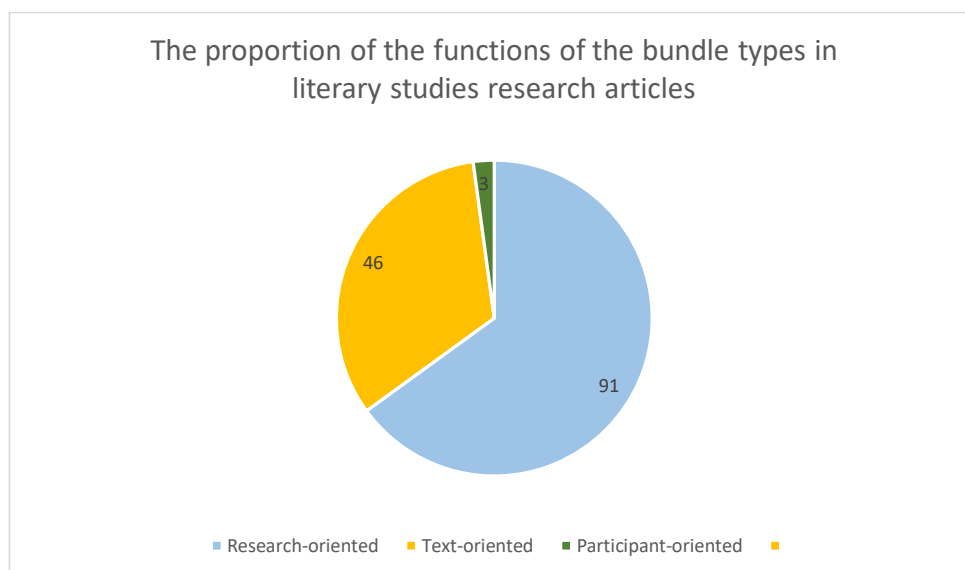
Table 4.10. Functional classification of lexical bundles in literary studies research articles.

Category	Bundle
Text-oriented	as a form of, as a kind of, as a means of, as a result of, as a way of, as a whole, as well as, as well as a, as well as its, as well as the, attention to the, be seen as, but it is, in contrast to the, in order to, in other words, in relation to the, in terms of the, in the case of the, in the context of the, in the light of, in the process of, in the words of, in which the, is not a, is not the, it means to be, on the one hand, on the other hand, on the part of, one of the most, out of the, such as the, the context of the, the course of the, the extent to which, the fact that the, the idea that, the meaning of the, the possibility of, the question of the, the relationship between the, the role of the, the same time the, the use of the, to the fact that
Research-oriented	a great deal of, a kind of, a number of, a sense of the, a series of, an attempt to, and in doing so, and in the, as part of the, at the center of, at the end of the, at the heart of, at the same time, at the same time the, at the time, in a letter to, in a state of, in a way, in a way that, in early modern England, in the absence of, in the face

	of, in the form of, in the image of, in the novel, in the play, in the poem, in the same, in the same way, in the same way that, in the wake of the, in the world, in this way the, it is a, it is in, it is not, it is the, of African American, of his own, of the body, of the book, of the city, of the human, of the new, of the New York, of the novel, of the past, of the play, of the poem, of the same, of the story, of the text, of the twentieth century, of the ways in which, of the world, of what it means to, over the course of, over the course of the, part of the, that he is, that there is no, the [years]s and [years]s, the absence of, the black arts movement, the death of, the early modern, the end of the poem, the figure of the, the history of the, the image of the, the kind of, the language of the, the long poem, the name of the, the New York, the New York Times, the power of, the publication of the, the rest of the, the story of, the structure of the, the twenty first century, the way in which, the ways in which, the work of, there is a, there is no, to be a, version of the, what it means to, what it means to be [X]'s use of the
Participant-oriented	as we have seen, can be read as, it is necessary to

As can be seen from the table, there are less participant-oriented bundles in the literary studies subcorpus than in psychology, especially if we consider the former's smaller size. Only 3 out of 140 lexical bundles are participant-oriented, whereas the number of text-oriented and research-oriented bundles is only slightly inferior to those in psychology (See Figure 4.6).

Figure 4.6. The number of types of lexical bundles for each function in Literary Studies research articles.



If we compare the percentages of the functional categories, the proportion of research-oriented bundles is higher in literary studies, and the percentage of text-oriented bundles is roughly equivalent to those in psychology: of the total number of 140 lexical bundles in the combined list, 91 (65%) are research-oriented, 46 (almost 33%) are text-oriented, and 3 (slightly over 1%) are participant-oriented. This seems to contradict the view that research-oriented bundles tend to be frequent in hard sciences (Hyland, 2008a). Two possible explanations for this fact can be offered. Firstly, as we saw, the functional classification of lexical bundles is always debatable; other analysts could obtain different results from the same corpus. The other tentative explanation is that even though the bundle types are more abundant in literary studies than in psychology (which is closer to natural sciences than literary studies), their

frequencies are much lower. In other words, literary studies writers employ a greater number of types but use these types less frequently than the psychology writers. This results in a higher number of tokens in psychology research articles, although the number of types is higher in literary studies.

Examples of each functional category are presented below.

Text-oriented bundles

(55) Luciana tries to teach Adriana how to find within herself the strength to overcome the jealousy and impatience that she feels as a result of her husband’s inattentiveness.

(56) Just as ‘bodkin’ may cue the more knowledgeable members of the audience to recognise a Caesar reference here, ‘the poor man’s contumely’ and ‘law’s delay’ can easily be read in the light of what we have already discussed as a reference to Sir Julius and his complaint of 1598.

(57) The possibility of latent considerations, for which this argument could offer cover, only emphasizes that this was a moment when the idea that wealth ought to be distributed more equally had political currency, especially when directed toward the king.

An example of the above-mentioned multifunctionality of lexical bundles is the following use of *the possibility of* as a research-oriented bundle:

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. Su autenticidad puede ser contrastada en la siguiente dirección https://sede.ull.es/validacion/	
Identificador del documento: 973669	Código de verificación: mnf6zN2e
Firmado por: DIANA BALASANYAN UNIVERSIDAD DE LA LAGUNA	Fecha: 30/06/2017 14:44:14
SALLY FRANCES BURGESS . UNIVERSIDAD DE LA LAGUNA	30/06/2017 21:06:24
ERNESTO PEREDA DE PABLO UNIVERSIDAD DE LA LAGUNA	14/07/2017 13:44:17

(58) Her story had run in the August 1952 issue and drawn favorable notices from editors at Knopf and Harper’s, but she was nonetheless modest about the possibility of winning the guest editorship for which she had subsequently applied.

Other examples of **research-oriented bundles** are:

(59) In Life of Pi, the lived messiness of propinquity is ultimately subordinated to an allegory of fabulation, of the power of storytelling to generate empathy, and especially of the cognitive autonomy of the human imagination, which can powerfully imagine the world as it is not.

(60) Further to this, while describing the death of Christ at the hands of the Romans, James concludes nevertheless that ‘he could not be a friend to Caesar, that was not his enemy’

(61) Moreover, as cycling and sexual beings, ALP and the letter chapter’s bicyclist become a part of the novel’s numerous series of sexcycles, and, in viewing Finnegans Wake as a microcosm of history itself, they become a part of sexcycles as manifested across time.

Participant-oriented bundles

(62) Poetry is seen [sic!], as we have seen time and time again, as an intermediary between heaven and earth.

(63) Another sentence in this chapter can be read as a reference to the generative role of Myers’s thought in Chapter Two.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. <i>Su autenticidad puede ser contrastada en la siguiente dirección https://sede.ull.es/validacion/</i>	
Identificador del documento: 973669	Código de verificación: mnf6zN2e
Firmado por: DIANA BALASANYAN UNIVERSIDAD DE LA LAGUNA	Fecha: 30/06/2017 14:44:14
SALLY FRANCES BURGESS . UNIVERSIDAD DE LA LAGUNA	30/06/2017 21:06:24
ERNESTO PEREDA DE PABLO UNIVERSIDAD DE LA LAGUNA	14/07/2017 13:44:17

(64) To make visible Nabokov's engagement with Myers's place in this debate, it is necessary to actively resituate Myers's work in its context.

As we have seen, the use of lexical bundles in literary studies is somewhat different from that in psychology, especially in terms of the rate of occurrence, type / token ratio, and distribution of functional and structural categories.

Below is a discussion of the bundles found only in the literary studies subcorpus.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.3.4. Lexical bundles present exclusively in literary studies research articles

Table 4.11 below is a list of the bundles found only in the literary studies subcorpus. The comparison has been conducted manually, using the combined lists (containing all the three-, four-, and five-word bundles) for the three disciplines and removing all the bundles which occur in any of the two other disciplines. This means that these bundles may occur in the other two subcorpora, but they were not retrieved by the corpus analysis software due to their low frequency.

Table 4.11. Lexical bundles found exclusively in literary studies research articles.

Bundle	Freq.	Range	Bundle	Freq.	Range
a kind of	83	37	of the New York	17	4
a sense of (the)	46	29	of the novel	39	12
a series of	48	29	of the past	44	18
an attempt to	34	21	of the play	69	15
and in doing so	13	8	of the poem	89	17
and in the	33	24	of the same	39	24
as a form of	18	13	of the story	31	12
as a kind of	23	17	of the text	41	18
as a means of	16	12	of the twentieth century	11	7
as a way of	13	8	of the world	42	20
as a whole	35	19	of what it means to	8	6
as we have seen	11	8	on the part of	12	10
at the center of	12	9	out of the	42	30
at the heart of	17	13	the [year]s and [year]s	10	7
attention to the	39	22	the black arts movement	12	4
be seen as	35	13	the context of (the)	71	29

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

but it is	33	22	the death of	31	17
can be read as	15	11	the early modern	33	9
in a letter to	12	7	the end of the poem	8	4
in a state of	10	9	the figure of the	17	5
in early modern England	56	14	the history of the	45	20
in the image of	10	8	the image of the	47	25
in the light of	11	8	the kind of	32	19
in the novel	51	12	the language of (the)	36	25
in the play	41	12	the long poem	36	4
in the poem	35	15	the meaning of (the)	32	21
in the process of	10	8	the name of (the)	44	13
in the same way (that)	14	11	the New York	62	11
in the wake of (the)	20	11	the New York Times	11	10
in the words of	11	8	the possibility of	39	25
in the world	32	18	the power of	37	21
in this way (the)	36	18	the publication of the	10	6
it is necessary to	10	5	the question of (the)	60	21
it means to be	11	5	the story of	34	19
of African American	43	6	the structure of the	13	5
of his own	51	29	the twenty first century	10	5
of the body	49	17	the work of	80	33
of the book	42	14	version of the	32	15
of the city	31	10	what it means to	16	7
of the human	35	12	what it means to be	11	5
of the new	55	19	X's use of (the)	51	29

Similar to the previous section's findings, this 'exclusive list' contains many topic-specific bundles, such as *of the play*, *of the past*, *of the book*, *can be read as*, etc. The occurrence of the toponym *New York* in several bundles is remarkable, pointing to the major role of this city in the literary life of the English-speaking world, although its pedagogical value may be debatable. The occurrence of the sequence *of African American* also indicates a major trend in contemporary literature and a central topic in literary studies. However, there are several topic-neutral bundles which are particularly frequent in the literary studies corpus: *a kind of*, *an attempt to*, and *in*

the, as a whole, as we have seen, etc. The analysis of concordance lines for these sequences reveals a few topic-specific features in the context of this discipline, although these bundles theoretically could collocate with entirely different sets of words in other contexts. Thus, *a kind of* collocates with *social relay, life, house, community, barefoot doctor, triumph over death, death*, and other abstract and concrete nouns; *an attempt to* is found in the company of such words and phrases as *re-establish the hierarchy, break free, change the situation, be faithful to his own community*, etc.; *as a whole* refers to works of literature, such as poems and novels, but also to *nation, community, race, city, era*, and many other nouns. *As we have seen* can be used in virtually any piece of academic writing as a rhetorical step of summarising what has been said before, whereas *and in the* can be regarded as a “structural only” (Cortés, 2008: 47) bundle which can be used in a variety of contexts. However, the occurrence of these bundles in the other subcorpora is minimal, if any at all.

The existence of discipline-specific bundles which do not seem to be motivated by the communicative purposes of the genre requires further research

This section addressed the frequency, structural types, and functions of lexical bundles in literary studies research articles, and presented a list of the lexical bundles exclusive to this discipline. The next section will examine the lexical bundles in linguistics research articles.

4.4. LEXICAL BUNDLES IN LINGUISTICS RESEARCH ARTICLES

4.4.1. The frequency of lexical bundles in Linguistics research articles

A search for three-, four-, and five-word bundles yielded a total of 111 bundles in Linguistics research articles: 55 three-word bundles, 50 four-word bundles, and 6 five-word bundles. The complete list of these bundles can be found in Appendix 3.

Eleven of these bundles occurred more than 100 times per million words (see Table 4.12), all of them being three-word bundles.

Table 4.12. The most frequent lexical bundles in linguistics research articles.

Bundle	Freq.	p.m.w.	Bundle	Freq.	p.m.w.
in terms of	219	273	in this study	105	131
the use of	217	271	the fact that	102	127
in order to	215	267	in the following	93	116
as well as	178	222	part of the	93	116
the present study	157	196	one of the	92	115
a number of	121	151			

The fact that none of the four-word bundles has made it to the top may be explained by the smaller overall number of types in linguistics than in the other two disciplines, as well as by the smaller number of very frequent bundles: there are three times more high-frequency lexical bundles in literary studies and four times more in psychology than in linguistics. Almost half of the bundles are fragments of *of*-phrases, and all of them belong to the ‘phrasal’ category.

In contrast to the literary studies frequent bundles list, there are no topic-specific bundles among those most frequent in linguistics. This finding can be explained by the much smaller number of the latter.

Predictably, the number of frequent four-word bundles is even smaller: only five of these sequences in linguistics research articles occur more than 50 times per million words (Table 4.13).

Table 4.13. The most frequent four-word bundles in Linguistics research articles.

Bundle	Freq.	p.m.w.	Bundle	Freq.	p.m.w.
on the other hand	63	79	on the basis of	45	56
in the case of	50	62	at the same time	44	55
in the present study	48	60			

All of these bundles are prepositional phrase fragments, and none of them is topic-specific.

Only six five-word bundles were retrieved from this subcorpus, two times fewer than in the other two disciplines. Table 4.14 summarises this information.

Table 4.14. The five-word bundles in linguistics research articles.

Bundle	Freq.	p.m.w.	Bundle	Freq.	p.m.w.
at the beginning of the	17	21	at the end of the	14	17
on the other hand the	16	20	in the case of the	13	16
as can be seen in	15	19	English as a foreign language	12	15

Two of the five-word bundles, *on the other hand the* and *at the end of the*, are longer versions of two extremely frequent four-word bundles. If they were to be treated as one type, the number of four-word bundles would be even smaller. It has been argued that the study of four-word bundles renders the most reliable results (Ädel & Erman, 2012; Chen & Baker, 2010; Hyland, 2008a) as many four-word bundles contain the smaller three-word bundles and are part of the larger five-word bundles, thus covering most part of the multi-word sequences in a genre. However, in less formulaic disciplines such as linguistics, it seems more plausible to address bundles of all sizes to obtain the full picture, since the bundles in this discipline are not numerous and can be subjected to manual analysis. ‘Less formulaic’ in this case refers to a smaller number of types than in other disciplines. The issue of the ‘formulaicity’ of the three disciplines will be discussed in Chapter six.

As we have seen, the overwhelming majority of lexical bundles in linguistics are ‘phrasal’, similar to the other two disciplines. The structural characteristics of lexical bundles in linguistics are discussed below.

4.4.2. The structure of lexical bundles in linguistics research articles

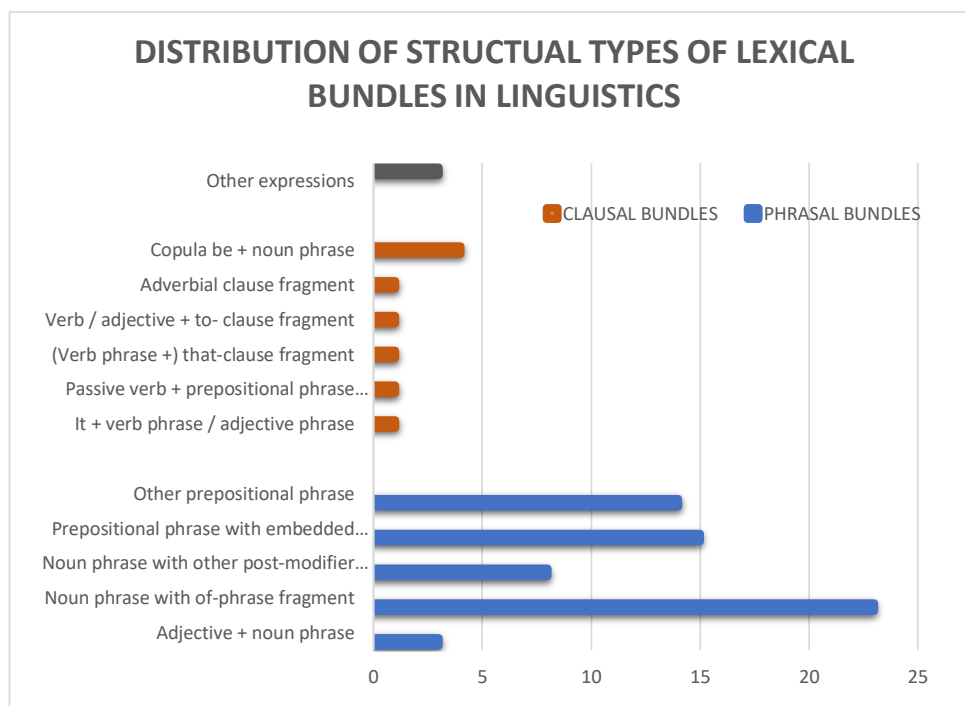
Table 4.15 shows the classification of lexical bundles in Linguistics research articles according to their structure, and Figure 4.6 shows the number of types in each structural category. It should be noted that, as in the previous sections, the lexical bundles included in this classification are those from the combined list which comprises the prototypical three-, four-, and five-word bundles found in Linguistics research articles. In other words, all the 111 bundles were included in the list, and then the subsumed bundles were removed.

It is not surprising that most of the bundles found in Linguistics are ‘phrasal’, as in the other two disciplines, although the difference between the ‘phrasal’ and ‘clausal’ type numbers is somewhat more dramatic: 63 ‘phrasal’ bundles (84% of the total number of types on the list) as opposed to 9 ‘clausal’ ones (12% of the total) and 3 ‘other expressions’ (4% of the total). Figure 4.7 illustrates this distribution.

Table 4.15. Structural classification of lexical bundles in linguistics research articles.

Structure	Examples
Adjective / noun + noun phrase	the current study, the most frequent, the other hand the
Noun phrase with <i>of</i> -phrase fragment	a number of, a small number of, a range of, a wide range of, each of the, of the present study, of the study, of this study, one of the, part of the, some of the, the analysis of the, the beginning of the, the development of, the end of the, the importance of, the nature of the, the number of, the rest of the, the results of the, the role of, the total number of, the use of the
Noun phrase with other post-modifier fragments	English as a foreign language, in relation to the, the extent to which, the relationship between, the ways in which, to the fact that the, with a view to, with respect to the
Prepositional phrase with embedded <i>of</i> -phrase fragment	as a result of, as part of a, at the beginning of the, at the end of the, at the time of, by the use of, for the purpose of, from the perspective of, in terms of the, in the case of the, in the context of, in the field of, in the form of, in the use of, on the basis of
Other prepositional phrase	at the same time, between the two, for the control group, in order to, in other words the, in the following example, in the present study, in the two, in this case, in this study, in which the, on the one hand, on the other hand, on the other hand the
<i>It</i> + verb phrase / adjective phrase	it is important to
Passive verb + prepositional phrase fragment	can be seen in
Copula <i>be</i> + noun phrase	to be able to
(Verb phrase +) <i>that</i> -clause fragment	that there is a
Verb / adjective + <i>to</i> -clause fragment	found to be
Adverbial clause fragment	as can be seen, as in the, based on the, due to the
Other expressions	as well as the, such as the, there is no

Figure 4.7. The number of types in each structural category of lexical bundles in linguistics research articles.



Although the corpus used in this study is too small to make any generalisations regarding the structural type proportions for the discipline as a whole, it is clear that the ‘phrasal’ bundles in linguistics are much more frequent than ‘clausal’ bundles: there are 7 times as many ‘phrasal’ bundles than ‘clausal’ ones, and 21 times as many ‘phrasal’ bundles than ‘other expressions’. In addition, many of the phrasal bundles

are among the most frequent, i.e., they outnumber the other categories both in type and in token.

Despite the relatively small size of the corpus, a clear tendency can be outlined: linguistics uses at least two times fewer lexical bundles than psychology or literary studies.

The examples below illustrate the formulaic and predominantly phrasal character of academic discourse in linguistics research articles.

- (65) On the other hand, the role of metadiscourse markers will be studied in order to assess the attainment of persuasion in the texts.
- (66) However, a small number of texts had relatively high numbers of potential errors.
- (67) The extent to which collaborative blogs are used by scholars when writing for publication is as yet unclear.
- (68) Just like other syntactic constructions, the xagam construction continues to evolve in the language as a result of the use people make of it in interaction.
- (69) A similar scoring procedure was applied for the control group, i.e. these students were also awarded a maximum of 20 points for content and 20 points for writing skills: text organisation, vocabulary and grammar knowledge, and spelling.

(70) As interviews can be seen as involving interactive constructions of meanings, it is important to consider my position as an interviewer and a researcher.

(71) As can be seen above, there appears to be a wish on the part of these participants to retain their own local accent in order to express their lingua-cultural identity.

(72) I began to notice that the student-led presentation could provide students with opportunities to participate in the cultural practice of the class, as well as to challenge them by requiring them to participate without clear instructions regarding the teacher's expectations.

The last example is interesting because it shows that *as well as*, one of the most frequent lexical bundles in many disciplines and genres, can colligate with virtually any syntactic units: in this particular case, with a verb, a rather infrequent colligation pattern in the three disciplines, although not impossible. Such cases may be helpful to novice researchers, both L1 and L2, as they can provide insights on the usage patterns regarded as acceptable in the target discourse community.

From the above analysis it can be concluded that linguistics relies heavily on 'phrasal' bundles, although the overall number of lexical bundles in this discipline is significantly smaller than in the other two disciplines. Below is a discussion of the functions of lexical bundles in linguistics research articles.

4.4.3. Functional classification of lexical bundles in linguistics

research articles

Of the 75 bundles in the combined list, 33 (47%) bundles turned out to be text-oriented, 39 (49%) were research-oriented, and only 3 (4%) can be classified as participant-oriented, with the above-mentioned reservations regarding the functional classification of lexical bundles (see Section 4.1). Table 4.16 lists all the bundles in each functional category.

Table 4.16. Functional classification of lexical bundles in linguistics research articles.

Category	Example
Text-oriented	as a result of, as well as the, based on the, can be seen in, due to the found to be, in order to, in other words the, in relation to the, in terms of the, in the case of the, in the context of, in the field of, in the following example, in the present study, in this case, in this study, in which the, of the present study, of the study, of this study, on the basis of, on the one hand, on the other hand, on the other hand the, such as the, that there is a, the analysis of the, the current study, the other hand the, the results of the, to the fact that the, with a view to, with respect to the
Research-oriented	a number of, a range of, a small number of, a wide range of, as in the, as part of a, at the beginning of the, at the end of, at the same time, at the time of, between the two, by the use of, each of the, English as a foreign language, for the control group, for the purpose of, from the perspective of, in the form of, in the two, in the use of, one of the, part of the, some of the, the beginning of the, the end of the, the development of, the extent to which, the importance of, the most frequent, the nature of the, the number of, the relationship between, the rest of the, the role of, the total

	number of, the use of the, the ways in which, there is no, to be able to
Participant-oriented	as can be seen, as can be seen in, it is important to

The proportional distribution of the functions is shown in Figure 4.8. The figure shows a roughly equal distribution of bundles across the research-oriented and text-oriented categories, the former comprising a slightly larger number of types.

The small number of lexical bundles in linguistics research articles suggests that this discipline is much less formulaic than psychology or literary studies. In addition, the number of tokens in linguistics research articles is significantly inferior to that in the other two disciplines: this becomes clear upon comparing the frequencies of the most frequent lexical bundles in the three disciplines. linguistics employs significantly less types of bundles, and it also employs them relatively less frequently.

A possible explanation for the fact that linguistics research articles contain less formulaic expressions than texts in the other two disciplines may be explained by the expectations of the discourse community and the generic conventions present in this discipline. The communicative functions realised by lexical bundles in psychology and literary studies may be performed by other textual elements in linguistics, but this question is beyond the scope of this study.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

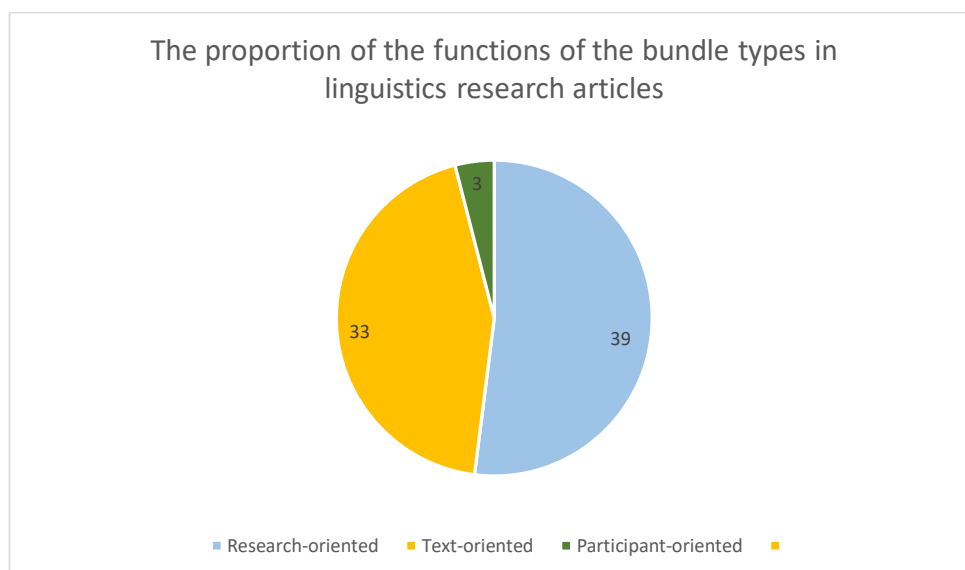
SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Figure 4.8. The number of types representing each function of lexical bundles in linguistics research articles.



It may be argued that linguistics uses less formulaic expressions due to the expectations of its discourse community members who believe that linguists are expected to produce more novel, creative language instead of relying on formulae. Although I am not aware of any study of lexical bundles in literary studies, it would not be unreasonable to suppose that this expectation would be even stronger for literary studies.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.3.4. Lexical bundles exclusive to linguistics research articles

Table 4.17 below presents the full list of all the bundles that occur only in linguistics research articles, retrieved via manual comparison of the three combined lists of lexical bundles. For a detailed description of the procedure see Section 4.2.4 above.

Table 4.17. Lexical bundles found only in Linguistics research articles.

Bundle	Freq	Bundle	Freq	Bundle	Freq
a range of	51	for the purpose of	17	of this study	50
a small number of	17	found to be	52	one of the	92
as can be seen	21	from the perspective of	20	the analysis of the	24
as can be seen in	15	in relation to the	20	the beginning of the	21
as in the	57	in terms of (the)	37	the current study	56
at the beginning of (the)	17	in the field of	17	the development of	55
based on the	69	in the following (example)	18	the most frequent	50
by the use of	16	in the two	54	the number of	67
can be seen in	23	in the use of	22	the other hand the	16
due to the	57	in this case	50	the total number of	19
each of the	55	in this study	105	to be able to	21
English as a foreign language	12	of the present study	28	with a view to	16
for the control group	17	of the study			

A total of 38 bundle types were identified as occurring exclusively in linguistics research articles. Of these, only one can be classified as topic-specific: *English as a foreign language*.

The remaining bundles can be regarded as topic-neutral. The examples below illustrate their topic-neutral character:

- (73) Linguistic research on the characteristics of these reports has been conducted in a range of contexts, with genre move structures and other core language elements identified in ethnographically focused New Rhetoric genre studies.
- (74) Though based on a small number of exemplars, it provides a linguistic-based account of an authentic contract negotiation process.
- (75) In the next few paragraphs we describe each of the factors that influence Accessibility as proposed by Ariel (1990), after which we test them with respect to cataphora.
- (76) In the following example, shared knowledge is disrupted as blog participants try to follow the notation used by previous authors (what would be described as “standard notation” in an RA).

Linguistics is not only the least formulaic of the three disciplines, it also employs a very large number of unique topic-neutral bundles which occur only in this discipline. As observed in the previous sections, this does not mean that these bundles are altogether absent in the other disciplines, rather, they do not occur frequently enough

to be retrieved by the corpus analysis software. Thus, linguistics seems to be ‘an outlier’ in terms of recurrent word combinations.

It is also interesting to note that linguistics uses the largest number of word-combinations referring to the study itself: *in this study, in the study, of the present study, of the current study, of the study*, etc. A tentative explanation for this finding may lie in the conventions of the discourse community particularly in a discipline where practitioners are likely to be very conscious of the need for clarity and the coherent organisation, hence the abundance of elements referring explicitly to the structure of the text and guiding the reader through it.

This section addressed the frequency, structure, and functions of lexical bundles in linguistics research articles. The next section will discuss the similarities and differences of the use of lexical bundles in psychology, literary studies, and linguistics.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

**4.5. LEXICAL BUNDLES ACROSS DISCIPLINES: FREQUENCY,
STRUCTURE, AND USE OF LEXICAL BUNDLES IN
PSYCHOLOGY, LITERARY STUDIES, AND LINGUISTICS**

4.5.1. Differences in the use of formulaic language across the three disciplines

This section will address the differences in formulaic language observed in the three disciplines.

The first obvious difference is the quantity of lexical bundles found in each discipline: both for the initial lists (the bundles retrieved through separate searches for three-, four-, and five-word bundles) and on the combined list (a unified list generated by removing bundles which are fragments of longer sequences and adding all the remaining bundles of different lengths) the numbers are strikingly different.

The search in the psychology corpus yielded 254 different bundles types which

resulted in a combined list of 172 bundles, for literary studies, the initial list contained 183 bundles, and the combined list, 79, whereas for linguistics, the initial list had only 111, and the combined list 80. As can be seen from these numbers, the initial linguistics list suffered fewer changes than those of the other two disciplines. This is due to the smaller number of three-word bundles on the initial list for linguistics. Many of the shorter sequences are parts of four- and five-word bundles and did not make it to the combined list, but in linguistics the number of three-word bundles was initially 55, as opposed to the 126 three-word bundles in the psychology subcorpus and 96 in the literary studies corpus, hence many of them could not be combined with longer sequences either because they were not fragments (*in order to*) or because the frequency of the three-word bundle was not comparable to that of the four-word bundle (*as well as* occurred 178 times, whereas *as well as the* occurred only 57 times). Table 4.18 shows the number of bundles retrieved for each discipline, number of types and tokens, and the search criteria for each bundle size.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Table 4.18. The number of lexical bundle types in psychology, linguistics, and literary studies research articles.

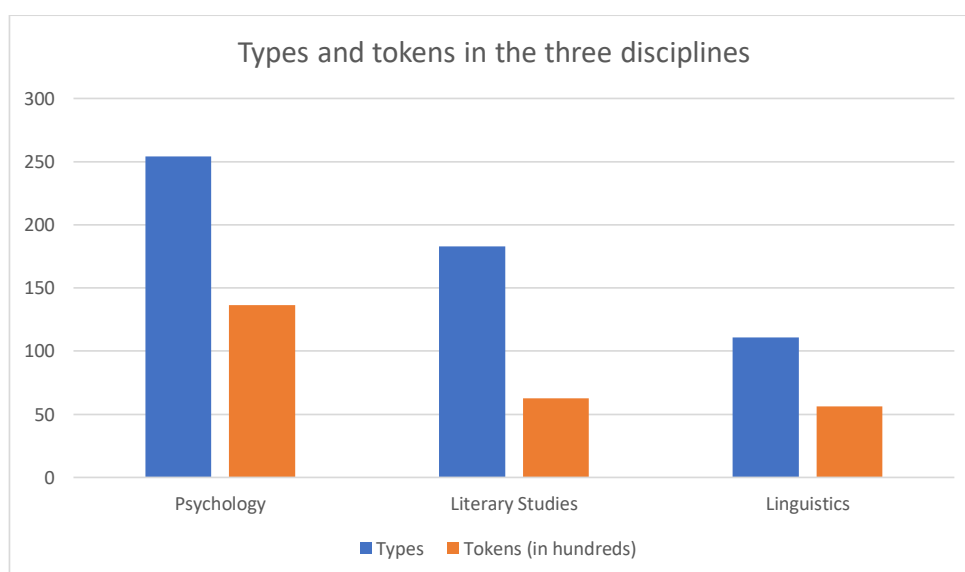
	Psychology	Linguistics	Literary studies	
three-word	Min. freq.	50	50	31
	Min. freq. p.m.w.	62.5	62.5	62
	Min. range	4	4	4
	Types / tokens	126 / 10.417	55 / 4278	96 / 4841
four-word	Min. freq.	16	16	10
	Min. freq. p.m.w.	20	20	20
	Min. range	4	4	4
	Types / tokens	115 / 3069	50 / 1244	76 / 1276
five-word	Min. freq.	12	12	8
	Min. freq. p.m.w.	15	15	16
	Min. range	4	4	4
	Types / tokens	13 / 169	6 / 87	11 / 129
Total tokens	254 / 13.655	111 / 5.609	172 / 6.246	

As can be seen from the table, psychology is the most formulaic of the three disciplines, and linguistics has the least number of types, although they occur with considerable frequency.

It was argued in the previous section that the genre conventions and discourse community expectations in linguistics may require less formulaicity and more novel utterances. It is believed that linguists are sensitive towards the language and expect good writing to be creative, even if it is academic writing. This hypothesis should

also hold true for literary studies researchers whose object of study is creative writing by definition. However, at first glance, it may seem that literary studies writers employ more formulaic language than their counterparts in linguistics, and the hypothesis does not hold true.

Figure 4.9. Lexical bundle types and tokens in psychology, literary studies and linguistics.



The picture becomes clearer when we look at the type / token ratio in the three disciplines: psychology uses the greatest number of types and tokens, but it also uses the same type more frequently than linguistics or literary studies. The frequencies of many bundle types in psychology are 3-4 times greater than those of linguistics or literary studies bundles. Literary studies uses a greater variety of bundle types than linguistics, but with lower frequencies. Indeed, as Fig. 4.9 shows, the tokens used by

250

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

literary studies researchers are almost equal in number to those used by linguists, although literary studies uses almost twice as many types. In other words, literary studies writers employ a greater variety of bundles than linguists, but use them two times less frequently. However, it must also be remembered that the corpus of literary studies research articles is almost 32% smaller than that of linguistics or psychology, and it is not clear what results would be obtained from an 800.000-word corpus in literary studies.

Of the three disciplines, linguistics may seem the least formulaic in terms of bundle types, but literary studies, although showing more bundle types, uses each bundle fewer times. This can be illustrated through type / token ratios, as shown in Table 4.19.

Table 4.19. Type / token ratios for the lexical bundles in the three disciplines.

Discipline	Types	Tokens	Type / token ratio
Psychology	254	13.655	0,018
Literary studies	183	6.246	0,029
Linguistics	111	5.609	0,02

A greater type / token ratio shows less repetition of the same types, i.e., greater variation in the use of different types, and hence less formulaicity. According to this parameter, linguistics is more formulaic than literary studies, but slightly less formulaic than psychology.

Despite the difference in subcorpus size, the data seem to support the idea that literary studies tends to be less formulaic. Thus, of the three disciplines, the ‘harder’

science is the more formulaic and the ‘softer’ one seems to be the least formulaic, although the difference between psychology and linguistics is minimal.

It is worth comparing the findings of the psychology subcorpus with those of Farvardin, Afghari and Koosha (2012), who extracted four-word lexical bundles from a three million word corpus of research articles in physics and obtained 188 different types of bundles and a total of 28.547 tokens. This is much more than the 126 types and 4.065 tokens in the psychology subcorpus (800.000 words) in this study. The frequency threshold used by Farvardin, Afghari and Koosha was identical to the one used in the present study, i.e. 20 occurrences per million words, which means that the only variable is the discipline. The results are thus directly comparable. Normalised numbers for tokens would be 7.612 tokens per million words for physics research articles and 5.081 tokens per million words for psychology research articles, 1,5 times more bundle tokens in physics than in psychology. The number of types cannot be compared through normalising to one million words because type does not depend so much on corpus size as on the frequency cut-off point, which is identical in the two studies.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Table 4.20. The most frequent lexical bundles in psychology, literary studies, and linguistics.

PSYCHOLOGY			LITERARY STUDIES			LINGUISTICS		
	Bundle	Norm. freq.		Bundle	Norm. freq.		Bundle	Norm. freq.
1	as well as	507	1	as well as	392	1	in terms of	273
2	the number of	324	2	in order to	302	2	the use of	271
3	the effects of	269	3	the end of	220	3	in order to	267
4	in terms of	242	4	the fact that	180	4	as well as	222
5	one of the	182	5	of the poem	178	5	the present study	196
6	in which the	181	6	one of the	174	6	a number of	151
7	such as the	181	7	a kind of	166	7	in this study	131
8	more likely to	176	8	it is the	166	8	the fact that	127
9	a number of	171	9	it is a	164	9	in the following	116
10	consistent with the	167	10	the work of	160	10	part of the	116
11	the development of	155	11	in terms of	158	11	one of the	115
12	the fact that	154	12	end of the	150			
13	a function of	151	13	it is not	148			
14	in order to	151	14	part of the	148			
15	the role of	150	15	in which the	146			
16	the effect of	149	16	the context of	142			
17	based on the	147	17	of the play	138			
18	on the other	146	18	the end of the	130			
19	the presence of	144	19	the New York	124			
20	with respect to	140	20	in other words	122			
21	the relationship between	130	21	the question of	120			
22	according to the	126	22	to be a	120			
23	the absence of	126	23	in the context	118			
24	there is a	126	24	the same time	118			

Salazar's (2014) study of lexical bundles in health sciences research articles revealed a total of 769 types and 37.909 tokens in a two million word corpus, which is 18.954

253

tokens per million words for bundles ranging from three to six words. The total number of tokens for psychology bundles of all sizes in the present study is 13.655, but the results are not directly comparable since Salazar's study used a different frequency threshold (10 times per million words). However, the difference between the number of types identified in the two disciplines is noteworthy: 769 for health sciences as opposed to 254 types for psychology.

As for the structural characteristics, we have seen that psychology employs a greater number of verb phrase structures than the 'softer' disciplines, employing 43 'clausal' bundles, followed by literary studies with 14 different verb phrase structures. The most 'phrasal' of the three is linguistics, with only 9 'clausal' types of bundles and 63 'phrasal' bundle types.

Both linguistics and literary studies seem to neglect participant-oriented bundles, whereas psychology makes a much more extensive use of them. Although not very numerous in number, participant-oriented bundles are much more frequent in psychology than in the other two disciplines. Psychologists seem to be more concerned with engaging the reader and interacting with him or her than the writers of the other two disciplines.

This section attempted to argue in favour of a discipline-specific approach to studying and teaching lexical bundles. It is hoped that the arguments presented above demonstrate that the differences between the disciplines are not negligible, and that

a study of lexical bundles should ideally address each discipline separately in order to achieve accurate and pedagogically viable results.

The next section will turn to the bundles shared by the three disciplines and their patterns of usage.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

4.5.2. Bundles common to psychology, literary studies, and linguistics

A final search was performed to identify the four- and five-word lexical bundles common to the three disciplines. The final list was examined to remove any bundles which occurred predominantly in one discipline, as these were considered discipline-specific. The bundles which were shared only by two disciplines were also removed.

Table 4.21 shows the list of all the bundles identified as common to all the three disciplines, together with their raw and normalised frequencies. Several observations can be made with respect to this list.

The bundle *as a function of* was removed because it was underrepresented in the literary studies subcorpus (only 3 occurrences in 550.000 words). The same applied to *in the absence of* which also occurred 3 times in the linguistics subcorpus. The bundle *et al found that* showed a very high frequency (68 in per 2,15 million words), but occurred mainly in the psychology subcorpus, with 9 occurrences in the linguistics subcorpus and none in the literary studies. It was, therefore, removed from the list.

Table 4.21. List of the bundles shared by the three disciplines.

No	Bundle	Freq.	p.m.w.	No	Bundle	Freq.	p.m.w.
1	on the other hand	187	87	10	in terms of the	95	44
2	in the context of	178	83	11	as a result of	80	37
3	at the same time	137	64	12	it is important to	74	34
4	as well as the	135	63	13	the fact that the	69	32
5	the end of the	127	59	14	in the form of	67	31
6	on the basis of	120	56	15	with respect to the	67	31
7	the extent to which	107	50	16	on the one hand	62	29
8	in the case of	101	47	17	at the end of the	61	28
9	at the end of	95	44				

The bundle *it is possible that* was a borderline case with 7 occurrences in the linguistics subcorpus and 4 in literary studies, while the remaining 56 all occurred in psychology. It was decided to remove this bundle due to the fact that it was predominant in psychology, and occurred less than 5 times in Literary Studies. Similarly, *the size of the* was removed because it occurred 58 times in the psychology subcorpus and only 3 times in each of the other two subcorpora. By contrast, *with respect to the* was maintained since it met the requirement of occurring in at least 5 texts in each subcorpus, although its presence is much more marked in Psychology than in the other two disciplines. *Are more likely to* and *in the present study* were underrepresented in literary studies (1 and 0 occurrence respectively), and were also removed from the list.

Thus, the final list of shared bundles contains 17 bundles, only one of which is a five-word bundle: *at the end of the*, whose shorter equivalents, *at the end of* and *the end*

of the, are also present. These were treated as different bundles since their rates of occurrence are different. If these bundles were combined into one single sequence, the number of shared types would be even smaller. Examples of the use of this bundle in each of the disciplines follow.

- (77) The third part investigates the end of the period of silence, and considers how the call-taker re-establishes contact with the caller and resumes the thread. [Linguistics]
- (78) His cajoling worked, and “Fresh Air” was in print before the end of the year. [Literary studies]
- (79) Possible explanations for these mixed results will be returned to toward the end of the section on experimental studies. [Psychology]
- (80) At the end of the data collection period, there was a review of the field notes and the data collected to determine whether or not it was necessary to go back to the subjects for more information in order to better understand and analyse the data. [Linguistics]
- (81) Race emerges at the end of the story, when the narrator has definitively refused to see the mask. [Literary studies]
- (82) A more detailed discussion of these contributions is provided at the end of the current article. [Psychology]

Despite the difference in the rate of occurrence, there seems to be no divergence in the use of this bundle in either of its versions, except that *the end of the* is less likely

to appear in sentence-initial position, whereas *at the end of the* can be found in virtually any position in a sentence. It is interesting to note, however, that both *at the end of the* and *the end of the* seem to perform different functions depending on the discipline. Although in psychology they can signal time or process, their occurrence as text-deictic markers is more prominent in this discipline, as in (78) and (81). In literary studies and linguistics, these bundles more often perform a research-oriented function.

Of the remaining 13 bundles, 8 seem to be text-oriented (*on the other hand, in the context of, at the same time, as well as the, as a result of, the fact that the, with respect to the, on the one hand*), and only one can be classified as participant-oriented, namely *it is important to*. It is also the only verb phrase fragment on the list. Some examples of these follow.

(83) Poor communication skills, on the other hand, can result in patient dissatisfaction and complaints (e.g. Taylor, Wolfe, & Cameron, 2002).
[Linguistics]

(84) On the other hand, this claim could be seen as conflating the heterogeneity of contemporary poetry informed not only by digital technologies of production and circulation, but also by a generation of experimental poets excluded from the anthology's field of influence.
[Literary Studies]

- (85) On the other hand, the current results support the idea of a fully componential network architecture allowing for the application of the conversion strategy. [Psychology]

It is interesting to note the difference of frequency of *on the other hand* and *on the one hand*, two expressions typically taught together as parts of a single discourse-organising frame. The data show that *on the other hand* is much more frequent and often occurs separately. A more surprising finding is that *on the one hand* also occurs without its ‘closing’ counterpart 11 times across the three subcorpora:

- (86) On the one hand, this phrase emphasizes the past as distant and final, but it also indicates that Mr. Sanchez is done with the concept of pastness. Humiliated by his inability to move, he thinks, “it is not yet time to quit.” This “not yet time” evolves into the “all the time” that Mr. Sanchez then uses to describe the things he does with his body, like riding unbuckled in his wheelchair (16). [Literary Studies]

The aim of this section was to describe the bundles common to psychology, literary studies, and linguistics, and the similarities and differences in their use of lexical bundles. It was argued that a discipline-specific approach could provide important pedagogical insights for instructing novice academic writers. It was also argued that a corpus-based approach could be useful for such writers if they were instructed to use it for their own purposes.

The next chapter will describe an experimental workshop offered to a group of such writers, its background, process and results.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

CHAPTER FIVE

LEXICAL BUNDLES IN NOVICE ACADEMIC WRITING

262

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

5.1. INTRODUCTORY REMARKS

The previous chapter addressed the use of lexical bundles in psychology, literary studies, and linguistics. It showed that the bundles shared by the three disciplines are not particularly numerous, and that both the density and the composition of multi-word sequences in the three disciplines differ greatly. At the end of the previous Chapter I also argued that both EAP classrooms and individual learners would benefit greatly from the introduction of corpus-based teaching techniques, particularly in the teaching of lexical bundles. In this Chapter, I will describe the methodology devised to help both EAP instructors and individual learners to improve their writing skills through the use of these techniques.

As was shown in Chapter two, the challenge of the “native-like selection” (Pawley & Syder, 1983) lies in the scope of phraseology, i.e., in the native-like selection of lexico-grammatical patterns rather than of isolated words or grammatical features (Römer, 2009). The recent years have witnessed a great deal of high-quality research concerned with the study of formulaic expressions from a variety of perspectives, including academic writing, learner writing, genre-based approaches, Intercultural Rhetoric and so on.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Among the different types of formulaic expressions, one particular type, lexical bundles, attract much attention due to the pervasiveness and ubiquitous character of the formulaic sequences of this type and the important communicative and discourse-building functions they perform. Studies of lexical bundles in learner writing demonstrate the apparent difficulties that these formulaic expressions suppose for novice and L2 writers (Cortés, 2002, 2004, 2008; Huang, 2015; Salazar, 2011, 2014).

Despite their seeming simplicity and semantic transparency, these formulaic sequences show considerable variation in the writing of L1 and L2 researchers, as well as in expert and novice writing. It has been argued that a particular genre can be identified, among other features, through the analysis of lexical bundles it contains (Hyland, 2008b). Successful academic writing is expected to comply with the generic conventions and expectations of the discourse community, and that expertness is a status achieved through producing texts which are highly prototypical of the given genre and are further validated by the discourse community (Swales, 1990). In this context, the acquisition of lexical bundles becomes an important part of the genre-specific features which need to be acquired by the novice and aspiring members of the discourse community.

A range of studies have addressed the issue of teaching lexical bundles or generating pedagogically useful lists for learners, instructors, and novice researchers (Ackermann & Chen, 2013; Cortés, 2006; Erman et al., 2013; Salazar, 2011, 2014; Shahriari, 2017; Shrestha, 2017; Simpson-Vlach & Ellis, 2012). Many studies have

also shown the importance of teaching the lexical bundles which the learners are likely to encounter in their coursework or reading materials (Byrd & Coxhead, 2010; Hyland, 2008b). This approach implies that the list of lexical bundles used for teaching should be extracted from texts representing the genres and the discipline the learners are studying (Eriksson, 2010; Farvardin, Afghari & Koosha, 2012; Hyland, 2008a), as was shown in the previous Chapter.

The acquisition of lexical bundles is important for a number of reasons:

- lexical bundles are retrieved directly from authentic texts,
- they are the most frequent units occurring in a given register or genre,
- they perform important discourse functions,
- they are objective corpus-driven data and do not depend on intuitions or guesses of teachers and course authors,
- although not structurally complete, they have structural and functional correlates which can facilitate learning,
- they are genre- and discipline-specific, thus belonging within a particular framework and taught with particular purposes, which makes them an ideal target for advanced level EAP classes focussing on certain contexts and environments.

However, despite their significant role in discourse, formulaic sequences in general and lexical bundles in particular are not given much attention in teaching contexts (Granger & Meunier, 2008a). Kennedy (2008) explains this by several possible

reasons, namely: the fuzziness of the concept of multi-word units, the tension among language teachers regarding form-focussed and message-focussed approaches, the absence of a “tried and true ‘method’ for teaching phraseology” (p. 38), the idea of phraseology being best learned “implicitly” (p. 38), a number of open questions which still remain unanswered in phraseology, the uncertainty as to how much exposure and in what form is needed for the language learners to acquire phraseological knowledge.

Coxhead (2008) analyses the reasons for the fact that these extremely frequent sequences are not easily acquired by learners and names the following:

- students do not want to spend their time on memorising formulaic sequences,
- students perceive learning formulaic sequences as a more difficult task than learning isolated words,
- students lack proper knowledge of the words constituting the formulaic sequences they are expected to learn,
- students are unwilling to take risks by using an expression which is new to them and prefer to rely on more familiar words,
- students think they are expected to produce a particular type of language, in particular, they hesitate to "sound academic" (p. 158),
- students are concerned about plagiarism and think they should not replicate word combinations they find in other authors' work.

Coxhead concludes by suggesting several strategies and approaches for teaching lexical bundles based on advances in phraseological research, using the tools and techniques available in classrooms, such as the digital environment, and adapting the teaching contents and forms to the learners' needs. Besides, Coxhead suggests that "[l]istening to student voices is important" (p. 159).

Therefore, a learner-centred methodology would seem appropriate for the acquisition of lexical bundles. The methodologies used to teach lexical bundles do not seem to render particularly encouraging results, as was shown in Chapter two. Formulaic sequences have often been taught as if they were simple vocabulary items, using exercises and activities typically employed to teach lexis, mostly based on Nation's (2000) widely cited three stages of vocabulary learning: noticing, retrieval, and generating. Furthermore, most pedagogically-oriented studies have relied on ready-made lists which were presented to learners as finite and complete sets, with the tacit implication that there is nothing more to learn in terms of lexical bundles in the given genre. This approach may induce the learners to regard lexical bundles as a closed set of linguistic items and to treat them as the object of the post-training test, rather than a useful resource for building their discourse.

Lexical bundles typically contain frequent words with which learners are likely to be familiar, hence, the task of acquisition does not seem to be related so much to learning lexis as lexico-grammar. Furthermore, the unwillingness of students to memorise whole sequences is understandable, but it does not seem to be a question of

memorising so much as one of using the familiar expressions, the final stage in Nation's (2000) framework. Durrant (2009) observes that:

“the fact that the words are superficially familiar to learners may make them all the more problematic, since learners may not even notice when they have not understood them properly.” (p. 164).

This does not seem to be a problem in the acquisition of lexical bundles, since, as was mentioned above, they are not composed of particularly complicated words.

Another aspect of the trainings focussed on teaching lexical bundles is that they typically include theoretical material, such as their syntactic structure or functions. The latter, as we have seen in the previous chapter, is an extremely subjective concept and is unlikely to be of practical use to novice researchers of sciences other than linguistics.

On the other hand, teachers and researchers in the field of EAP have developed a number of tools which seem to be more appropriate for advanced learners, such as corpus-based methodologies designed both for self-directed learning and EAP classroom use. Two publications are remarkable in this respect: Thurston and Candlin's (1997) workbook for essay writing, and Cargill and O'Connor's (2009) guide on writing research articles. The former uses concordance lines to illustrate the use of the words selected as learning material. The latter contains clear and helpful guidelines for compiling and analysing a personalised corpus to improve novice researchers' writing skills.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

These two publications prompted the idea of instructing novice L2 researchers to collect their own mini-corpora comprising texts which they would consider representative of their target genre and discipline, and analyse them for specific features using corpus analysis software. It was hoped that this approach would raise the students' awareness of lexical bundles and would help them to improve their writing skills significantly. It was thought that by looking at the frequencies of lexical bundles novice researchers could decide how frequently a given bundle should be used in their writing, while consulting concordance lines might give them a better understanding of the collocates and usage patterns of the formulaic expressions.

Such an approach would make it possible not only to 'listen to student voices' (Coxhead, 2008, p. 158), but also to enhance their autonomy as learners. It would also make it possible to look at sets of sequences instead of learning one bundle at a time, an important advantage in the conditions of constant time pressure in university classrooms. Another advantage of this approach would be the holistic picture obtained from analysing a whole corpus, albeit a small-scale and personalised one.

The idea of the workshop was partly based on the Vygotskian (1987) concept of the Zone of Proximal Development. It was thought that if the learners were provided with adequate tools and an efficient methodology for small-scale corpus analysis, they would be able to improve their writing skills continuously and in a self-directed way, and could gradually gain expertise by replicating the phraseological patterns used by successful members of their target discourse community.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The results of this experimental study are reported in this chapter.

The study involved two phases: a preparatory stage, whereby L2 novice academic writing was analysed in order to reveal the patterns of usage of lexical bundles, and the workshop which was aimed at instructing L2 writers to collect and analyse their own mini-corpora based on the ‘model’ texts highly prototypical of their discipline and topic.

The next section will describe the preparatory stage of the study focussing on an analysis of the learner writing in the selected disciplines.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

5.2. LEXICAL BUNDLES IN L2 NOVICE ACADEMIC WRITING

The preliminary study consisted in identifying the characteristic features of L2 novice writing through a small-scale contrastive study. For this purpose, the PhD corpus was analysed using AntConc (Anthony, 2010) to identify the most frequent clusters and use them as target bundles for a contrastive analysis of the TFG'15-16 corpus which comprised TFGs written by the students of the ULL in the previous years. The lexical bundles identified in the native novice (PhD) subcorpus were searched in the L2 texts to determine their occurrence and usage patterns. The contrastive analysis of the two comparable subcorpora made it possible to obtain a general understanding of the lexical bundle usage strategies in L2 academic texts.

The search for lexical bundles resulted in a total of 84 bundles which in turn were analysed to remove any unrelated or uninformative items. The final list of L1 Novice bundles is presented in Table 5.1.

Table 5.1. The most frequent lexical bundles in the L1 novice corpus.

3-word bundles	Freq. p.m.w.	4-word bundles	Freq. p.m.w.	5-word bundles	Freq. p.m.w.
as well as	441	at the same time	181	at the end of the	59
in order to	397	the development of the	122	in the context of the	44
the development of	392	the end of the	113	at the center of the	39
in terms of	318	in the context of	108	at the start of the	39
the nature of	230	as well as the	98	in the [year]s and [year]s	34
a number of	225	the way in which	98		
part of the	225	the [year]s and [year]s	93		
the concept of	220	the extent to which	88		
in which the	205	the nature of the	88		
in relation to	196	in terms of the	83		
one of the	191	at the end of	78		
some of the	176	the context of the	78		
the fact that	167	in the same way	69		
the end of	162	on the other hand	64		
as part of	157	the use of the	64		
such as the	157	at the heart of	59		
the relationship between	157	in the case of	59		
the work of	151	the influence of the	59		
as well as	441	at the same time	181		
in order to	397	the development of the	122		

As can be seen from the Table, the L1 novice bundle list shares many bundles with the expert corpus (discussed in Chapter four), and the frequencies of many of the bundles are similar to those in the Expert corpus.

These bundles were then searched in the L2 novice corpus to determine cases of overuse, underuse, or unusual collocates. The L2 novice (TFG'15-16) subcorpus and the L1 novice (PhD) subcorpus are described in detail in Chapter three and in Appendix 1.

The analysis of L2 novice writing offered several insights which helped to shape and focus the workshop offered to a group of students who were writing their TFGs.

Below are the lists of underused, overused² and equally used bundles. A brief discussion of the identified unusual collocates follows.

It should be noted that these two subcorpora were contrasted only to reveal general patterns of usage and reach an understanding of L2 strategies of lexical bundle usage in order to obtain a point of reference for the introductory session on lexical bundles.

As can be seen from Table 5.2, 41 bundles in total appear less frequently in the L2 corpus than in the texts of the L1 writers. Some bundles appeared in the L2 texts, but upon close examination turned out to be parts of quotations, as in the case of *at the centre of* and *the extent to which*. The latter was used 4 times in the L2 corpus, all of which were quotations, with 3 instances accounting for the same quotation in different texts.

² The terms 'underused' and 'overused' in this study refer to lexical bundles which show strong deviation from the L1 subcorpus. By 'strong deviation' a difference of at least 40% is meant. However, no statistical tests were performed to determine whether the deviation is statistically significant, as these data were used exclusively for the purpose of obtaining general information regarding L2 novice writers' phraseological competence.

Table 5.2. Bundles underused in the L2 corpus.

Bundle	L1 corpus	L2 corpus	Bundle	L1 corpus	L2 corpus
the development of	80	41	in this way	29	10
in terms of	65	45	discussion of the	21	4
the nature of	47	7	understanding of the	23	16
a number of	46	22	the form of	20	7
the concept of	45	27	the heart of	20	1
the relationship between	32	15	in the context of	22	8
as part of	32	18	the extent to which	18	0
the work of	31	7	at the heart of	12	0
a variety of	25	10	the influence of the	12	3
the notion of	24	10	the majority of the	12	5
of the world	23	12	at the centre of	11	0
the construction of	23	4	the centre of the	11	2
in this study	22	3	at the start of the	8	0
of the new	22	9	within the context of	10	0
a form of	21	3	in the context of	9	1
a set of	21	8	at the centre of the	8	0
many of the	21	9	as will be discussed in	5	0
of the author	21	7	ideas about the nature of	4	0
the field of	21	14	it is interesting to note	4	0
the focus of	21	4	on the part of the	4	1
the majority of	21	7			

The bundle *the heart of*, which also showed a high frequency in the expert corpus, seems to be a favourite of L1 PhD writers. It collocates with abstract nouns such as *problem, distinction, journalism, ventures, social practices, teacher-pupil interaction*, etc., and with verbs like *attack, strike, pulse, get to*. In the L2 corpus this bundle occurs only once and is used in its almost literal meaning: *the heart of Harlem*.

The bundle *on the part of the* collocates with animate beings, primarily humans in the L1 corpus: *The determination on the part of the joint authors; interrogation primarily on the part of the interviewer*, etc. In the L2 corpus its only instance appears in company with inanimate objects: *the change of state is simultaneous to a change of position on the part of the first element*.

Table 5.3. Bundles used more frequently in L2 corpus.

Bundle	L1 corpus	L2 corpus	Bundle	L1 corpus	L2 corpus
in order to	81	138	the idea of	25	35
part of the	46	80	the importance of	23	43
one of the	40	160	on the other hand	13	32
the fact that	34	75	the study of	21	44
some of the	36	60	in other words	20	31
there is a	29	53	at the end of	16	45
in the same	27	39	in the case of	12	59
use of the	26	52	at the end of the	12	28
end of the	25	50	the idea of	25	35

Table 5.3 may be read as exemplifying the phenomenon of the “phraseological teddy bears” (Salazar, 2011) after Hasselgren’s (1994) famous label, “lexical teddy bears”. L2 and L1 novice writers have been shown to rely heavily on a limited number of lexical bundles based on their primings, and neglect others, and it is not always clear why this or that bundle has been chosen as a ‘teddy bear’. In some cases, these choices may be due to L1 transfer (Paquot, 2013; Rica, 2010), or to previous explicit instruction (Lindstromberg, Eyckmans, & Connabeer, 2016), in others no viable explanation can be offered.

A group of target bundles was found to occur with similar frequency in the L1 and L2 subcorpora. These are presented in Table 5.4 below.

Table 5.4. Bundles used with similar frequencies in L1 and L2 subcorpora.

Bundle	L1 corpus	L2 corpus	Bundle	L1 corpus	L2 corpus
as well as	90	77	the process of	30	25
the use of	70	68	aspects of the	22	15
in relation to	40	49	at the same time	37	42
it is not	38	33	the way in which	20	17
the role of	36	38	in the same way	14	13
such as the	32	25	in the same way that	5	3

The use of these bundles in the two subcorpora is surprisingly similar. It seems that not only the frequencies, but also the habitual collocates tend to converge in the two groups. One exception is the bundle *aspects of the*, which in L2 writing collocates with *United States, Women's Movement, teaching implementation, teaching process, textbook's content, second language, Spaniards, region, unconscious, life, novel, community, 1950s, and American society*. The L1 corpus, on the other hand, offers such collocates as *context of practice, changes in higher education, system, individual artist's approach, thesis, mind's reaction to experience, study, playing of McCoy Tyner, path taken at this stage, two activities, community programme, investigation, generative movement, paradigm, natural or biological world, formal sciences, written-up interviews, literary interview*. The use of the verbs in the two

subcorpora is similar: *show, reveal, represent, evaluate, criticise*, etc. in L2 and *examine, reflect, outline*, etc. in L1.

The patterns of use of *the way in which* mostly coincide in the two subcorpora, with the exception that in L2 corpus *the way in which* is followed by the personal pronoun *we* on two occasions.

The examples above allowed me to gain an understanding of the differences between L1 and L2 novice writers' use of lexical bundles, and particularly, to develop an understanding of my expectations with respect to the group of learners I was going to work with. The L2 corpus selected for this purpose was produced by a comparable group of undergraduate students belonging to the institution and department where I would deliver the workshop on lexical bundles.

These L1 texts were not intended to be taken as models for the L2 group to imitate, although they represented a level of academic literacy which most of the workshop participants would presumably aspire to achieving as the next step in their academic development. The aim of this small-scale study was to draw a comparison between the two groups of novice researchers in order to determine whether there was a gap between the two groups in terms of their language evolution (Biber, Gray & Poonpon, 2011), and to obtain a general idea of the extent to which the two groups diverge in the use of lexical bundles. I chose to use the lexical bundles extracted from the PhD corpus because there clearly was a gap, and it could be reduced by using the PhD corpus data by way of examples. On the other hand, using the Expert corpus could

make the first encounter with the concept of lexical bundles somewhat intimidating for the workshop participants. This is why novice L1 writing was selected as a basis for the introductory session.

Based on the lists of underused and overused bundles, I designed several activities and a PowerPoint presentation for the introductory session. The handout and the presentation can be found in Appendix 4.

Below is a description of the introductory session and the workshop.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

5.3. INTRODUCING LEXICAL BUNDLES TO LEARNERS

5.3.1. The introductory session

The first session on lexical bundles with 13 undergraduate students included a brief introduction to the concept of lexical bundles in academic writing followed by a number of activities designed to raise the awareness of the students regarding the nature of lexical bundles. This introduction highlighted the definition, structure, and functions of lexical bundles, as well as their role as building blocks of discourse (see Appendix 4 for more detail).

The first activity consisted in examining two lists of lexical bundles, one of which contained some of the bundles found to be underused by the undergraduate students of the previous two years, whereas the second list contained several lexical bundles which were found to be overused by the same writers. Table 5.5 shows the bundles included in this activity. The lists were labelled ‘List 1’ and ‘List 2’, and the students did not know what categories they represented. The undergraduate participants of the sessions were asked to judge the two lists in terms of their relevance to academic writing and to determine which of the sets of lexical bundles would be most likely to appear in an academic text. The students almost unanimously judged the overused

bundles list as ‘more academic’. This finding suggests that the participants of the session would be likely to repeat the practice of relying heavily on the bundles overused and underusing the sequences used rarely by their predecessors. After that the handout was given to the students, and they were asked to comment on the underused bundles and say if they would use them in their own writing or not. Most of the underused bundles, according to the participants, were unlikely to appear in their own writing.

Table 5.5. Bundles used in Activity 1.

<p>LIST 1 the development of a number of the concept of the relationship between as part of the work of a variety of the notion of the construction of a form of a set of</p>	<p>LIST 2 in order to part of the one of the the fact that in the same use of the the idea of the importance of the study of in other words in the case of</p>
---	--

Some of the bundles aroused doubts as to their usage and relevance, as in the case of *the form of*, but the absolute majority were reported as clear, easy to understand and quite usable. *The heart of* was labelled as the least ‘academic’ of all: the students seemed to interpret it literally and as referring to feelings rather than to the essence or the central and most important part of a concept or phenomenon. This explains the

total absence of this lexical bundle in L2 writing while in L1 dissertations it occurs 20 times per 200.000 words, and in the Expert corpus it occurs 17 times in 13 different texts (approximately 21 times per million words).

The next activity was based on 4 sentences from published academic research containing four of the target bundles and the same sentences with a slight modification and with omitted target lexical bundles (see Appendix 4). The students were asked to judge which of the sentences were more likely to appear in a published academic text. They judged the original sentences containing the target bundles as more likely to appear in a published text. When asked if they would use these target bundles in their own writing, most of the students admitted that it was not very likely.

The final activity was of traditional ‘fill in the gap’ type, with eight sentences from the Native-speaker PhD dissertations corpus in which the target bundles were substituted with a gap (see Appendix 4). The participants first discussed each target bundle and its possible functions, after which they filled in the gaps. The students were also asked to justify their choice of each lexical bundle, thus inducing them to reflect on the nature and possible functions of the target bundles. When asked whether it had seemed difficult for them, they judged the activity as easy and simple to perform. However, many of them thought that they would not use the target bundles in their writing.

The students were given a list of target bundles, as well as an abridged list of overused bundles, and instructed to remember to use the underused ones and substitute the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

overused ones where necessary. Additionally, they were invited to take part in the workshop to learn to collect their own mini-corpora and generate personalized lists of lexical bundles for their TFGs.

This introductory session provided several insights. On the one hand, it is obvious that novice writers do not consider lexical bundles to be particularly difficult to understand, remember or use. However, the participants of the session would be likely to follow the patterns of usage adopted by the TFG writers of the previous years, underusing a wide range of multi-word sequences and overusing a handful of ‘phraseological teddy bears’ (Salazar, 2011). The reasons why a particular phrase becomes a ‘teddy bear’ are unclear, but the following tentative explanations can be suggested:

- The students perceive some lexical bundles as more ‘scholarly’ than others, perhaps due to the salience of these lexical bundles, the existence of L1 equivalents, or depending on the time and context of having acquired them,
- Novice writers are not aware of the repetitions of the same phrase across their writing and hence do not take action to avoid it,
- The students think they need to produce novel language and are thus reluctant to use expressions which seem too familiar to them,
- Novice writers, especially in the humanities, feel that they need to be original and use complicated and elaborated language, therefore, they avoid the use of expressions which sound like clichés to them,

- TFG writers are anxious about the number of words they need to produce, and where possible, prefer lengthy expressions rather than using single words or shorter expressions (consider the great overuse of *in order to* instead of *to*).

These are only some of the possibilities that need further investigation if we are to fully understand the use of lexical bundles in L2 student writing.

Another introductory session was held with a much smaller group of 6 students, also TFG writers. This session was much shorter due to time constraints and included the brief introduction and a discussion of the activities prepared for the session. Two out of six students said that the bundles in list 1 (which had been found to be underused by L2 writers) looked more academic, while two other students found both lists acceptable. The bundle (*at the heart of*) was labelled as ‘not very academic-sounding’ by this group as well.

After discussing the bundles, the students were given the list of target bundles together with a small list of the most overused lexical bundles. It was suggested that they pay conscious attention to the bundles on the list at the stage of editing their TFGs by running the search utility in their word processing software and consider adding or replacing some of the bundles according to their needs and purposes.

The participants had no previous knowledge of lexical bundles and showed considerable interest in learning about them. The participants reported that the presentation and the information provided were extremely useful to them and expressed willingness to follow the recommendations.

The introductory session provided crucial information regarding the students' knowledge and awareness of lexical bundles. For instance, it revealed that the students had no difficulty in understanding the meaning of the lexical bundles, arguably with the exception of *(at) the heart of* whose literal meaning was obviously known to them, but they expressed doubts as to its usage.

The information obtained during the introductory session helped to shape the workshop as an autonomous learning activity, as will be described below.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

5.2.2. The lexical bundles workshop

Prior to the workshop, an introductory email was sent to all the participants containing a brief explanation of the nature and characteristics of lexical bundles and instructions to download AntConc and AntFileConverter (Anthony, 2010, 2015), the two pieces of free software which were to be used during the workshop. The email was sent in English to the English Language and Literature students, and in Spanish to the PhD students from the Department of Psychology. The texts of both emails with the instructions can be found in Appendix 4.

The workshop was intended for advanced undergraduate students of the Department of English and German Philology of the University of La Laguna, and for PhD students of the Department of Psychology of the same University. Only a proportion of the TFG writers participated in the introductory session, but all the participants received preliminary information on lexical bundles through the introductory email.

The aim of the workshop was to instruct the participants to compile a personalised mini-corpus of ‘model’, prototypical texts in their topic and to use the cluster tool of AntConc corpus analysis software to analyse those texts for lexical bundles.

During the workshop, the participants worked on their own mini-corpora to extract a list of bundles which were later commented on by the group.

The workshop was conducted several times with different groups of participants. All the groups were instructed to download the papers or book chapters which they thought representative and worth imitating, to convert them into .txt files and remove most of the text which was unrelated to the actual content of the article or book chapter, such as headings, footnotes, references, etc. It was not possible to clean the text files entirely as it would have taken up too much time, but the greater part of the extratextual elements were removed.

After cleaning the files from irrelevant elements, the participants pasted the texts into a word processing file to obtain the word count, opened the files in AntConc and specified the search criteria (size of the clusters, minimum frequency, and minimum distribution). Bundles of different lengths were searched separately, i.e., separate searches were conducted for three-, four-, and five-word bundles with different frequency and distribution parameters.

The participants were asked to comment on the most interesting findings and consult the concordance lines for the ones that seemed interesting to them.

Informal feedback was collected after the workshop. Most of the participants reported no or very little prior knowledge of the existence of lexical bundles and pointed out that both the workshop and the introductory session were extremely useful. Some participants observed that the software would be useful for analysing their own writing.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

A formal feedback questionnaire was also sent to the participants, to which only six people responded. The opinions of all the respondents were extremely favourable: the workshop was rated as ‘Good’ or ‘Very good’, the lexical bundle approach seemed ‘Very useful’ or ‘Useful’, 3 respondents thought they would use the technique in the future, while the other 3 reported that they were already using it. The reasons for such a low response rate are unclear, but a tentative explanation is that the students might have been preoccupied with completing their TFGs and did not have time to respond.

The overall number of the participants was 32, of which seven TFG writers submitted their texts for analysis. Several PhD students submitted their lists of lexical bundles extracted during the workshop, and one participant submitted pre- and post-workshop abstracts which will be discussed below.

The next section reports the use of lexical bundles by the workshop participants.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

5.4. LEXICAL BUNDLES IN THE WRITING OF THE WORKSHOP

PARTICIPANTS

The texts submitted by the workshop participants were searched for underused and overused lexical bundles in order to determine whether there was any change in the use of lexical bundles.

Overall, very little variation was observed with respect to the writing of the undergraduate students of the previous years who had not participated in the workshop. These results are summarised in Table 5.6.

The frequency of 14 bundles has increased substantially (*a number of, the concept of, it is important to note, the context of, in this way, etc.*), while in about 25% of the underused bundles (11 types) the frequency has decreased (*understanding of the, the form of, at the centre of, etc.*). There is no apparent correlation with any variable: none of these bundles is topic-specific, they perform different functions, some are characteristic of expert writing, but others are not, some were included in the list of suggested target bundles, but others were not. In addition, the size of the TFG'17 subcorpus is, unfortunately, too small to make any viable generalisations.

Table 5.6. Change in the frequency of underused bundles in the writers who participated in the workshop.

Bundle	L2'17		L2 15-16		Bundle	L2'17		L2 15-16	
	Freq.	p.m. w.	Freq.	p.m. w.		Freq.	p.m. w.	Freq.	p.m. w.
the development of	12	160	41	170	the majority of	1	13	7	29
in terms of	5	67	45	187	in this way	10	133	10	41
the nature of	3	40	7	29	discussion of the	0	-	4	17
a number of	9	120	22	91	understanding of the	2	27	16	67
the concept of	10	253	27	112	the form of	1	13	7	29
the relationship between	2	27	15	62	the heart of	0	-	1	4
as part of	3	40	18	75	in the context of	4	53	8	33
the work of	1	13	7	29	the extent to which	0	-	0	-
a variety of	3	40	10	41	at the heart of	0	-	0	-
the notion of	9	120	10	41	the influence of the	1	13	3	12
of the world	8	107	12	50	the majority of the	0	-	5	21
the construction of	9	120	4	17	at the centre of	0	-	0	-
in this study	3	40	3	12	the centre of the	0	-	2	8
of the new	1	13	9	37	at the start of the	0	-	0	-
a form of	6	80	3	12	within the context of	0	-	0	-
a set of	7	93	8	33	at the centre of the	0	-	0	-
many of the	10	133	9	37	as will be discussed in	0	-	0	-
of the author	6	80	7	29	ideas about the nature of	0	-	0	-
the field of	4	53	14	58	it is interesting to note	1	13	0	-
the focus of	1	13	4	17	on the part of the	0	-	1	4

The data for overused bundles are summarised in Table 5.7.

Table 5.7. Change in the frequency of overused bundles in the writers who participated in the workshop.

Bundle	L2'17		L2 15-16		Bundle	L2'17		L2 15-16	
	Freq.	p.m.w.	Freq.	p.m.w.		Freq.	p.m.w.	Freq.	p.m.w.
in order to	72	960	138	676	the idea of	10	133	35	172
part of the	14	187	80	392	the importance of	9	120	43	211
one of the	57	760	160	784	on the other hand	10	133	32	157
the fact that	29	387	75	368	the study of	1	13	44	216
some of the	17	227	60	294	in other words	2	27	31	152
there is a	24	320	53	260	at the end of	12	160	45	221
in the same	12	160	39	191	in the case of	18	240	59	289
use of the	10	133	52	255	at the end of the	4	53	28	137
end of the	12	160	50	245					

Of the 17 overused bundles, eight are used more moderately than in the previous years. However, the frequency of three bundles has increased, while the remaining six are mostly unchanged. This finding may indicate that learners have become more conscious regarding their writing and are more careful not to use the same expression many times. It can also be argued that some of the bundles which are used more cautiously have equivalents or semi-equivalents, therefore, the learners have been able to replace them. However, the closest equivalent of *in order to* is a single word, *to*, and the students may have decided to use the longer expression considering it

more formal (or trying to fill up pages with more words), hence the increase in the frequency of this bundle.

Ten of the bundles showing positive variation were on the overused bundles list presented during the introductory session. This is an encouraging finding showing a conscious effort on the part of the learners to reduce their overuse of bundles.

This is a tentative interpretation, and further research is needed to test it, although two observations should be made at this point with respect to research on raising awareness and increasing proficiency in the context of an EAP classroom.

Firstly, a study aimed at obtaining measurable and absolutely objective results is impossible to conduct due to the multiple input channels present in any environment. Any learner (and any human being, for that matter) receives all kinds of input from very different sources, and one cannot make absolute claims regarding the origin of the knowledge or skills acquired by learners. The workshop participants received input during the introductory session and the workshop, but they also received other types of input, particularly, through reading professional literature and attending other classes and seminars, looking up information online, sharing information with peers, etc. Thus, although the workshop participants can be said to have slightly improved their use of lexical bundles (particularly in the overused bundles), we cannot claim that it was a direct consequence of the workshop, as they were exposed to other types of input, and their primings could have changed as a result of other factors. Hoey (2005) convincingly argues that our primings change constantly as a

result of numerous and diverse interactions with all kinds of linguistic environments. This could have been the case with the workshop participants, whose lexical bundles usage could have been changed as a consequence of reading academic texts, or taking a course in academic writing, or any other kind of input. For example, during the writing course they did as part of their TFG preparation, they were explicitly instructed to use the *Manchester Academic Phrasebank* (Morley, Phrasebank.manchester.ac.uk, n.d.) as a resource for constructing their discourse, and this instruction may have produced a change in their writing strategies. A brief description of the *Manchester Academic Phrasebank* can be found in Section 5.4 below.

The second observation has to do with the fact that linguistic evolution is a life-long endeavour and should be studied as such. An ideal large-scale longitudinal study should comprise a period of time superior to ten years, for instance, beginning in the first year of college and stretching on to a post-Doctoral position. It should also cover a large number of subjects, since it cannot be predicted which of the participants will continue through to a post-Doctoral position and who will drop out at some stage of the experiment. Even in such a short-term study as the one reported here, many of the participants dropped out of the experiment, for a variety of reasons. An ideal study concerned with measurable results should ensure that all the other variables are unchanged so as to be able to attribute the results to the variable of teaching input. In other words, an ideal study is impossible. An attempt to describe a case study of the

linguistic evolution of one of the participants is made in the next section, in spite of the above-mentioned caveats.

This notwithstanding, in the case of the present study we can say that there is a certain degree of comparability between the TFG'15-16 and the TFG'17 subcorpora in the sense that the number of possible variables is reduced as much as possible, although the difference in the results is difficult to interpret without more information.

This section examined the results of the workshop by contrasting the frequency of target bundles in the writing of the workshop participants with that of learners who had not participated in the workshop. The findings show a little improvement in the underused bundles and some improvement in the overused bundles. However, the size of the corpus does not allow for generalisations.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

5.5. A CASE STUDY OF LINGUISTIC EVOLUTION

This case study was inspired by the personal experience of one of the participants of the workshop who gave a personal account of his linguistic evolution and explained the factors that had influenced it.

Kevin (the name has been changed for privacy considerations) began his studies in the department of English and German Philology at the ULL. As an undergraduate student, he was a very self-confident young man and never doubted his linguistic competence. By contrast, some of his instructors were aware of certain lacunae in his productive knowledge: they thought his English pompous and baroque. They tried to suggest improvement techniques, but their suggestions were politely rejected.

During Kevin's MA studies, one of the professors made an explicit comment regarding the oddity of his written and spoken production, which triggered Kevin's pursuit of better strategies. He began to consciously search for model texts and ways to improve his production. One of his techniques was copying model texts by hand. The meticulousness and tediousness of this task shows Kevin's determination to improve his writing: he was prepared to make extraordinary effort to achieve his goal.

In the course of his search for improvement techniques, Kevin turned to the *Manchester Academic Phrasebank* (Morley, Phrasebank.manchester.ac.uk), a corpus-based resource designed for novice researchers seeking helpful phrases to organise their writing. The Manchester Academic Phrasebank was inspired by the work of John Swales (1990) and his description of the rhetorical structure of the research article. The website is divided according to the IMRAD pattern reflecting the sections of a research article (Introduction, Methods, Results and Discussion), and each of the sections is organised on the basis of possible moves and steps typically present in each of the sections. Each move or step is represented by a number of expressions which perform the rhetorical function of the move, e.g., “Establishing the importance of the topic for the world or society”, which would correspond to the Swalesian Move of “Establishing a niche” (Swales, 1990) is followed by a number of phrases such as "X plays a vital role in the metabolism of ...", "X plays a critical role in the maintenance of ...", "Xs have emerged as powerful platforms for ...", "X is essential for a wide range of technologies.", etc. The website also provides helpful expressions based on the communicative functions typically realised in research articles, such as ‘Being critical’, ‘Citing previous research’, ‘Expressing caution’, etc.

Kevin began to use the *Manchester Academic Phrasebank* which, according to his testimonial, helped to structure his writing and, more importantly, his thinking. He found phrases that prompted him to reconsider how to organise his texts, guiding him

through the rhetorical structure of the paper and highlighting the next step he should make.

Both Kevin and his professors observe that his writing has improved substantially as a result of conscious search for improvement and a number of successful strategies.

He showed great interest in the workshop and worked through it very actively. After the workshop he sent me two abstracts which he had written after attending the workshop. one of the abstracts was 229 words long, the other had 404. While the shorter did not contain any lexical bundles, the second abstract (the longer one) contained three of the target bundles identified in the PhD subcorpus: *based on the* (occurring twice), *relationship between the*, and *as a result of*. This is another encouraging finding, especially given the background of this student.

The case of Kevin is somewhat ambiguous. It is not clear whether his own personalised corpus contained any other bundles that he had used, but I was unable to detect them due to their absence in my data. On the other hand, three lexical bundles cannot constitute a basis for generalised, though their presence in Kevin's texts is significant. The fact that the shorter abstract does not contain any lexical bundles sequences that I could identify may be due to its length, but other explanations are possible.

Kevin is still on his way to fluent and expert-like writing, and it would be interesting to examine the outcome of his efforts in the future, as it has been done hitherto.

It would also be instructive to obtain more self-reflections of this type from other learners. This could be an issue for further research.

This chapter described the findings obtained from the introductory session and the workshop on lexical bundles, as well as presented a brief case-study on one of the participants' linguistic evolution. The next chapter will discuss and summarise the results of the present study.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

CHAPTER SIX

DISCUSSION AND CONCLUSIONS

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

6.1. LEXICAL BUNDLES IN PSYCHOLOGY, LITERARY STUDIES, AND LINGUISTICS

As stated in the introductory Chapter of this thesis, the purpose of this study was twofold. Firstly, its goal was to identify and analyse lexical bundles in Psychology, Literary Studies, and Linguistics research articles and outline general patterns of usage, differences and similarities between these three disciplines. The second broad objective was to devise a learner-centred methodology for novice L2 writers aimed at developing their phraseological competence and improving their writing skills.

This chapter will address the research questions specified in chapter one and attempt to provide answers to each of them based on the results obtained from the study.

Section 6.1 addresses the frequency, structure and use of lexical bundles in Psychology, Literary Studies, and Linguistics research articles.

Section 6.2 discusses the results of the introductory session on lexical bundles, the workshop, and presents a case study of one of the participants.

6.2. THE FREQUENCY, STRUCTURE, AND FUNCTIONS OF LEXICAL BUNDLES

This section will provide answers to the research questions posed in chapter one based on the findings of this study.

The first research question was concerned with the types of lexical bundles in each discipline. The findings of this study show that the three disciplines employ different sets of lexical bundles, sharing a relatively small number of types. The types of lexical bundles identified for each of the disciplines are shown in chapter four. The main finding of this chapter is that there is considerable variation across disciplines, even if we discard topic-specific bundles.

As was indicated in Chapter four, of the three disciplines, psychology clearly employs the largest number of bundles, both in type and token. This finding can be regarded as reflecting the view that lexical bundles are typically more abundant in natural sciences (Hyland, 2008a) than in arts and humanities. As has been observed in Chapter one, even though psychology is a social science, it is more akin to the natural sciences in some regards than it is to linguistics and literary studies. It was

thus my expectation that a social sciences discipline such as psychology would show a greater variety and number of lexical bundles as compared with the other two disciplines.

As we saw, psychology uses more lexical bundles with a considerably higher frequency than literary studies, which employs slightly fewer bundles far less frequently. Linguistics, on the other hand, relies on a small number of formulaic sequences but employs each sequence relatively frequently, as different from literary studies relying on a large number of types, but making rarer use of them, and from psychology, which counts on a large number of types and employs them frequently.

The similar type / token ratios in psychology and linguistics show that these disciplines share certain lexico-grammatical features. This finding is indirectly supported by the fact that Coxhead's (2000) Academic Word List places psychology key words in the section of arts arguing that this discipline has been shown to share more syntactic features with arts than with social or natural sciences (Biber, 1989).

Biber et al. (1999) state that three-word bundles are almost ten times more frequent than four-word bundles, and four-word bundles are in turn almost ten times more frequent than five-word bundles. However, the findings of this study deviate from this pattern: the number of three- and four-word bundles in all the three disciplines show a very small difference, although five-word bundles do occur with a frequency similar to the one indicated by Biber et al. This can be due to very conservative frequency cut-off points selected for the present study.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The research question regarding the structural and functional characteristics is also addressed in detail in chapter four. Findings confirm the phrasal character of academic prose (Ädel & Erman, 2012; Biber, Conrad & Cortés, 2004; Chen & Baker, 2010): the absolute majority of lexical bundles in all the disciplines are nominal or prepositional groups, referred to as ‘phrasal’ bundles. Moreover, 39% of the psychology bundles, 50% of the literary studies bundles, and 48% of the linguistics bundles contain the preposition *of*.

Lexical bundles in literary studies present an unusual result in terms of structural types. The phrasal nature of academic prose has been confirmed by numerous studies, but the number of ‘clausal’ bundles in literary studies is unexpectedly high: 91% of all the bundles are ‘phrasal’. This finding shows the complexity of literary studies research articles in terms of syntactic structure which relies heavily on lengthy nominal and prepositional phrases. As shown by Biber and Gray (2010), academic discourse is complex in the sense of employing unusually heavy nominal structures, but even for academic discourse 91% of nominal structures among the most frequently used ones is a very large amount. This finding explains the difficulties of novice literary studies writers in constructing their texts (and reading those of others!) and provides a powerful pedagogic resource both for instructors and learners.

The ‘clausal’ bundles identified in the corpus, besides being less numerous than the ‘phrasal’ ones, show one more important feature: the abundance of weak verbs – predominantly *to be* – and the use of modal verbs, particularly *can* and *should* (as in

it should be noted that). This finding may have important pedagogical implications for novice researchers, particularly in terms of supplying them with a stock of hedging devices.

The structural analysis of lexical bundles revealed a new category which has not been mentioned in any study that I am aware of, namely, the adjective / noun + noun structure. This can be explained by the topic-specific nature of all the bundles in this category: these structures may have appeared in other studies, but could be discarded, since most linguists discard topic-specific bundles as irrelevant for their purposes.

Another finding worth noting is the total absence of anticipatory *it*+ verb phrase / adjective phrase in the literary studies subcorpus. As explained above, these structures usually represent stance bundles and often function as hedging devices (e.g., *it is likely*). Their absence in literary studies signals an important feature of this discipline, and reflects the fact that the discipline is concerned largely with dealing with abstract concepts which are not gradable in terms of likelihood, possibility, or other similar continua.

Related to this finding is the unusually high number of bundles classified as ‘Other expressions’ in literary studies. This might suggest that literary studies writers employ other structural categories which are different from those occurring in previous research due to the fact that this discipline has been underrepresented in the studies of lexical bundles. Some of the possible structural categories for this discipline can be: *wh*-clause fragment, existential *there* phrase, or *and*-phrase. These

structures seem to serve the communicative purposes of this particular discipline better than those devised on the basis of natural and social sciences corpora and can shed light on the construction of meaning in literary studies. However, a much larger corpus would be needed to draw any conclusions. This is another question that merits a further study.

A somewhat unexpected finding is that psychology uses the largest number of participant-oriented bundles. It is worth noting that of the 15 participant-oriented (stance) bundles, only one contains a personal pronoun: *our understanding of the*, the other 14 instances being mostly introduced by anticipatory *it*, which is an indicator of depersonalised stance. However, literary studies also has only one bundle containing a personal pronoun, *as we have seen*, while among the linguistics bundles such pronouns are absent. Although the greater part of participant-oriented lexical bundles in psychology convey ‘depersonalised’ stance, the presence of such a substantial number of participant-oriented bundles is indicative of an intention to interact with the reader. This is in contrast to linguistics and literary studies which show a much smaller amount of participant-oriented bundles; only 3 types in each discipline were found. The scarcity of participant-oriented bundles contradicts the findings of Hyland (2008a), who showed that the proportion of participant-oriented bundles in applied linguistics is 18,6%. A possible explanation for this divergence is that Hyland’s corpus contained texts from a variety of genres; apart from research articles, it was composed of PhD dissertations and MA / MS theses.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

The number of research-oriented bundles in literary studies also seems to contradict Hyland's (2008a) finding regarding this category. Although his corpus did not include literary studies texts, he states that research-oriented bundles are more common in natural sciences than in the humanities. Three observations can explain this discrepancy: the subjective character of the functional classification, the abundance of topic-specific bundles in this category, and the inclusion of bundles of different sizes in the present study. As explained above, topic-specific bundles were not removed from this analysis because they are an important part of the discipline-specific discourse, the focus of this study. In Hyland's study, however, topic-specific bundles were removed. In addition, Hyland studied only four-word bundles, whereas I look at three-, four-, and five-word bundles.

The findings show that the 'rule of thumb' that many EFL learners have been taught to use does not always hold true in authentic data. Consider the example of *on the one hand* and *on the other hand* which in many EAP classrooms are taught as inseparable discourse organisers: the use of the former is supposed to inevitably trigger the latter. However, the data show that both can occur on their own, although *on the one hand* does so less frequently than *on the other hand*. Such examples can provide valuable information to L2 users in their struggle to produce a native-like and expert-like text. Concordance lines can be especially useful in this regard, as they reflect the primings of a word or a phrase in a reduced space (Hoey, 2005; Sinclair, 1991). For instance, the concordance lines for *at the same time* show that its usual collocates across the three subcorpora are *while*, *however*, and *yet*, indicating a clear

element of contrast, and not very often time reference. The concordance lines for *it is important to* reveal the most frequent collocates of this sequence to be *note, point out, emphasise, stress, acknowledge*, etc., a wealth of nearly-synonymic expressions for a desperate novice writer available at one (or two) clicks.

As for the bundles found in only one discipline, the main finding is that certain strings appear to be discipline-specific, but not topic-specific, i.e., several discipline-specific bundles are used only in a given discipline without any apparent reason. They could be used in many other disciplines, but for some reason are primed to occur only in one.

As was stated earlier, the number of unique lexical bundles in psychology is almost twice as large as that in literary studies and three times larger than in linguistics. This makes linguistics the least formulaic discipline in terms of types of bundles.

Many of these discipline-exclusive bundles are topic-specific, as was to be expected, though others are neutral sequences which could be intuitively predicted to occur in any discipline. This finding implies that the formulaic nature of a genre, register, or discipline (the quantity and semantic characteristics of formulaic sequences present in it) is determined not only by its communicative purposes, but also by the accepted practices of the discourse community.

It could be said that each discipline has its own set of sequences which are commonly used by the expert members of its discourse community with no apparent reason: widely popular common expressions selected due to the shared primings of the

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

members of the discourse community, and not based on the communicative purposes or any formal characteristics of the genre. Consider the presence of such bundles as *are consistent with the, and colleagues found that, it is likely that, our understanding of the, it is possible that, for a review see*, to name but a few. The latter is especially interesting as it does not occur at all in the literary studies or linguistics subcorpora, while the others did occur, although not frequently enough to be included in the lists generated for these disciplines. The semantic structure of these bundles does not prevent them from occurring in any other discipline, but the findings of this analysis show that expert writers in the field of psychology seem to have a special predilection for these bundles, whereas writers in other disciplines show preference for other sequences. This notwithstanding, it is important to note that the majority of these discipline-specific bundles are also topic-specific, and the number of unmotivated exclusive bundles, although not negligible, is smaller than that of the topic-specific bundles.

It is, of course, important to bear in mind that this study addresses only three disciplines, and the label ‘exclusive’ is used in relation to the other two disciplines, and not to the whole body of academic writing. It may be that these ‘exclusive’ bundles are rather frequent in other disciplines as well, clearly a subject for further research.

Although the findings of exclusive bundles cannot be generalised due to the limited range of disciplines studied, the tendency clearly shows that in each of the disciplines

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

a number of bundles are preferred over other similar expressions for reasons that I have not been able to identify. An explanation for this finding may be that these bundles have become part of the disciplinary communication strategies due to multiple repetitions, possibly in the work of expert members of the discourse community, and have been replicated by other members to the extent of becoming standardised. It seems that not only individuals, but also disciplines have their own primings which are passed down to subsequent generations of researchers and are institutionalised through continuous repetition. As observed in Chapter four, this is an issue for further research.

The findings of this study show variation in lexical bundles use, structure, and frequency across disciplines. Some of the parameters vary greatly, others do not seem to differ significantly, but overall these findings show the importance of a discipline-specific approach to lexical bundles.

The research questions regarding the pedagogical applications of the knowledge on lexical bundles were addressed in chapter five. Its findings are discussed below.

The small-scale contrastive study performed to identify L2 novice researchers' pattern of usage of lexical bundles confirmed the view that L2 learners tend to overuse certain bundles and underuse others (Ädel & Erman, 2012; Allen 2001; Chen and Baker 2010; Parkinson & Musgrave, 2014, etc.). Several cases of uncommon use or "misuse" (Pan, Reppen & Biber, 2016) were also identified in the L2 corpus, as discussed in the previous chapter. There were also 12 bundles which were used with

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

similar frequency in the L1 and L2 subcorpora, and showed remarkable similarities in the patterns of usage, with a few exceptions.

As mentioned in chapter five, lexical bundles do not represent a comprehension difficulty for advanced L2 learners who tend to show a good receptive knowledge of these sequences, at least in terms of the words constituting them. However, several lexical bundles seem to be understood too literally by the learners, and they may tend to use them as separate words rather than as whole unanalysed chunks (Wray, 2002), which may lead to possible misuse of these bundles, for instance, as in the case of *at the heart of*. This finding supports Durrant's (2009) observation that superficial knowledge is even more problematic than complete ignorance, as it prevents the learner from identifying the gap in her knowledge and fill it adequately (Durrant, 2009).

It was argued in the previous chapter that in this respect concordance lines can be particularly helpful, as recommended for classroom use and self-study by a number of researchers (Cargill & O'Connor, 2009; Sinclair, 1991; Thurston & Candlin, 1997) since they present significant advantages, such as multiple encounters with the linguistic phenomenon to be studied.

It was hoped that a corpus-based methodology would yield better results than had been obtained by previous studies. Unfortunately, the corpus of texts produced by L2 participants turned out to be too small to draw unambiguous conclusions. In addition, the results obtained from the TFG'17 corpus did not show a clear tendency in one

direction or another. Four contradictory types of change were detected: increase in the number of certain previously overused bundles (undesirable), decrease in the number of several previously underused bundles (undesirable), increase in the number of some previously underused bundles (desirable), and decrease in the number of several previously overused bundles (desirable). Although the positive change is estimated as slightly greater than the undesirable variation in the linguistic behaviour of the participants, the difference does not seem to be due to any variable that I am aware of and cannot be easily explained.

The workshop participants showed great interest and responsiveness to the methodology, as was the case with other studies (Erman et al., 2013; Jones & Haywood, 2004), but were unable to submit their work before the completion of this study. Following their progress over the next months could shed more light on the viability of the proposed method.

One obvious limitation of the study is the very limited time available for instruction. Previous studies have conducted trainings ranging from 10 hours to 10 weeks, and it has been suggested that much more time was needed to achieve substantial results (Cortés, 2006; Jones & Haywood, 2004). The assumption that a two-hour workshop could be enough to instruct the participants was based on the difference in teaching methodology and the student-centred character of the workshop. Although the feedback was positive, and several participants reported using the lexical bundle

approach for their own writing, it has not been possible to collect texts from these participants. This should be an object of further study.

It will be recalled that the mini-corpora compiled by the participants varied greatly in terms of focus and quality. This finding shows the importance of the learners' receptive knowledge and their proficiency as readers for self-directed study. More proficient readers are able to choose better collections of texts and in this way obtain better results.

Extensive reading may also result in a more proficient and expert-like writing, thus becoming one of the factors influencing learners' overall proficiency. In general, changes in linguistic behaviour may be due to a multitude of factors, and no methodology can be claimed to be entirely responsible for them.

This finding suggests that any study concerned with the quantification of teaching outcomes should ideally be longitudinal and embrace a period of several years (ideally dozens of years), depicting the development of the learner's competences at different moments along the timeline of her studies.

The case study of Kevin's linguistic evolution partly illustrates this idea, although it is not complete and particularly detailed. The fact that he had begun to use lexical bundles in his writing after the workshop may be due to a number of reasons unrelated to the instruction, or even to chance. However, it is likely that he has done so consciously, avoiding the use of creative constructions in favour of accepted and

widely used discourse practices, as a result, at least partly, of his participation in the workshop.

This section discussed the findings of the previous chapters and attempted to answer the research questions specified in chapter one. The next section will provide concluding remarks regarding the study and its findings.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

6.3. CONCLUSIONS

This study analysed lexical bundles in psychology, literary studies, and linguistics research articles to examine their frequency, structure, and use in the three disciplines in order to determine the viability of a discipline-specific approach to the teaching of lexical bundles and attempted to devise a corpus-based methodology for EAP classrooms and self-directed learners.

The main finding of the study is that lexical bundles vary not only across broad discipline fields, as had been shown by previous research (Biber, 2006; Hyland, 2008a), but also across disciplines belonging to the same or contiguous fields. The level of formulaicity characteristic of each discipline, the almost institutionalised set of formulaic features and their use are an important part of the generic and disciplinary conventions for each discourse community, and the knowledge of these conventions is crucial for novice researchers aspiring to be accepted in a particular discourse community.

The distinction between ‘phrasal’ and ‘clausal’ bundles made it possible to discover important features of the nature of academic discourse in different disciplines, such as the overwhelmingly nominal character of literary studies as opposed to the more

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

interactive, participant-oriented nature of psychology research articles. It is difficult to say whether literary studies and linguistics writers abstain from using participant-oriented bundles in order to distance themselves from the reader and make the writing more objective, or whether these disciplines use other resources as mechanisms of interaction with their readers. In any case, the reasons for such differences between the disciplines are unclear, although they also may be attributed to ‘disciplinary primings’.

Other differences can be explained by the communicative purposes of the genre in each discipline. Thus, the existence of a large number of bundles in literary studies which structurally do not fit in any of the previously devised categories suggests that this discipline needs a separate set of structural categories which will better reflect its phraseological patterns, an issue worth further research. One such new category was explored in this study, namely, the adjective / noun + noun phrase category, represented entirely by topic-specific bundles.

The inclusion of topic-specific bundles may be debatable, but it seems to be pedagogically justified, as in an adequately composed corpus these topic-specific bundles can become a powerful resource for novice researchers, as they can point to sets of multi-word terms typically used together in a given discipline, subdiscipline and topic and help novice researcher organise not only their writing, but also define the scope of their study.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Looking at lexical bundles of different length allowed a closer look at disciplines which show low levels of formulaicity, such as linguistics. On the other hand, the large number of bundles extracted from the psychology subcorpus made them very difficult to handle manually and may have obstructed in-depth analysis. It may be that lexical bundles should be analysed not only separately for each discipline, but also the study itself should be designed depending on the discipline. Standardised universal approaches may be useful for other purposes, such as revealing general patterns and contrasting variables. For pedagogical purposes, however, one-size-fits-all methodologies seem to render less satisfactory results.

The personalised approach to teaching lexical bundles was received enthusiastically on the part of the learners and evaluated extremely positively. Most of the participants were able to extract a list of lexical bundles which was judged as relevant to their topic and described as a powerful resource for future writing. The reasons for not submitting post-workshop texts for analysis are unclear, but a possible explanation may be that the participants did not manage to finish their texts before this study was completed. Another tentative explanation is that the method did not prove to be efficient in the process of writing. Those who write texts on a regular basis know that there is a difference between knowing what one wants to write and actually writing it. Personal circumstances should not be disregarded either.

The texts comprising the TFG'17 subcorpus are few and diverse. It is impossible to draw generalised conclusions, although the overall tendency points to a positive

change in the writing strategies of the workshop participants with respect to the TFG'15-16 subcorpus texts. The causes of the changes in the linguistic behaviour of the workshop participants need to be addressed in a future study, although some tentative explanations can be offered at this point.

The most marked tendency seems to be that of decrease in the number of overused bundles, an issue which raised great interest among the workshop participants. The concept of “linguistic teddy bears” (Hasslegren, 1994) was taken up enthusiastically. The list of overused bundles presented during the introductory session confirmed that the group perceived these bundles as ‘more academically acceptable’ and hence was likely to overuse the bundles included on the list. For this reason it was decided to include this information in the workshop and explicitly recommend that the participants should check their work for possible overused bundles. The fact that more than half of the overused bundles were used less frequently in the TFG'17 subcorpus is encouraging. The tendency suggests that at least some of the participants might have followed this recommendation and run the search utility in their word processing software to remove or replace possible redundancies. However, the list of target overused bundles provided to them was very short (see Appendix 4) and could not contain all the possible overused types, therefore, it was hoped that the workshop would provide a more sustainable solution.

It is difficult to determine the role of the workshop in the output of the participants, but the drastic reduction in the number of several overused bundles and their presence

in the target bundles list as overused bundles suggest that the introductory session had been pedagogically more valuable than the workshop itself.

It is possible that the undergraduate students followed the explicit recommendation to avoid overusing those particular bundles and instead of using the corpus analysis tool to incorporate new bundles in their writing. This seems all the more probable given the time constraints related to submitting their TFGs. Another possible explanation is that undergraduate students have a lesser degree of autonomy and are more accustomed to following directions than performing analysis. Although both the undergraduate students and the PhD students were very responsive, their behaviour during the workshops was slightly different: the undergraduates showed less autonomy and less interaction. The case with PhD students might have been different, but unfortunately, only one PhD student submitted a post-workshop text.

His contribution, although valuable, is too short to draw definitive conclusions, but the story of his linguistic evolution and his conscious efforts to improve is even more valuable.

Kevin's experience parallels that of most L2 learners, my own being a case in point. As an undergraduate student in my home country, Armenia, I was very successful and consequently, extremely self-confident. I used stylistically marked, colourful words and elaborated syntactic constructions and was admired for that by my instructors, old-school post-Soviet linguists whose model texts were the classics of

British and American fiction. I was praised for using the word *gregarious* on a daily basis.

My self-confidence began to fade as I began to realise how disadvantaged I was during encounters with speakers of English who were able to take two simple words and put them together with disarming elegance. I was aware of my deficiency, but could not describe it adequately. Was it lack of vocabulary? I did my best to replenish my word stock with all sorts of fancy words, but it did not seem to be a successful strategy. Then I began to think that my problem was rooted in my ignorance of colloquial expressions, which I searched and collected avidly. This strategy also yielded poor results. I worked on different aspects of my knowledge of English, but the desired native-like fluency was not achieved.

After dozens of years of learning English, I finally found the solution in the seminal work of Pawley and Syder (1983): what I had thought to be an art was actually a statistical measure. Native language sounds natural because it is based on frequency. This was a revelation.

I had been instructed to avoid clichés and be creative, and this pursuit of originality led me farther and farther away from naturally-sounding speech. Currently my next step is to unlearn what I have learned and to take a closer look at ‘trite’ expressions.

This study is part of my conscious attempt to discard the non-productive patterns and adopt a productive, corpus-based approach to speech production.

REFERENCES

Ackermann, K., & Chen, Y. (2013). Developing the Academic Collocation List (ACL) e a corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12, 235-247.

Ädel, A. (2014). Selecting quantitative data for qualitative analysis: A case study connecting a lexicogrammatical pattern to rhetorical moves. *Journal of English for Academic Purposes*, 16, 68–80. <https://doi.org/10.1016/j.jeap.2014.09.001>

Ädel, A., & Erman, B. (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), 81–92. <https://doi.org/10.1016/j.esp.2011.08.004>

Adolphs, S., & Durow, V. (2004). Sociocultural integration and the development of formulaic sentences. In N. Schmitt (Ed.), *Formulaic sequences. Acquisition, processing, and use*. Amsterdam: John Benjamins. 107-126.

Alali, F. & N. Schmitt (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal* 3.2, 153–180.

Allen, D. (2011). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS Learner Corpus. *Komaba Journal of English Education* 1: 105–127.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. Cowie (Ed.), *Phraseology* (pp. 101–122). Oxford: Oxford Clarendon Press.

Amosova, N.N. 1963. *Osnovi angliyskoy frazeologii*. Leningrad. (Амосова, Н.Н. Основы английской фразеологии. Ленинград).

Anthony, L. (2010). *AntConc* (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

Anthony, L. (2015). *AntFileConverter* (Version 1.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>

Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. London. Longman.

Bhatia, V. K. (2008). Genre Analysis, ESP and professional practice. *English for Specific Purposes*, 27. 161–174.

Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3–43.

Biber, D. (2005). Paquetes léxicos en textos de estudio universitario: Variación entre disciplinas académicas. *Revista Signos*, 38(57), 19–29.

Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written*

Registers [Studies in Corpus Linguistics 23]. Amsterdam: John 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. <https://doi.org/10.1075/ijcl.14.3.08bib>

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

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech*. 71–92. Frankfurt/Main: Peter Lang.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>

Biber, D., Conrad, S. M., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., Urzua, A. (2004). Representing Language Use in the University: Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9 (1), 2-20. <https://doi.org/10.1016/j.jeap.2010.01.001>

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. London: Longman.

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In Schmitt, N. (ed.). 2004. *Formulaic Sequences. Acquisition, processing and use*. Amsterdam: John Benjamins. 227–259.

Boyadzhieva, E. (2016). Learner-centered Teaching and Learner Autonomy, *Procedia - Social and Behavioral Sciences*, 232. 35-40, <http://dx.doi.org/10.1016/j.sbspro.2016.10.008>

British National Corpus. (n. d.) Available at: <http://www.natcorp.ox.ac.uk> [Accessed 29 May. 2017].

Brown, G., & Yule, G. (1983). Discourse analysis. Cambridge University Press.

Burgess, S. (2002). Packed houses and intimate gatherings: Audience and rhetorical structure in Academic Discourse. In Flowerdew, J. (ed.). Academic discourse. London. Longman. 196-215.

Burgess, S. (2017). Accept or contest: A life-history study of humanities scholars' responses to research evaluation policies in Spain. In Cargill, M. & Burgess, S. (Eds.) Publishing research in English as an Additional Language. Practices, pathways and potentials. The University of Adelaide Press. Available at www.adelaide.edu.au/press 13-31.

Byrd, P. & Coxhead, A. (2010). *On the other hand*: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5, 31-64.

Canagarajah, S. (2007). 'Lingua Franca English, multilingual communities, and language acquisition'. *The Modern Language Journal*. 91/5: 923-39.

Cargill M., & Burgess, S. (2017). Publishing research in English as an Additional Language. Practices, pathways and potentials. The University of Adelaide Press. Available at www.adelaide.edu.au/press

Cargill, M., & O'Connor, P. (2009). Writing scientific research articles: Strategy and steps. Oxford: Wiley-Blackwell.

Chen, Y-H., Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14, #2, 30-49.

Connor, U. (2004). Intercultural Rhetoric Research: Beyond Texts. *Journal of English for Academic Purposes*. 3, 291-304.

Connor, U. & Moreno, A. (2005). Tertium Comparationis: A vital component in contrastive research methodology. In P. Bruthiaux, D. Atkinson, W. G. Eggington, W. Grabe, & V. Ramanathan (eds), *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*. Clevedon, pp. 153-164. England: Multilingual Matters.

Connor, U., & Rozycki, W. (2013). Culture and language for specific purposes. In C. Chapelle (Ed), *The Encyclopedia of Applied Linguistics* (pp.1-4). New York: John Wiley and Sons.

Conrad, S., & Biber, D. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56–71. <https://doi.org/10.1515/9783484604674.56>

Cortes, V. (2002). Lexical bundles in freshman composition. In Reppen, R., Fitzmaurice, S. M., Biber, D. (eds.) *Using Corpora to Explore Linguistic Variation*. (131-145). John Benjamins Publishing Company. Amsterdam/Philadelphia.

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. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3(1), 43–57. <https://doi.org/10.3366/e1749503208000063>

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, 33-43.

Cosme, C. & Guilquin, G. (2008). Free and bound prepositions in a contrastive perspective: The case of *with* and *avec*. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. 259-274. Amsterdam: John Benjamins.

Cowie, A. (1998). Phraseological Dictionaries: Some East-West Comparisons. In A P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications*, Oxford Clarendon Press, 208-228.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34: 213–238.

Coxhead, A. (2008). Phraseology and English for Academic Purposes. Challenges and opportunities. In Granger, S., Meunier, F. (eds.): *Phraseology in foreign language learning and teaching*. John Benjamins. Amsterdam / Philadelphia. 149 – 161.

Coxhead, A. (2015) ‘Replication research in pedagogical approaches to formulaic sequences: Jones & Haywood (2004) and Alali & Schmitt (2012)’. *Language Teaching*, 1–11. doi: 10.1017/S0261444815000221

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129–147.

Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics in Language Teaching*, 49, 1–26.

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer*. 67–79. London: Longman.

Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins. 87–106.

Dudley-Evans, T. (2000). Genre analysis: a key to a theory of ESP? *Ibérica*, 2, pp. 3–11.

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes*, 28, 157–169. <http://dx.doi.org/10.1016/j.esp.2009.02.002>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics*, 47, 157–177.

Ellis, N. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*. 24/2: 297-339.

Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics and TESOL. *TESOL Quarterly* 42(3): 375– 396.

Elsevier Scopus. Elsevier. Available at: <https://www.elsevier.com/> [Accessed 01 Jun 2017]

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

Eriksson, A. (2012) Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In Thomas, J. & Boulton, A. (Eds.). *Input, Process and Product: Developments in Teaching and Language Corpora*, James Thomas & Alex Boulton (eds), 195–211. Brno: Masaryk University Press.

Erman, B., Lewis, M., & Fant, L. (2013). Multiword structures in different materials, and with different goals and methodologies. In Romero-Trujillo, J. (ed.). *Yearbook of corpus linguistics and pragmatics 2013*. 77 – 103. 1st ed. Dordrecht: Springer.

Farvardin, M., Afghari, A., Koosha, M. (2012). Analysis of four-word lexical bundles in physics research articles. *Advances in Digital Multimedia (ADMM)*, 1(3), 134–139. Retrieved from <http://worldsciencepublisher.com/journals/index.php/ADMM/article/view/436>

Fiedler, S. (2017). Phraseological borrowing from English into German: Cultural and pragmatic implications. *Journal of Pragmatics*, 113, pp.89-102.

Firth, J. (1962). *Papers in linguistics 1934-1951*. Blackwell. Oxford.

Flowerdew, J. (2008). Scholarly writers who use English as an Additional Language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes*, 7 (2), 77-86.

Flowerdew, J. (2010). Use of signalling nouns across L1 and L2 writer corpora. *International Journal of Corpus Linguistics*, 15(1), 36–55. <https://doi.org/10.1075/ijcl.15.1.02flo>

Flowerdew, L. (2002). Corpus-Based Analyses in EAP. In J. Flowerdew (Ed.). *Academic Discourse*. London. Longman. 95–114.

Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes*, 32(4), 208–220. <https://doi.org/10.1016/j.esp.2013.06.001>

Fontenelle, T. (1998). Discovering Significant Lexical Functions in Dictionary Entries. In A. P. Cowie (Ed.). *Phraseology*. Oxford: Oxford Clarendon Press.

Gilmore, A. (2015). Research into practice: The influence of discourse studies on language descriptions and task design in published ELT materials. *Language Teaching*, 48(4), 506–530. <https://doi.org/10.1017/S0261444815000269>

Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications*, Oxford Clarendon Press, 125-143.

Graddol, D. (2006). *English next*. London. British Council retrieved from http://vigdis.hi.is/sites/vigdis.hi.is/files/images/einangrun_enskumaelandi_folks.pdf

Granger, S. (1998a). The computer learner corpus: a versatile new source of data for SLA research. In Granger, S. (1998). *Learner English on Computer*. 3-18. London / New York. Longman.

Granger, S. (1998b). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications*, Oxford Clarendon Press, 145-160

Granger, S., & Lefer, M.-A. (2016). From General to Learners' Bilingual Dictionaries: Towards a More Effective Fulfilment of Advanced Learners' Phraseological Needs. *International Journal of Lexicography*, 29(3), 279–295. <https://doi.org/10.1093/ijl/ecw022>

Granger, S., Meunier, F. (Eds.). (2008a). *Phraseology in foreign language learning and teaching*. John Benjamins. Amsterdam / Philadelphia.

Granger, S. & Meunier F. (Eds.). (2008b). *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. 28-49. Amsterdam: John Benjamins.

Gries, S. (2008). Phraseology and linguistic theory: a brief survey. In S. Granger, & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. 3-25. Amsterdam & Philadelphia: John Benjamins.

Handl, S. (2008). Essential collocations for learners of English. The role of collocational direction and weight. In Granger, S., Meunier, F. (eds.): *Phraseology in foreign language learning and teaching*. John Benjamins. Amsterdam / Philadelphia. 43 – 66.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4, 237–260.

Hatami, S. (2015). Teaching Formulaic Sequences in the ESL Classroom. *TESOL Journal*, 6 (1), 112–129. <https://doi.org/10.1002/tesj.143>

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

Howarth, P. (1996). Phraseology in English academic writing. Some implications for language learning and dictionary making. Tübingen: Max Niemeyer Verlag.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *In Phraseology: Theory, Analysis and Applications*. Oxford: Clarendon Press. 161–186.

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System*, 53, 13–23. <https://doi.org/10.1016/j.system.2015.06.011>

Hunston, S. (2002). Pattern grammar, language teaching, and linguistic variation: Applications of a corpus-driven grammar. In Reppen, R., Fitzmaurice, S. M., Biber, D. (eds.) *Using Corpora to Explore Linguistic Variation*. (167-186). John Benjamins Publishing Company. Amsterdam/Philadelphia.

Hyland, K. (2008a). *As can be seen*: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27, 4-21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62. <https://doi.org/10.1111/j.1473-4192.2008.00178.x>

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32(2012), 150–169. <https://doi.org/10.1017/S0267190512000037>

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*. 31, 58-69.

Hyland, K. & Hamp-Lyons, L. (2002). EAP: issues and directions. *Journal of English for Academic Purposes*. 1 (1) 1-12.

Johnston, K. (2017). Lexical bundles in applied linguistics and literature writing: a comparison of intermediate English learners and professionals. Unpublished MA

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

dissertation. Portland State University. Paper 3482.

<https://doi.org/10.15760/etd.3358>

Jiang, N., & Nekrasova, T. M. (2007). The Processing of Formulaic Sequences by Second Language Speakers. *The Modern Language Journal*, 91(3), 433–445.

<https://doi.org/10.1111/j.1540-4781.2007.00589.x>

Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences. Acquisition, processing and use*. Amsterdam: John Benjamins. 269-291.

Kazemi, M., Kohandani, M., & Farzaneh, N. (2014). The impact of lexical bundles on how applied linguistics articles are evaluated. *Procedia - Social and Behavioral Sciences*, 98, 870–875.

<https://doi.org/10.1016/j.sbspro.2014.03.494>

Kennedy, G. (2008). Phraseology and language pedagogy: Semantic preference associated with English verbs in the British National Corpus. In Granger, S., Meunier, F. (eds.): *Phraseology in foreign language learning and teaching*. John Benjamins. Amsterdam / Philadelphia. 21–42.

Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In Schmitt, N. (ed.). 2004. *Formulaic Sequences*. Language Learning & Language Teaching. Amsterdam: John Benjamins. 37-54.

Kuteeva, Maria ; Mauranen, Anna. / Writing for publication in multilingual contexts: An introduction to the special issue. *Journal of English for Academic Purposes*. Vol. 13. 1-4.

Kuteeva, M., & Negretti, R. (2016). Graduate students’ genre knowledge and perceived disciplinary practices: Creating a research space across disciplines. *English for Specific Purposes*. 36-49. <https://doi.org/10.1016/j.esp.2015.08.004>

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, 39, 45–61. <https://doi.org/10.1016/j.esp.2015.03.003>

Leech, G. (1998). Preface. In Granger, S. (ed.) *Learner English on Computer*. XIV – XX. London / New York. Longman.

Lillis, T., & Curry, M.J. (2015). The politics of English, language and uptake: The case of international academic journal article reviews. *AILA Review*, 28(1), 127-150. DOI:10.1075/aila.28.06lil

Lindstromberg, S., Eyckmans, J., & Connabeer, R. (2016). A modified dictogloss for helping learners remember L2 academic English formulaic sequences for use in later writing. *English for Specific Purposes*, 41, 12–21. <https://doi.org/10.1016/j.esp.2015.08.002>

Louw, B. 2000. Contextual prosodic theory: Bringing semantic prosodies to life. In *Words in Context: A Tribute to John Sinclair on His Retirement*, Chris Heffer & Helen Sauntson (Eds.), 48–94. Birmingham: University of Birmingham.

Martín Martín, P. A. (2003). A genre-analysis of research paper abstracts in experimental social sciences. *English for Specific Purposes* 22 (1), 25-43.

Martin, W. (2008). A unified approach to semantic frames and collocational patterns. In: S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective*, Amsterdam/ Philadelphia: John Benjamins, 51-65.

Matukhin, D. & Bolgova, D. (2015). Learner-centered Approach in Teaching Foreign Language: Psychological and Pedagogical Conditions, *Procedia - Social and Behavioral Sciences*, 206, 148-155, <http://dx.doi.org/10.1016/j.sbspro.2015.10.044>

Mel'čuk, I. (1998) Collocations and Lexical Functions. In A.P. Cowie (ed.): *Phraseology. Theory, Analysis, and Applications*, Oxford: Clarendon Press, 23-53.

Meunier, F. (1998). Computer tools for the analysis of learner corpora. In Granger, S. (1998). *Learner English on Computer*. 19-37. London / New York. Longman.

Miller, G. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), pp.81-97.

Milton, J. (1998). Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment. In S. Granger (Ed.), *Learner English on computer* (186–198). London: Longman.

Moreno, A. I., Rey-Rocha, J., Burgess, S., López Navarro, I., & Sachdev, I. (2012). Spanish researchers' perceived difficulty writing research articles for English-medium journals: The impact of proficiency in English versus publication experience. *Ibérica*, 24, 157-184. Retrieved from <http://www.aelfe.org/?s=presentacio>

Morley, J. (n. d.). Manchester Academic Phrasebank. Available at: <http://www.phrasebank.manchester.ac.uk> [Accessed 20 May. 2017].

Moon, R. (1998a). Frequencies and forms of phrasal lexemes in English. In Cowie, A. P. (ed.). *Phraseology. Theory, Analysis, and Applications*. 79–100. Oxford: Oxford University Press.

Moon, R. (1998b). Fixed expressions and idioms in English: A corpus-based approach. Oxford Clarendon Press.

Moon, R. (2008). Dictionaries and collocation. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. 313-309. Amsterdam: John Benjamins.

Monti, J., Barreiro, A., Elia, A., Marano, F., & Napoli, A. (2011). Taking on new challenges in multi-word unit processing for machine translation. Proceedings of the Second International Workshop on Free/Open-Source Rule-Based Machine

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Translation (2011: Barcelona). < [Http://hdl. Handle. net/10609/5646](http://hdl.handle.net/10609/5646), (January), 11–19. Retrieved from <http://openaccess.uoc.edu/webapps/o2/handle/10609/5646>

Mudraya, O., Piao, S. S. L., Rayson, P., Sharoff, S., Babych, B., Löftberg, L. (2008). Automatic extraction of translation equivalents of phrasal and light verbs in English and Russian. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. 293-309. Amsterdam: John Benjamins.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Neely, E., & Cortes, V. (2009). *A little bit about: Analyzing and teaching lexical bundles in academic lectures*. *Language Value*, 1(1), 17-38.

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

Oakey, D. (2002). Formulaic Language in English Academic Writing. A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines In R. Reppen, S. M. Fitzmaurice, D. Biber (eds.). *Using Corpora To Explore Linguistic Variation*. (111-130). John Benjamins Publishing Company. Amsterdam/Philadelphia.

Paltridge, B. (1997). *Genre, Frames and Writing in Research Settings*. John Benjamins Publishing Company. Amsterdam/Philadelphia.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60–71. <https://doi.org/10.1016/j.jeap.2015.11.003>

Paquot, M. (2013). Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*, 18(3), 391 – 417.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14, 48–59. <https://doi.org/10.1016/j.jeap.2013.12.001>

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory native like selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.). *Language and communication* 191–230. London: Longman.

Peacock, M. (2012). High-frequency collocations of nouns in research articles across eight disciplines. *Ibérica: Revista de La Asociación Europea de Lenguas Para Fines Específicos (AELFE)*, 23 (2012), 29–46. <https://doi.org/doi:10.1075/ijcl.16.1.04pea>

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84–94.

Peters, E., & Pauwels, P. (2015). Learning academic formulaic sequences. *Journal of English for Academic Purposes*. <https://doi.org/10.1016/j.jeap.2015.04.002>

Phrasebank.manchester.ac.uk. (2017). Academic Phrasebank. [online] Available at: <http://www.phrasebank.manchester.ac.uk/> [Accessed 25 May. 2017].

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220-231.

Read, J. & Nation, P. (2004). Measurement of formulaic sequences. In Schmitt, N (Ed.), *Formulaic Sequences Acquisition, processing and use*. Amsterdam: John Benjamins. 23-35.

Reppen, R., Fitzmaurice, S. M., Biber, D. (eds.). *Using Corpora to Explore Linguistic Variation*. John Benjamins Publishing Company. Amsterdam/Philadelphia.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Rica, J. P. (2010). Corpus analysis and phraseology: transfer of multi-word units. *Linguistics and the Human Sciences*, 6, 321-343.

Römer, U. (2009). The inseparability of lexis and grammar. *Annual Review of Cognitive Linguistics*, 7(1), 140-162.

Sabban, A. (2008). Critical observation on the culture-boundness of phraseology. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. 229-241. Amsterdam: John Benjamins.

Salazar, D. (2011). Lexical bundles in scientific English: A corpus-based study of native and non-native writing. Unpublished doctoral dissertation. Retrieved from http://diposit.ub.edu/dspace/handle/2445/35023%5Cnhttp://diposit.ub.edu/dspace/bitstream/2445/35023/2/DJLS_DISSERTATION.pdf

Salazar, D. (2014). Lexical bundles in native and non-native scientific writing. Applying a corpus-based study to language teaching. John Benjamins Publishing Company. Amsterdam/Philadelphia.

Sánchez, P. 2013. “Lexical bundles in three oral corpora of university students.” *Nordic Journal of English Studies* 13(1): 187–209.

Sánchez, P. (2014). Lexical bundles in Biology: Differences between textbooks and research articles. *Revista Española de Lingüística Aplicada* 27:2, 492–513. doi 10.1075/resla.27.2.12san

Schmitt, N. (ed.). 2004. *Formulaic Sequences. Acquisition, processing and use*. Amsterdam: John Benjamins.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins. 127–152.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins. 1-22.

Schmitt, N., Dörnyei, Z., Adolphs, S., Durow, V. (2004). Knowledge and acquisition of formulaic sequences: a longitudinal study. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins. 55–86.

Schmitt, N. & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins. 173–189.

Shahriari, H. (2017). Comparing lexical bundles across the introduction, method and results sections of the research article. *Corpora*, 12(1), 1-22. <http://dx.doi.org/10.3366/cor.2017.0107>

Shrestha, P. N. (2017). Investigating the learning transfer of genre features and conceptual knowledge from an academic literacy course to business studies: Exploring the potential of dynamic assessment. *Journal of English for Academic Purposes*, 25, 1–17. <https://doi.org/10.1016/j.jeap.2016.10.002>

Siepmann, D. (2008). Phraseology in learners’ dictionaries. What, where and how? In Granger, S., Meunier, F. (eds.): *Phraseology in foreign language learning and teaching*. John Benjamins. Amsterdam / Philadelphia. 185 – 202.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512. <https://doi.org/10.1093/applin/amp058>

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

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

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015. Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Sinclair, J. (2008). The phrase, the whole phrase, and nothing but the phrase. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. 407-410. Amsterdam: John Benjamins.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429-458.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*. 214-225. <https://doi.org/10.1016/j.jeap.2013.05.002>

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

Swales, J. M., (1990). *Genre Analysis. English in academic and research settings*. New York. Cambridge University Press.

Swales, J. M. (1997). English as Tyrannosaurus rex. *World Englishes* 16(3): 373-382.

Swales, J. M. (2004). *Research Genres. Explorations and applications*. Cambridge University Press.

Thompson, P. (2013). Learner-centred education and ‘cultural translation’. *International Journal of Educational Development*, 33 (1), 48-58. <http://dx.doi.org/10.1016/j.ijedudev.2012.02.009>

Thurston, J., & Candlin, C. (1997). *Exploring academic English: A workbook for student essay writing*. Sydney: NCELTR Publications.

Tognini-Bonelli, E. (1996). Towards translation equivalence from a corpus linguistics perspective. *International Journal of Lexicography*, 9 (3). 197-217.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work (Studies in corpus linguistics, v. 6)*. John Benjamins Publishing Company.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence from Self-Paced Reading and Sentence Recall Tasks. *Language Learning*, 61(2), 569–613. <https://doi.org/10.1111/j.1467-9922.2010.00622.x>

Underwood, G., Schmitt, N. & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use*. 153–173.

Vaughan, E. & Clancy, B. (2013). Small corpora and Pragmatics. In Romero-Trillo (Ed.) *Yearbook of Corpus Linguistics and Pragmatics 2013: New Domains and Methodologies*, Springer Science+Business Media Dordrecht. 53-73. DOI 10.1007/978-94-007-6250-3_4

Vincent, B. (2013). Investigating academic phraseology through combinations of very frequent words: A methodological exploration. *Journal of English for Academic Purposes*, 12(1), 44–56. <https://doi.org/10.1016/j.jeap.2012.11.007>

Vinogradov, V. V. (1947), 'Ob osnovnuikh tipakh frazeologicheskikh edinit v russkom yazyke', in A. A. Shakhmatov, 1864-1920. *Sbornik statey i materialov* (Moscow: Nauka), 339-64. (Виноградов, В.В. Об основных типах фразеологических единиц в русском языке. Сборник статей и материалов 1864-1920гг. под редакцией А.А. Шахматова. 1947. Москва. Наука).

Vygotsky, L.S. (1987). *The Collected Works of L.S. Vygotsky: vol. 1, Problems of General Psychology. Including the Volume Thinking and Speech.* (N. Minick, Trans.) Plenum, New York.

Wible, D. (2008). Multiword expressions and the digital turn. In Granger, S., Meunier, F. (eds.): *Phraseology in foreign language learning and teaching*. John Benjamins. Amsterdam / Philadelphia. 163 – 181.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2004). ‘Here’s one I prepared earlier’ Formulaic language learning on television. In Schmitt, N. (ed.). *Formulaic Sequences. Acquisition, processing and use*. Amsterdam: John Benjamins. 249–268.

Wray, A. (2013). Research timeline: Formulaic language. *Language Teaching* 46.3, 316–334.

Wray, A., & Perkins, M. R. (2000). The Functions of Formulaic Language: An Integrated Model. In *Language & Communication*, 20(1), 1–28. [https://doi.org/10.1016/S0271-5309\(99\)00015-4](https://doi.org/10.1016/S0271-5309(99)00015-4)

Wulff, S. (2010). *Rethinking idiomaticity*. London: Continuum.

Yorio, C. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & K. Obler (Eds.), *Bilingualism across the lifespan*. Cambridge: Cambridge University Press. 55–72.

APPENDICES

339

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

APPENDIX 1**THE PHD AND TFG'15-16 SUBCORPORA**

1. The PhD subcorpus

	Author	Title	Year	University
LINGUISTICS	Heather Alice Sturman	Immigrant Socialization: Literacy Development among Girls in a U.S. High School	2012	University of California
	David Mark Smith	Linguistic Change in the Galician Speech Community	2000	University of Liverpool
	Ryan M. Nefdt	The Foundations of Linguistics: Mathematics, Models, and Structures	2016	University of St. Andrews
	Paul Robert McPherron	Internationalizing Teaching, Localizing English: Language Teaching Reforms through a South Chinese University	2008	University of California
	Holly A. Lakey	The grammar of fear: morphosyntactic metaphor in fear constructions	2016	University of Oregon
LITERARY STUDIES	Robert Arthur Gillespie	Shades of an urban frontier: historical resonances in the cities of Black and Anglophone SF	2015	University of Iowa ³
	Sarah Fay	The American tradition of the literary interview, 1840 – 1956: A Cultural history	2013	University of Iowa
	Judith Claire Coleman	Holy vessels, tyrants, fools, and blind men : performing antinomianism and transgressive agency in English drama	2013	University of Iowa
	Andrew Crooke	In praise of peasants : ways of seeing the rural poor in the work of James Agee, Walker Evans, John Berger, and Jean Mohr	2013	University of Iowa
EDUCATION	Patricia Ann Tunstall	New educational assessment and the construction of reality	1999	University of London
	David George Holloway	Practitioner research and professional development: their contribution to an understanding of curriculum and	2001	University of Portsmouth

³ Different PhD dissertations in Literary Studies were downloaded from online libraries of other universities, but were discarded due to a number of reasons, such as their structure which seemed to deviate dramatically from the IMRAD pattern and lacked any explicitly signalled organisation; inadequate length (either too long or too short to be comparable to other texts in the subcorpus), impossibility of converting them into .txt format due to document protection restrictions, their availability online and copyright restrictions, etc. The Literary Studies theses provided by the University of Iowa seemed the most appropriate for the purposes of this study and were selected in spite of the drawback of providing little variation.

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

		organisational change in the post-compulsory education sector		
	Charles W. Beale	From jazz to jazz education: an investigation of tensions between player and educator definitions of jazz	2001	University of London
	Christopher Denis Wiltsher	The concept of education in higher education in England 1960-1997 with special reference to adult continuing education.	2004	N/A
	Emily Pringle	The artist as educator: an examination of the relationship between artistic practice and pedagogy within contemporary gallery education	2004	Loughborough University

2. The TFG'15-16 subcorpus

Author	Title	Year
Juan Carlos Rodríguez González	'Advance': Meaning, Syntax and the Influence of Metaphors in a Verb of Movement	2015
Cristina Barreto Tomé	A functional lexematic analysis of separate verbs: paradigmatic and syntagmatic features	2015
Beatriz Sánchez Ramos	The Heidi Chronicles as Illustration of the Second Feminist Wave in the United States	2015
Laura Álvarez Machado	The Guts as Illustration of the Importance of Humour in Life	2016
Ruth Gómez Love	The Figure of the Fallen Woman in Emmeline, the Orphan of the Castle and Memoirs of Modern Philosophers: The Journey from Innocence to Damnation	2015
Francisco Yeray Estévez Cabrera	The Great Gatsby: A Queer Approach	2016
Kevin García González	The 1920s in the United States: An Account of the Filmic Representation	2015
Irene Cubas Régulo	The comic hero and his attitude towards women in a star called Henry and oh, play that thing	2015
N/A	Antilanguage in the songs of The Stranglers	N/A
Verónica Concepción Ramos	CLIL - translating theory into practice: acquiring language in relation to the process of learning language	2016
Natalia González González	The implementation analysis of Lengua Inglesa I: An analysis of its effectiveness in developing communication skills in English	2015
Beatriz González Reyes	Toni Morrison's Home: A Portrayal of the 1950s for African Americans	2015

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

Leticia Siverio González	The use of monsters in Beowulf	2015
Haidee Mora Padilla	The Syntactic and Semantic Interface of English Cut Verbs	2015
Dimas Enrique Acosta Rodríguez	The Non-Violent Message In Ernest Hemingway's Writing	2015
Lourdes González Hernández	The Mapping of the Self A Jungian Analysis of Atwood's Surfacing	2015
Jairo Adrián Hernández	The Lotus Flower that Grows out of the Mud: an Approach to Postcolonial Feminism in Bharati Mukherjee's Jasmine	2015
Félix Aday Rodríguez Alonso	The Loss of El Dorado: A Different Approach to Colonial History?	2015
Fabián García Díaz	Verbs of 'preparing something for eating by heating it in a particular way': a lexicological analysis	2015
Pedro Daniel Albericio Martín	The interaction of syntax and semantics in the lexicological analysis of combine verbs	2015

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

APPENDIX 2

LEXICAL BUNDLES EXTRACTED FROM THE PHD CORPUS

3-word bundles

Bundle	Frequency	Bundle	Frequency
as well as	90	use of the	26
in order to	81	a variety of	25
the development of	80	the idea of	25
in terms of	65	the influence of	24
the nature of	47	the notion of	24
a number of	46	of the world	23
part of the	46	the construction of	23
the concept of	45	the importance of	23
in which the	42	aspects of the	22
in relation to	40	in this study	22
one of the	39	the [X]th century	22
some of the	36	way in which	22
the fact that	34	a form of	21
the end of	33	a set of	21
as part of	32	many of the	21
such as the	32	of the author	21
the relationship between	32	on the other	21
the work of	31	the field of	21
the process of	30	the focus of	21
in this way	29	the majority of	21
that it is	29	the study of	21
there is a	29	discussion of the	20
in the [years]s	28	in other words	20
in the same	27	the form of	20
		the late [year]s	20

4-word bundles

Type	Frequency	Type	Frequency
at the same time	37	the use of the	13
the development of the	25	at the heart of	12
the end of the	23	in the case of	12
in the context of	22	the influence of the	12
as well as the	20	the majority of the	12
the way in which	20	at the center of	11
the [year]s and [year]s	19	the center of the	11
the extent to which	18	at the start of	10
the nature of the	18	in relation to the	10
in terms of the	17	in the s and	10
at the end of	16	in the united states	10
the context of the	16	of the development of	10
in the same way	14	the role of the	10
on the other hand	13	within the context of	10

5-word bundles

Type	Frequency
at the end of the	12
in the context of the	9
at the center of the	8
at the start of the	8
in the [year]s and [year]s	7
during the [year]s and [year]s	6
within the context of the	6

APPENDIX 3

LEXICAL BUNDLES EXTRACTED FROM THE EXPERT CORPUS

LEXICAL BUNDLES IN PSYCHOLOGY

3-word bundles

Bundle	Freq.	Bundle	Freq.	Bundle	Freq.
a function of	121	in the same	70	the amount of	66
a number of	137	in this article	57	the association between	52
a range of	59	in this section	61	the basis of	74
a review see	59	in which the	145	the comparison of	55
a set of	85	included in the	68	the context of	96
a variety of	84	individual differences in	83	the development of	124
according to the	101	is associated with	94	the difference between	57
account for the	66	is based on	51	the distribution of	55
activation of the	67	is consistent with	87	the effect of	119
all of the	71	is possible that	56	the effects of	215
are likely to	69	is that the	97	the extent to	68
are more likely	58	it has been	54	the fact that	123
as a function	89	it is not	85	the idea that	89
as a result	88	it is possible	82	the impact of	57
as well as	406	likely to be	84	the importance of	87
associated with the	58	may not be	51	the influence of	66
at the same	50	more likely to	141	the level of	60
based on the	118	most of the	57	the magnitude of	58
be used to	56	number of studies	51	the nature of	53
been shown to	56	of the model	91	the number of	259
between the two	56	of the stimulus	60	the other hand	94
consistent with the	134	of the target	65	the presence of	115
depends on the	58	of the task	51	the processing of	61
differences in the	60	of the two	70	the proportion of	51
each of the	81	on the basis	72	the relationship between	104
effect sizes were	59	on the other	117	the results of	72
extent to which	68	one of the	146	the role of	120
for a review	73	part of the	51	the set of	77
for example the	60	related to the	74	the size of	74
found that the	74	shown in figure	69	the strength of	78
function of the	54	size of the	69	the type of	65
have shown that	50	some of the	73	the use of	73
in addition to	85	such as the	145	the value of	68

345

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

in line with	74	suggests that the	56	there is a	101
in order to	121	support for the	52	there is no	62
in other words	70	that can be	62	to account for	80
in response to	81	that is the	51	to be a	56
in terms of	194	that it is	86	to respond to	51
in the absence	63	that there is	52	trial to trial	53
in the context	90	the ability to	64	understanding of the	59
in the literature	62	the absence of	101	ventromedial prefrontal cortex	50
in the present	55	the activation of	68	well as the	69

4-word bundles

Bundle	Freq.	Bundle	Freq.	Bundle	Freq.
a function of the	37	in a way that	16	on the basis of	69
a great deal of	17	in addition to the	17	on the basis of the	17
a recent meta analysis	17	in contrast to the	17	on the one hand	19
a risk factor for	27	in line with the	27	on the other hand	94
a wide range of	29	in terms of the	46	one of the most	23
an important role in	16	in the absence of	63	our understanding of the	19
and colleagues found that	19	in the case of	28	over the course of	21
are consistent with the	19	in the context of	83	play a role in	16
are likely to be	20	in the face of	21	risk factor for depression	17
are more likely to	50	in the form of	29	studies have shown that	16
as a function of	89	in the meta analysis	16	that there is a	19
as a function of the	22	in the next section	19	the absence of a	19
as a result of	41	in the number of	19	the activation of the	30
as long as the	16	in the presence of	22	the basis of the	20
as shown in figure	25	in the previous section	17	the degree to which	16
as the number of	17	in the sense that	17	the difference between the	22
as well as the	68	in the united states	19	the distance between the	18
as well as to	16	in this meta analysis	16	the end of the	26
at the end of	25	in this section we	30	the extent to which	68
at the level of	22	is based on the	16	the fact that the	27
at the same time	41	is consistent with the	40	the location of the	19
at the time of	20	is likely to be	17	the magnitude of the	27
can be found in	27	is more likely to	25	the nature of the	20
can be used to	33	is possible that the	18	the presence of a	19
consistent with the idea	16	is shown in figure	19	the results of the	17
et al as well as	16	is that it is	16	the role of the	22
et al for example	16	it can be seen	17	the shape of the	17
et al found that	59	it is important to	45	the size of the	58
et al showed that	20	it is likely that	16	the strength of the	33
for a review see	59	it is possible that	56	the value of the	22
for each of the	17	it is possible that the	18	there is evidence that	18
from trial to trial	16	it is possible to	21	these findings suggest that	17
has been shown to	20	it may be that	16	to account for the	37

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

have been found to	16	it should be noted	21	to the development of	22
have been shown to	25	more likely to be	22	to the extent that	33
have been used to	16	of individual differences in	22	to the fact that	17
in a study of	17	of the effects of	19	we found that the	16
in a variety of	17	of the number of	20	were more likely to	27
				with the idea that	24

5-word bundles

Bundle	Freq.	Bundle	Freq.
a risk factor for depression	12	it is important to note	14
as a function of the	22	it is possible that the	18
at the end of the	15	it should be noted that	14
consistent with the idea that	15	on the basis of the	17
is important to note that	12	on the other hand the	15
it can be seen that	15		

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

APPENDIX 4

MATERIALS OF THE INTRODUCTORY SESSION AND THE WORKSHOP

Introductory email with instructions (English)

Dear TFG writers,

Here is some introductory information on the workshop.

At the end of the session, each of you will have her own personalised list of lexical bundles, all set and ready to use.

We will need one piece of software for that, which is called AntConc, it is absolutely free and very user-friendly. You can download it here:

<http://www.laurenceanthony.net/software/antconc/>

Below is a description of what we will do during the session.

1. We will choose 5-7 model texts which we think are representative of our field and whose style we would like to follow.
2. We will paste them into txt files for the software to be able to read them.
3. We will process the texts in accordance with corpus compiling principles so that they render the best possible results.
4. Using AntConc, we will generate a list of lexical bundles based on your model texts.
5. Using AntConc again, we will have a look at concordance lines for some lexical bundles to see how they are used in the model texts.

348

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

THE ACQUISITION OF LEXICAL BUNDLES IN EAP

The whole workshop should take less than an hour, if everything goes well.

I'll ask you to note down some details regarding your corpus and to send me the list of lexical bundles that comes out. I will also ask you to send me the final versions of your TFGs for my research. I might be able to provide some helpful comments on how to improve them further. The analysis will be completely anonymous and no substantial part of your work will be reproduced without your permission.

It would be fantastic if you could download and install [AntConc](#) before the session, but if not, no problem, we'll do it in situ.

I hope this email has been helpful and now you have a better understanding of what we are going to do. I also hope it will be useful for you.

See you tomorrow!

Best regards,

Diana

349

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

Introductory email (Spanish)

Gracias por confirmar su asistencia al taller del viernes.

Este es un correo introductorio para explicar un poco qué es lo que vamos a hacer en el taller, qué herramientas necesitamos y qué resultados cabe esperar.

Paquetes léxicos

Como ya saben, vamos a hablar de unos elementos textuales que se llaman paquetes léxicos (lexical bundles en inglés). Tienen muchos otros nombres, pero eso para nosotros es irrelevante.

Lo que sí es relevante es que dichos paquetes léxicos son las frases o secuencias de palabras más recurrentes en un tipo de texto y un campo determinados. Es decir, un artículo o un resumen de artículo en el campo de psicología probablemente tendría una serie de frases que son características de este género y esta disciplina. Y no estoy hablando de palabras clave. Los paquetes léxicos son **frases incompletas, transparentes** (su significado es la suma del significado de los elementos que contiene), y, sobre todo, **muy frecuentes**, por ejemplo, *it has been suggested that, as can be seen, as well as, at the end of the.*

Parecen fáciles y no tienen nada especial, pero por eso mismo muchas veces descuidamos ponerlos en el texto, y el texto pierde en cohesión y se hace más parecido a un texto escrito por un novato.

Como ya he dicho, los paquetes léxicos se definen para cada campo por separado y para cada género por separado. Esto quiere decir que una frase puede ser paquete léxico para biología, pero no para literatura, no porque no se use del todo, sino porque no se usa con tanta frecuencia en una disciplina como en la otra.

En este taller voy a proponer que creamos nuestras propias listas personalizadas de paquetes léxicos, basándonos en unos textos modelos que elegirá cada participante. Esto nos permitirá tomar los textos que para nosotros son representativos y extraer las frases más recurrentes de ellos para estudiarlas y poder usarlas en nuestros propios textos. Les adelanto que esto NO ES PLAGIO, y si quieren en el taller hablaremos de ello.

Entonces, ¿cómo vamos a extraer esos paquetes léxicos?

Primero necesitamos que participante prepare sus textos modelo, unos 8-10 artículos o capítulos de libro que son cruciales para su tema y cuyo estilo le gustaría replicar.

Por eso les pido que descarguen los pdfs de sus textos de referencia (el Punto Q suele proveer materiales excelentes) y los tengan a mano en sus portátiles cuando vayan al taller.

Nosotros convertiremos esos pdfs en archivos de textos para poder usarlos con el programa extractor de paquetes léxicos.

Luego usaremos el programa extractor AntConc para extraer los paquetes léxicos y estudiar cómo se han usado en los textos modelo.

Por eso les pido que descarguen y tengan a mano el programa informático AntConc. Es gratuito y muy sencillo de usar, ya lo verán. Se puede descargar en este enlace:

<http://www.laurenceanthony.net/software.html>

En el mismo enlace hay otro programa que se llama AntFileConverter. Pueden descargarlo también para convertir los pdfs en archivos de texto. Hay otras formas de hacerlo, así que no insisto, pero esta parece ser la más rápida.

Si todo va bien, este viernes ya tendrán su propia lista de paquetes léxicos, listos para usar en sus trabajos y mejorar sus textos en inglés sin mucho esfuerzo.

RESUMEN

Les pido que hagan 2 cosas para este taller:

- 1. Preparen 8-10 textos de referencia digitalizados (descargarlos de internet si aún no los tienen)**
- 2. Descarguen el programa AntConc para extraer los paquetes léxicos.**

Si quieren, también pueden descargar el AntFileConverter para facilitar nuestro trabajo.

Nos vemos el viernes en la entrada de la Facultad de Filología.

Saludos,

Diana

ACTIVITIES FOR THE INTRODUCTORY SESSION

Read these sentences carefully. Which of them seem academically acceptable? Why?

We will treat differences in these descriptions as disagreements as to what this area of vocabulary is.	We will treat differences in these descriptions as disagreements about the nature of this area of vocabulary.
In a chapter which is concerned with the connection of the academic performance and written proficiency, the author tackles several questions.	In a chapter which is concerned with the relationship between the academic performance and written proficiency, the author tackles a number of questions.
In this expanding field, Thierry Fontenelle's research is especially noteworthy.	In this expanding field, the work of Thierry Fontenelle is especially noteworthy.
This is a useful point of reference for constructing a classification scheme.	This is a useful point of reference for the construction of a classification scheme.
In the past, I have done more narrowly-focused research with different nouns.	The work that I have done in the past has been with a set of nouns that is more narrowly focused.

Insert the corresponding bundle from the list below.

the nature of, the relationship between, the work of, the construction of, in this study / thesis, a set of, the focus of, the form of

1. The interviewer grants respondents the right to control _____ meaning.
2. The term 'educational assessment' refers to _____ appraisal or evaluation of children's work.
3. The activities are also each defined by _____ 'steps', a sequence of activities for teachers to develop mastery.
4. The research is centrally concerned with _____ the practice of educational assessment.
5. _____, I adopt all these approaches.

352

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

6. _____ Malcom Ross (1974), for example, argues that music education combines self- expression and self-discovery with highly disciplined and recreative technical achievement.

7. It is shortsighted to oversimplify _____ Puritans and antitheatricalists.

8. _____ this study on concepts will necessarily entail ignoring many aspects of the changes in higher education.

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

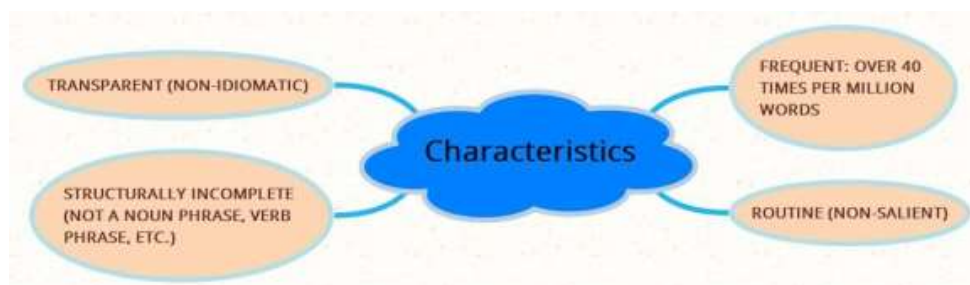
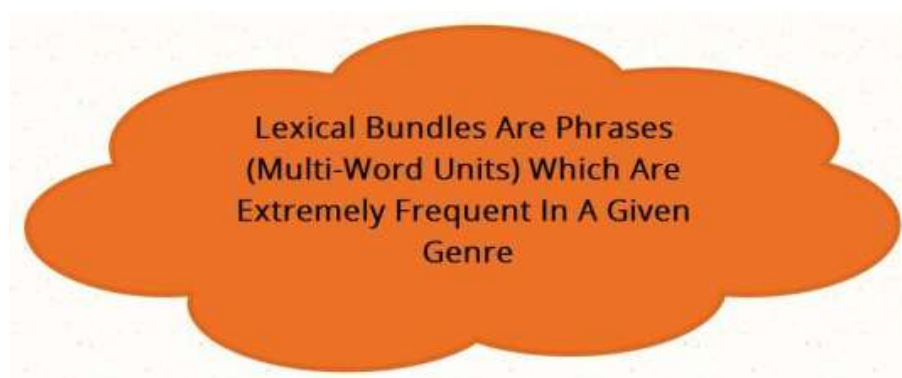
SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

POWERPOINT SLIDES FOR THE INTRODUCTORY SESSION



354

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
UNIVERSIDAD DE LA LAGUNA

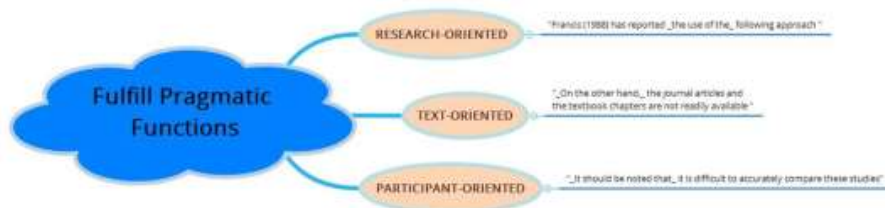
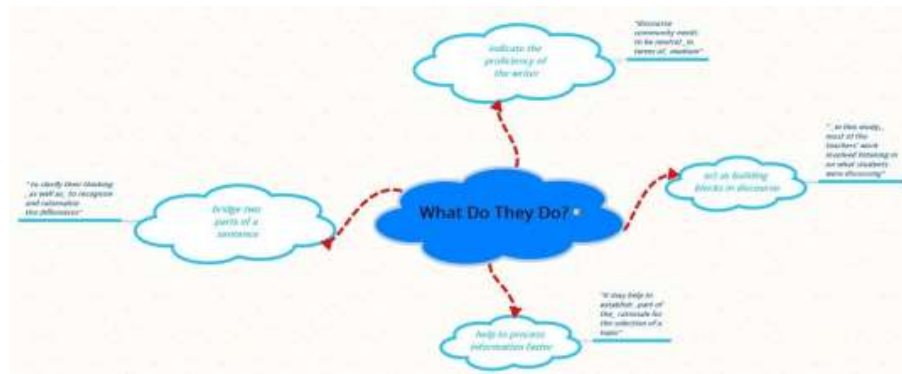
Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17



Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17

List 1	List 2
THE DEVELOPMENT OF	IN ORDER TO
A NUMBER OF	PART OF THE
THE CONCEPT OF	ONE OF THE
THE RELATIONSHIP BETWEEN	THE FACT THAT
AS PART OF	IN THE SAME
THE WORK OF	USE OF THE
A VARIETY OF	THE IDEA OF
THE NOTION OF	THE IMPORTANCE OF
THE CONSTRUCTION OF	THE STUDY OF
A FORM OF	IN OTHER WORDS
A SET OF	IN THE CASE OF

Este documento incorpora firma electrónica, y es copia auténtica de un documento electrónico archivado por la ULL según la Ley 39/2015.
 Su autenticidad puede ser contrastada en la siguiente dirección <https://sede.ull.es/validacion/>

Identificador del documento: 973669

Código de verificación: mnf6zN2e

Firmado por: DIANA BALASANYAN
 UNIVERSIDAD DE LA LAGUNA

Fecha: 30/06/2017 14:44:14

SALLY FRANCES BURGESS .
 UNIVERSIDAD DE LA LAGUNA

30/06/2017 21:06:24

ERNESTO PEREDA DE PABLO
 UNIVERSIDAD DE LA LAGUNA

14/07/2017 13:44:17