

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage



Carmen Pérez-Llantada*

University of Zaragoza, Faculty of Arts, Department of English and German Studies, 50009 Zaragoza Spain

A B S T R A C T

Keywords:

Academic writing
 English for academic purposes
 Research article genre
 Lexical bundles
 Register analysis
 Second language acquisition
 Interlanguage development

Formulaicity (i.e. knowledge of conventionalised multi-word combinations) in academic writing is not part of the native writer's innate language ability and is thus far from being a linguistic universal skill (Kachru, 2009; Wray, 2008). It can therefore be assumed that L2 academic writers find it particularly difficult to acquire native-like formulaic sequences. Building on this assumption, I use a 5.7 million-word corpus of expert academic writing to compare convergent and divergent usage of lexical bundles in three language variables, L1 English, L2 English and L1 Spanish. I identify core bundles (i.e., bundles shared by the three variables) and contend that writers' usage of these bundles is determined by register. I also compare the structures and functions of bundles specific to one or to two language variables to exemplify how these distinctive bundles build different pragmatic meanings in the texts. In identifying phraseological norms implicitly recognised by L1 writers, I argue that the use of bundles by the L2 writers deviates from L1 norms and conclude that, although they are expert writers, their formulaicity is 'hybrid', that is, largely, but not completely, native-like. I also discuss implications regarding L2 expert writers' interlanguage development and propose areas for pedagogical intervention.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Over the past decades, formulaicity (i.e., knowledge of conventionalised multi-word combinations) in academic writing has been investigated by several influential research strands. Taking a frequency-based approach, the North-American corpus linguistics school has taxonomised formulaic sequences (called 'lexical bundles') in both academic speech and writing. It has been consistently argued that each academic genre displays "a distinct set of lexical bundles, associated with [its] typical communicative purposes" (Biber & Barbieri, 2007, p. 265). Structural and functional descriptions of lexical bundles have also served to describe English academic writing in terms of grammatical compression, syntactic elaboration and degree of explicitness (Biber, 2009; Biber, Conrad, & Reppen, 1998; Biber & Gray, 2010; Biber, Johansson, Leech, Conrad, & Finegan, 1999). Formulaic language in American and British academic English varieties as well as in other varieties of academic writing such as Argentinean and Peninsular Spanish or Philippine, has also been typified within this research tradition (Biber, Conrad, & Cortes, 2004; Cortes, 2004; Liu, 2012; Pérez-Llantada, 2012; Salazar, 2010).

Formulaicity has also been investigated with a view to examining non-native English speakers' lexical bundle use and proposing pedagogical interventions for the teaching of English as a Foreign Language. Drawing on learner corpora, a number of lexico-grammatical, pragmatic and stylistic features in L2 English have been reported as deviant from L1 norms

* Fax: +34 976 76 15 19.

E-mail address: llantada@unizar.es.

(Granger, 1998; Granger & Meunier, 2008; Howarth, 1998; Meunier & Granger, 2008). There is also little dispute that advanced English learners use fewer formulaic sequences than their native-speaker counterparts (Granger, 1998) and that their language production exhibits “lack of register awareness, phraseological infelicities, and semantic misuse” (Gilquin, Granger, & Paquot, 2007, p. 319). In the field of English for Academic Purposes, phraseology research also maintains that L2 English learners with different proficiency levels overuse, underuse and misuse L1 English bundles and fail to understand their pragmatic functions according to L1 conventions (Ådel & Erman, 2012; Chen & Baker, 2010; Salazar, 2010; Staples, Egbert, Biber, & McClair, 2013).

Formulaicity has also been approached from the perspectives of psycholinguistics and language acquisition. It has been claimed that knowledge of academic formulas facilitates fluent language processing and that mastery of bundles equates to successful language production (Ellis, Simpson-Vlach, & Maynard, 2008; Schmitt, 2004; Wray, 2002). Research has also offered evidence that lexical phrases in L1 are learnt ‘as wholes’ and not as strings of individual words, that formulaic language is learnt incrementally and that fluent language users exhibit an ample repertoire of memorised language sequences (Ellis, 2008; Li & Schmitt, 2009). Li and Schmitt (2009, p. 86) further note that the absence of formulaic sequences in language production signals the “lack of mastery of a novice writer in a specific disciplinary community”.

Formulaicity in academic writing is not a language universal skill (Kachru, 2009; Wray, 2008). Both L1 and L2 academic writers may not only acquire formulaic sequences through formal instruction but also through non-formal incidental learning — e.g., extensive academic reading and repeated usage of patterns through extensive writing (Ellis, 2008; Li & Schmitt, 2009). Supporting Warren’s (2005, p. 38) claim that native-like mastery of idiomaticity is difficult to attain by foreign language learners, research with small-scale monolingual and multilingual corpora has demonstrated that formulaic language is associated with expert, and not novice, academic writing production (Cortes, 2004, 2008; Durrant & Mathews-Aydinli, 2011; Neff, 2008; Römer, 2009). To my knowledge, no corpus-driven studies to date have systematically contrasted the formulaicity of L2 English published writing *vis-à-vis* that of L1 English published writing using large-scale corpora. In an attempt to fill this gap, here I compare lexical bundles across three language variables of academic writing (L1 English, L2 English written by Spanish scholars and L1 Spanish). The aim is to ascertain to what extent formulaic language in L2 expert writing is native-like. The following questions helped to focus the investigation:

1. Which are the high-frequency lexical bundles in each language variable? What are the defining features of these bundles? Is choice of bundles determined by register/genre?
2. Which are core bundles shared by the three language variables? Which bundles are shared only by L1–L2 English and only by L2 English–L1 Spanish? Are these bundles similar or different structurally and functionally? How do these bundles build discourse meanings?
3. Finally, which bundles are distinctive to L1 English, L2 English and L1 Spanish? Do these bundles involve distinctive structures and functions? How do these bundles build discourse meanings?

In identifying the phraseological norms implicitly recognised by L1 writers, I discuss several possible reasons why the L2 English writers’ bundle usage deviates from L1 norms. It is argued here that, although the L2 English writers are published (and hence, expert) authors, their formulaic language is ‘hybrid’—largely, but not fully, native-like.

2. Methodology

The corpus used is the *Spanish–English Research Article Corpus* (SERAC 2.0), a 5.7-million word compilation of 1056 research articles (RAs) that comprises three sets of texts, each of them representing a ‘language’ variable. The first set of texts includes 360 L1 English RAs written by scholars from Anglophone-based contexts and published in peer-reviewed English-medium journals from different disciplinary fields. The second set comprises 336 L2 English original (not translated) RAs written by Spanish scholars and published in the same journals in which the L1 English texts were published. The third set includes 360 L1 Spanish RAs published by Spanish scholars in peer-reviewed Spanish journals targeted at a national-based scholarly readership. Selecting peer-reviewed journals was expected to guarantee that writers had experience in journal publications and thus familiarity with register/genre and style conventions in research writing. Table 1 shows the overall statistics.

Table 1
SERAC 2.0 statistics.

	L1 English	L2 English	L1 Spanish
Tokens (running words) in text	2,146,347	1,771,727	1,811,071
Types (distinct words)	54,184	51,020	70,190
Type/token ratio TTR	2.65	3.04	4.03
Standardised TTR	37.44	37.7	39.21
Standardised TTR std. dev.	62.42	62.75	61.40
Sentences	87,390	66,903	60,085
Mean (in words)	23.36	25.12	29.01
Std. dev.	15.13	16.01	19.15

Download English Version:

<https://daneshyari.com/en/article/360261>

Download Persian Version:

<https://daneshyari.com/article/360261>

[Daneshyari.com](https://daneshyari.com)