

# New Trends in Corpora and Language Learning

Edited by

Ana Frankenberg-Garcia,  
Lynne Flowerdew and Guy Aston



continuum

**Continuum International Publishing Group**

The Tower Building  
11 York Road  
London SE1 7NX

80 Maiden Lane  
Suite 704  
New York, NY 10038

www.continuumbooks.com

© Ana Frankenberg-Garcia, Lynne Flowerdew, Guy Aston and Contributors 2011

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without prior permission in writing from the publishers.

Ana Frankenberg-Garcia, Lynne Flowerdew and Guy Aston have asserted their right under the Copyright, Designs and Patents Act, 1988, to be identified as Author of this work.

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

ISBN: 978-1-4411-5996-0 (hardcover)

**Library of Congress Cataloging-in-Publication Data**

New trends in corpora and language learning / edited by  
Ana Frankenberg-Garcia, Lynne Flowerdew and Guy Aston.

p. cm. – (Corpus and discourse)

Includes bibliographical references.

ISBN 978-1-4411-5996-0 (hardcover)

1. Corpora (Linguistics) 2. Computational linguistics. 3. Discourse analysis.  
I. Frankenberg-Garcia, Ana. II. Flowerdew, Lynne. III. Aston, Guy. IV. Title.  
V. Series.

P128.C68N49 2010

418.0071–dc22

2010000321

Typeset by Newgen Imaging Systems, Pvt Ltd, Chennai, India  
Printed and bound in Great Britain

## Chapter 15

# Polishing papers for publication: palimpsests or procrustean beds?

John McKenny and Karen Bennett

This chapter analyses a corpus of academic papers written in English by established Portuguese academics that were presented to a language consultant for revision prior to submission for publication. The corpus was interrogated for the presence of certain discourse features and compared with a control corpus of published articles by native-English academics in a similar field. Its purpose was to examine the hypothesis that not only lexical and syntactic features, but also phraseological and discourse features of L1 may be transferred into L2, thereby undermining the 'naturalness' of the writing and raising an (invisible?) obstacle to international publication. The results reveal significant overuse of certain features by Portuguese academics, and a corresponding underuse of others. This may be due to differences in epistemological outlook, which raises issues of both a practical and an ideological nature for the reviser of those texts.

### 15.1. Introduction

Texts written by foreign academics for publication in international English language journals typically undergo a series of interventions before their final polished form is achieved. Amongst the various 'literacy brokers' (Lillis and Curry 2006) that make their mark on the text, one of the most crucial is the language reviser, who is recruited primarily to correct grammar and spelling mistakes, and ensure that the register is suitable for the journal in question.

While foreign researchers are usually quite happy for their prose to be doctored in this superficial way, there exists another dimension of revision that is more controversial. Senior academics with established reputations in

their own countries do not always take kindly to having their work meddled with at the *discourse* level – that is, when the reviser takes it upon herself to reformulate whole sentences or paragraphs, perhaps even excising chunks of text. Yet such radical surgery is sometimes necessary to bring the work into line with the norms of English academic discourse. That is to say, cultural differences at the discourse level are often unacknowledged by both foreign authors and Anglophone editors, leading to unfortunate misunderstandings. At worst, this may result in the text being rejected out of hand as ‘badly written’ (Frankenberg-Garcia 1990a).

The question of L1 transfer has of course been widely discussed within the EAP context, and in recent decades, this has gone beyond mere grammar and lexis to include aspects such as text organisation and cohesion. Since Kaplan (1966) first suggested that there are cultural differences in discursive or expository writing patterns, there has been a plethora of comparative studies from a variety of cultural perspectives (e.g. Duszak (ed.) 1997; Smith (ed.) 1987; Ventola and Mauranen (eds) 1996), culminating in the formal constitution of the discipline that is today known as Contrastive Rhetoric (Connor 1996). Hence, English academic writing has been compared to ‘teutonic, gallic and nipponic’ styles (Galtung 1981), German (Clyne 1987a, 1987b, 1988), Indian languages (Kachru 1987), Czech (Cmejrková 1996), Finnish (Mauranen 1993), Polish (Duszak 1994), Norwegian (Dahl 2004) and Russian/Ukrainian (Yakhontova 2006), to name but a few. Despite the undoubted value of these studies, none of them have gone so far as to suggest that there may actually be a different epistemological framework underlying the scholarly discourse of non-Anglophone cultures. Yet in the case of Portugal and Spain, the powerful influence of the Catholic Church, supported by conservative political regimes, ensured that the Scientific Revolution – which radically altered attitudes to knowledge in seventeenth-century England, with far-reaching consequences for discourse (Halliday & Martin (eds) 1993: 2–21, 54–68; Martin 1998) – did not happen. Instead, an older text-based humanities tradition remained central to the education system right through to the late twentieth century.<sup>1</sup>

As a result, the kind of discourse typically used in humanities writing in Portuguese displays features that are quite different from those prescribed in English academic discourse style manuals (Bennett 2009).<sup>2</sup> While English values succinctness, clarity and objectivity, much Portuguese humanities writing is characterised by a general ‘wordiness’ and redundancy, a preference for a high-flown erudite register over the demotic (evident in both

syntactic and lexical choices) and a tendency towards abstraction and figurative language. There are also important differences as regards textual organisation, such as a propensity for indirectness, with the main idea often embedded, adorned or deferred at all ranks (Bennett 2010).<sup>3</sup>

It would not be surprising, then, if some of these features were to get transferred into Portuguese authors' English texts, even when those authors are very proficient in that language. For as with other kinds of social semiotic, questions of value frequently operate at the subconscious level and are 'taken for granted' in a way that grammar and spelling are not. Hence, it is not easy for the language reviser to convince her client that fewer words are better than many in English, or that a simple straightforward sentence is valued above a complex ornate one. Similarly, the editors of English-language journals may not realise that the presence of such features may be due to cross-linguistic transfer, rather than to ineptitude.

This chapter uses corpus analysis to procure empirical evidence of cross-cultural transfer at the discourse level in English-language research articles produced by Portuguese scholars. The study involves the comparison of two corpora, one consisting of English academic articles in the humanities written by Portuguese academics, and the other, similar articles from published journals by English L1 authors.

The project resulted from a collaboration between two language professionals from somewhat different backgrounds: a corpus linguist, and a professional language consultant/translator (who is also a researcher specialising in the critical study of academic discourse). This collaboration led to a few methodological tussles. For example, the corpus linguist, in accordance with accepted practice in his field (Sinclair 2004; Tognini Bonelli 2001), was in favour of approaching the data without preconceptions, following wherever it might lead; the language consultant, on the other hand, was interested in testing certain hypotheses that had been formulated over the course of her professional practice. Ultimately this problem was resolved by dividing the study into two distinct phases, as described below.

## 15.2. Corpora and procedures

Each of the two corpora contained around 113,000 running words. The corpus of texts by Portuguese authors, dubbed *Portac*, was essentially opportunistic, as it was made up of texts that had been presented to the

language consultant for revision prior to submission for publication in English-language journals. These were all articles in the Humanities or Arts, and the authors were all senior Portuguese academics with a certain reputation in their home culture. Naturally, the authors' agreement was obtained prior to their inclusion in the corpus.

The control corpus (Controlit) consisted of published articles from journals which the Portuguese authors hoped to be or had been published in, and which were available electronically from our university library. We included only articles written by single authors who were native speakers of English, and which dealt with similar areas of interest to those of our Portac writers.<sup>4</sup>

Two software suites were used in a complimentary fashion. Wmatrix2 (Rayson 2003)<sup>5</sup> enables the investigator to compare two corpora and continually shift focus as trends become apparent; that is to say, researchers may quickly compare lexical and grammatical dimensions from the perspective of one or other corpus. Wordsmith Tools (Scott 2004) was used to carry out analyses not available in Wmatrix2, such as frequency counts of word clusters, and word/n-gram searches using wildcards (for example, for polysyllabic noun forms, a frequency list of all words ending in \*ion). Results of corpus comparisons are expressed in terms of Log Likelihood (henceforth LL).<sup>6</sup>

The analysis was undertaken in two phases. The first phase set out to test the intuitions of the language reviser and therefore focused on particular features that had been noted during the course of her professional practice. The second phase explored issues that had arisen unexpectedly during the first phase and was therefore more properly corpus-driven.

### 15.3. Results

The results of the first phase have already been described elsewhere (McKenny and Bennett 2009) and will thus be merely summarised here. They include the following:

(a) Overuse of nominalisation:

The writing of the Portuguese academics showed a high degree of nominalisation compared to the control corpus. At the level of individual words, there was an overuse of nouns, both singular (LL 25.17) and plural (LL 69.81), and, concomitantly, a greater use of indefinite and definite articles (LL 43.81 and LL 36.13 respectively).

## (b) Underuse of most pronouns:

There was marked underuse of pronouns in Portac, 6,154 (6.11% of all text) vs. 8,671 in Controlit (8.49%), giving a surprising LL of -394.98. This underuse of pronouns is more likely to be a consequence of nominalisation for, as Biber et al. (1999: 92) point out, after analyzing written corpora totalling 40 million words, 'a high frequency of nouns . . . corresponds to a low density of pronouns'. In particular, *he* (LL -232.00), *she* (LL -104.00), *him* (LL -96.00), *I* (LL -39.00), *me* (LL -37.00), *it* (LL -25.74) were all underused.

## (c) Overuse of 'we' and 'us':

On the other hand, *we* (LL 39.41) and *us* (LL 16.85) were overused. This cannot be attributed to multiple authorship, as all the articles in our corpora were written by a single author. This would seem to directly result from the transfer of L1 discourse habits.<sup>7</sup>

## (d) Overuse of the genitive:

There was also a startling overuse of the genitive, both singular and plural ('s and s') (LL -211.64), and also of the alternative construction using *of* to express the same relationship (LL -34.03). In many cases, this seemed to be directly related to the tendency to over-nominalise, as in the example *a comment on the possibilities of the play's staging*, which was reconstrued by the reviser using a clausal form (i.e. *a comment upon how the play might be staged*).

## (e) Underuse of subordination:

Using Wmatrix2 and the POS tagged versions of the corpora it was found that Portac writers underuse the subordinating conjunctions *if, because, unless, so, for, although, while* (LL -8.16). They greatly overuse coordinating conjunctions, such as *and, or, nor*, (LL 26.17), with the exception of *but* which is tagged separately in CLAWS7 as CCB, adversative coordinating conjunction (LL -14.74). This apparent underuse of subordination was somewhat surprising, and ran counter to the language reviser's intuitions. More corpus investigation, however, is needed of other subordinate structures such as non-finite, verbless, relative and correlative clauses.

## (f) Overuse of embedding structures:

On the other hand, the Portac writers tend to make greater use of certain kinds of embedding or matrix structures (*We can see that . . . ; It should*

*be pointed out that . . .*), particularly to carry epistemic stance. Searches using Wordsmith Tools for four variable structures (*It \* \* that; It is \* \* that; We \* that; We \* \* that*) all yielded higher frequencies for Portac than for Controlit.

(g) Preference for Latinate vocabulary:

Portuguese authors favour polysyllabic abstract nouns of Latinate origin. Using Wordsmith Tools to search for *-ion*, 2,184 instances were found in Portac compared to only 1,458 in Controlit (LL 163), while the results for *-icity*, *-isation* and *-ation* gave LL 7.07, LL 14.16 and LL 50.71 respectively. There was also a high frequency of the indefinite article *an* (LL 18.65), which, according to Hofland and Johansson (1982: 22), indicates a high proportion of Latinate word tokens. This is perhaps to be expected given the derivation of their mother tongue from Latin.

(h) Overuse of reformulation markers:

This particular search was stimulated by a study by Cuenca (2003) into the usage of reformulation markers in academic English compared to similar writing in Spanish and Catalan. A comparison of our two corpora revealed a higher occurrence of reformulation markers (*namely, i.e., e.g., that is, in other words*) in the writing of the Portuguese academics (LL 67.76). This is likely to result directly from the transfer of discourse conventions from the L1.

(i) Overuse of prepositions:

In the initial Wmatrix contrast of the two corpora, significant overuse of prepositions by the Portac writers was apparent (LL 46.32). As noted above, *of* was a main contributor to this overuse (LL 31.31). A closer scrutiny revealed that multi-word prepositions (Granger and Meunier (eds) 2008) also showed a difference between the two corpora (LL 13.19), of which the most significant were *with regard to, by means of, with reference to, in spite of, in view of, in connection with, by way of, in front of, in conjunction with, in common with*. As these multi-word prepositions bear a fairly close resemblance to compound prepositions frequently used in Portuguese, it is likely that this overuse results from a simple process of cross-linguistic transfer.

(j) Overuse of multi-word expressions:

Wmatrix automatically extracts what it calls multi-word expressions (MWEs) from corpora by using a large-scale semantically classified multi-word expression template database. The definition of MWE is fairly straightforward. Only word sequences in or predicted by the dictionary of the semantic tagger will count as an MWE. The decisions therefore have been taken in advance. Users of Wmatrix have the option of customising their MWE dictionary. One source of such enrichment is from perusing lists of N-grams extracted automatically by Wmatrix. N-grams are recurrent sequences of words ( $N > 2$ ) with a frequency of three or more as the default in Wmatrix. Wordsmith Tools uses the term *cluster* for the same notion.

More than 1,600 types of such MWEs were detected in both corpora. *Portac* had 5,756 tokens of MWEs as opposed to 4,772 tokens in *Controlit* (LL 92.10). This suggests that there are significant differences in the balance between novel and formulaic language in the two groups of writers. The more well-stocked the Wmatrix MWE dictionary becomes, the more accurate will be the automatic measure of the formulaicity in a corpus.

These initial findings on multiword expressions led us to believe that *Portac* authors may be less creative in their language use than the native speakers, and more dependent upon ready-made or formulaic expressions. Hence, the second phase of our study focused upon the phraseology used by our two groups of writers. We looked, firstly, at the lexical choices made by these authors in the light of word frequencies, and secondly, at some of the collocational patterns, as reflected in the N-grams found in the two corpora.

### 15.3.1. Lexical choices

Other corpus studies of the lexical choices made by advanced learners of English have suggested that they tend to underuse the most frequent words of the language. Fox and her colleagues on the COBUILD team found that the language of advanced foreign students 'is often too stilted, too formal and too high-level; and when it is analysed it is seen that the most common words are used less frequently and in fewer contexts than they would be by native speakers of English' (Fox 1998: 27).

Suspecting that the same phenomena might be found in the writing of our Portuguese authors, we compared frequencies in our two corpora of

the most frequent words in the BNC using Leech et al.'s (2001) frequency lists. The results are shown in Table 15.1:

**Table 15.1** Distribution of the most frequent words in the BNC in Portac and Controlit

	Portac	Controlit
100 most frequent words	59,618	65,892
101st–201st most frequent words	11,653	12,914
Total	71,271	78,806
Coverage	72%	79%

This would seem to seem to confirm Fox's claims. However, it is unclear whether this is due to the language learning career of Portac authors (in the sense that core words were only given attention at elementary and intermediate level of instruction and then neglected at higher levels) or to the transfer of discourse habits from the L1. That is to say, as Portuguese writing in the humanities tends to be very recondite and formal, it is possible that this underuse of everyday words reflects a deep-rooted sense that such vocabulary is inappropriate to the academic register.

### 15.3.2. Collocational patterns

When the collocational patterns in the two corpora are examined, the strength of the corpus-driven approach is clearly demonstrated. Wmatrix2 delivered a clear picture of the frequency of recurrence of N-grams. The results given in Table 15.2 show the startling discrepancy between the word sequencing of L2 writers and their L1 counterparts.

**Table 15.2** N-gram tokens in the two corpora

Ngram	Portac	Controlit	Log Likelihood
5 grams	330	272	6.01
4 grams	1661	1058	134.85
3 grams	10050	7288	304.85
2 grams	47263	43174	184.94

It seems that the Portac writers are using more recurrent sequences (of two, three, four words and so on), making their writing more predictable and less creative than that of the native writers in Controlit.

However, we must be careful, when talking about N-grams or clusters, not to make assumptions about the formulaicity of the language or the use of multiword expressions (MWEs). N-grams do not provide a direct measure of the degree of 'chunkiness' (de Cock 2000) contained in a text or corpus. An N-gram is merely a sequence of consecutive words generated by the syntax or is a fragment thereof. To get from N-gram lists to formulaic sequences, it is necessary to apply a sequence of filtering rules, the strictness of which will determine the number of N-grams which are pronounced to be prefabricated or formulaic sequences. In this case, the following criteria were used to filter N-grams to obtain formulaic sequences.

- Obvious non-idiomatic sequences were eliminated: e.g. most sequences beginning with *and*, *a*, *an*, *he*, *the* and *they* (with some exceptions such as *a lot* and *the world*);
- Free combinations bound only by morphosyntactic rules were retained if they had some additional pragmatic or textual function. *I think (that)* was retained whereas *they are* was not;
- Recurrent sequences ending in *a*, *an*, and *the* were eliminated (This differs from Biber et al.s' (1999) lexical bundle approach);
- Substitution tests of constituent words were used to ascertain whether there was a degree of fixity or formulaicity in a sequence;
- Sequences with syntactic unity and a clear grammatical function in their context (e.g. compound verbs, nouns, adjectives, adverbs, prepositions, connectives, quantifiers) were counted.

Taking the fifty most frequently occurring 4-grams from Portac and applying these filtering rules, we were left with eighteen sequences in Table 15.3 which might be considered formulaic.

**Table 15.3** Most frequent four-word formulaic sequences in Portac after filtering is applied

the end of the	17
one of the most	17
on the other hand	16
i would like to	14
at the same time	13
at the end of	13
as one of the	12
in the sense that	12

(Continued)

**Table 15.3** (Cont'd)

---

the return of the	11
with regard to the	11
the beginning of the	10
the extent to which	10
in the light of	10
in the case of	9
the middle of the	9
on the one hand	8
as well as the	8
return of the caravels	8
in the mysteries of	8
be found in the	8
can be said to	7
by a sense of	7
the history of the	7
the title of the	7
the work of art	7
as much as in	7
in the form of	7
un coup de d	6
the fact that the	6
the figure of the	6
the representation of the	6
for the first time	6
in the middle of	6
mill on the floss	6
as we have seen	6
the latter part of	6
of some of the	6
the mill on the	6
in the course of	6
the light of the	6
neither here nor there	6
as much as of	6
can be found in	6
at the beginning of	6
the way in which	6
in the work of	6

---

*(Continued)*

**Table 15.3** (Cont'd)

the case of the	6
of the work of	6
with a sense of	5
it is common to	5

Moreover, when the semantic tagger was used to extract the MWEs in the two corpora, the frequencies shown in Table 15.4 were obtained.

**Table 15.4** Comparison of multi-word expressions

	Portac	Controlit
MWE tokens overused in Portac	3861	1527
MWE tokens underused in Portac	1661	3253
Total in each corpus	5522	4470

The twenty MWEs overused by Portac writers in comparison to Controlit authors (all with LL > 8) are listed in Table 15.5. It is noticeable that they include a number of phrasal verbs.

**Table 15.5** Twenty MWEs overused by Portac authors

	Portac	Controlit
in_fact	87	14
with_regard_to	23	0
according_to	44	11
as_much_as	22	2
as_regards	13	0
in_the_picture	11	0
carried_out	10	0
in_question	17	2
due_to	14	1
brought_about	7	0
in_view_of	7	0
made_up	7	0
by_means_of	11	1
in_order_to	44	19
in_the_end	10	1
at_stake	6	0

(Continued)

**Table 15.5** (Cont'd)

put_forward	6	0
stand_for	6	0
sum_up	6	0
vantage_point	6	0
such_as	42	20

These are just the preliminary results of an area of study that could eventually prove to be very fruitful. They indicate that these Portuguese scholars tend to use more formulaic expressions in their L2 writing than their L1 counterparts, which could have important repercussions for language revisers and EAP professionals, as well as for the authors concerned.

#### 15.4. Discussion

Our results clearly suggest that the English writing of established Portuguese academics in the humanities differs in many respects from that of native speakers in the same field. In particular, the overuse of features such as nominalisation, embedding structures, Latinate vocabulary and reformulation markers, combined with the underuse of pronouns and everyday vocabulary, may have the cumulative effect of causing the prose to sound dense and abstract to native ears, while the prevalence of formulaic sequences may make it sound stilted and unnatural. In this section, we will discuss the extent to which such features may jeopardise these writers' chances of getting published in international journals, and how corpus linguistics might help remedy this situation.

It is important to distinguish between academic writing in science and in the humanities. Responding to the question of linguistic imperialism raised by Phillipson (1992), Pennycook (1994) and others, Tribble (2008: 308) contends that 'in professional and academic writing, both authorship and gate keeping authority have shifted and the production and evaluation of these texts is no longer a native speaker monopoly'. That is to say, in situations where English is being used as a *lingua franca*, native speaker accuracy is no longer an issue. Instead, Tribble argues, what counts is acceptability in the eyes of peers in the discourse community.

This claim may well be true for the sciences, where language is used as a transparent medium for transmission of a content that may be presented primarily in non-linguistic form (graphs, tables, diagrams, mathematical

formulae, etc). But in the humanities, where the content is frequently inseparable from the words used to convey it, the matter is much more complex, and failure to comply with the stylistic norms of the target culture will often be taken as a sign of intellectual ineptitude or sloppy thinking.

The problem is compounded by the fact that, unlike the discourse of science, which in most languages of the world is calqued from English, the discourse of the humanities may vary dramatically between cultures as a result of long native traditions. In the Romance cultures, as we have seen, there is evidence that quite different values govern the production of scholarly texts; hence, authors who blindly transfer discourse features from their own languages into English may unwittingly be striking a blow at the target-culture's whole epistemology, provoking in some cases an equally blind response from the gatekeepers of that culture.

Comparative corpus studies such as this one may be useful in raising awareness of the culturally contingent nature of academic discourse conventions. Within the ELT environment, such analyses have traditionally been used to make learners aware of the ways in which their prose differs from that of native speakers, with a view to instituting corrective pedagogies (see, for example, Biber et al. 1999; Fox 1998; Frankenberg-Garcia 1990b). However, there may also be a place for such studies within the critical tradition, which resists the Anglo-American academic hegemony in the interests of epistemological pluralism. Corpus-driven research can offer hard empirical evidence of cultural differences that may otherwise be conveniently overlooked by the gatekeepers of the hegemonic culture, and can perhaps open up the way towards less rigidly prescriptive attitudes to the construction of knowledge.

How does this affect the reviser of the academic text, caught as she is between two worlds? If, as has been suggested in this case, there are distinct discourse norms operating in the source and target cultures, should her role be to ruthlessly domesticate the foreign author's text, bringing it into line with the Procrustean demands of the target culture in order to ensure publication at all cost? Or should she rather seek to produce a kind of palimpsest that allows the thought patterns of the original to be glimpsed beneath the surface of the revised text? These are issues that have preoccupied translation scholars for some time now, and which are unlikely to be resolved while the Anglo-Saxon world maintains its hegemony over academic production.<sup>8</sup> However, corpus studies may have a part to play in raising awareness of hitherto-unperceived cultural differences, thereby encouraging greater acceptance of alternative ways of construing knowledge.

## Notes

- <sup>1</sup> Scientific ideas only began to filter into Portugal in the second half of the eighteenth century via a loose network of exiled intellectuals living abroad, known as the *estrangeirados* (Carneiro et al. 2000; Gomes 1995; Nunes 2002). Indeed, in Portugal, there were only two very brief periods before 1974 when Enlightenment views were implemented – the period in office of the Marquis of Pombal in the eighteenth century and the brief Republic of 1910–1926. For a more detailed exploration of the historical circumstances conditioning the development of Portuguese academic writing, see Bennett (forthcoming).
- <sup>2</sup> Contrastive studies by Martín Martín (2003), Moreno (1997), Mur Dueñas (2007a, 2007b) and Cuenca (2003) suggest that Spanish contains similar characteristics. This is to be expected, given the cultural proximity between the two countries.
- <sup>3</sup> See also Bennett (2006, 2007a, 2007b) for detailed analyses of excerpts of Portuguese humanities texts.
- <sup>4</sup> This was done on the basis of authors' names, which, though fallible, reduces the likelihood of using L2 writers in a corpus designed to represent L1 writing.
- <sup>5</sup> Available at <http://ucrel.lancs.ac.uk/wmatrix2.html>.
- <sup>6</sup> LL measures the likelihood that a difference between the observed frequency of an item and its expected frequency is not random. The higher the LL value, the more significant the difference between two frequency scores. An LL value of 3.8 or higher is significant at the level of  $p < 0.05$  and an LL of 6.6 or higher is significant at  $p < 0.01$ . See <http://ucrel.lancs.ac.uk/llwizard.html>.
- <sup>7</sup> Mur Dueñas (2007a) reports a similar phenomenon in Spanish.
- <sup>8</sup> For a discussion of these issues in the context of Translation Studies, see Venuti (1995).

## References

- Bennett, K. (2006), 'Critical language study and translation: the case of academic discourse', in J.F. Duarte, A.A. Rosa and T. Seruya (eds), *Translation Studies at the Interface of Disciplines*. Amsterdam and Philadelphia: John Benjamins, pp. 111–127.
- Bennett, K. (2007a), 'Galileo's revenge: ways of construing knowledge and translation strategies in the era of globalization'. *Social Semiotics*, 17, (2), 171–193.
- Bennett, K. (2007b), 'Epistemicide! The tale of a predatory discourse'. *The Translator*, 13, (2), 151–169.
- Bennett, K. (2009), 'English academic style manuals: a survey'. *Journal of English for Academic Purposes*, 8, (1), 43–54.
- Bennett, K. (2010), 'Academic discourse in Portugal: a whole different ballgame?' *Journal of English for Academic Purposes*, 9, (1), 21–32.
- Bennett, K. (forthcoming), *Academic Discourse, Hegemony and Translation*. Manchester: St Jerome Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999), *Longman Grammar of Spoken and Written English*. London: Longman.
- Carneiro, A., Simões, A. and Diogo, M.P. (2000), 'Enlightenment science in Portugal: The *Estrangeirados* and their communication networks'. *Social Studies of Science*, 30, (1), 591–619.

- Clyne, M. (1987a), 'Cultural differences in the organization of academic texts in English and German'. *Journal of Pragmatics*, 11, (2), 211–247.
- Clyne, M. (1987b), 'Discourse structures and discourse expectations: implications for Anglo-Germanic academic communication in English', in L. Smith (ed.), pp. 73–83.
- Clyne, M. (1988), 'Cross-cultural responses to academic discourse patterns'. *Folia Linguistica*, 22, (3–4), 457–475.
- Čmejková, S. (1996), 'Academic writing in Czech and English', in E. Ventola and A. Mauranen (eds), *Pragmatics and Beyond*. Amsterdam and Philadelphia: John Benjamins, pp. 137–153.
- Connor, U. (1996), *Contrastive Rhetoric: Cross-Cultural Aspects of Second-Language Writing*. Cambridge: Cambridge University Press.
- Cuenca M.-J. (2003), 'Two ways to reformulate: a contrastive analysis of reformulation markers'. *Journal of Pragmatics*, 35, (7), 1069–1093.
- Dahl, T. (2004), 'Textual metadiscourse in research articles: A marker of national culture or of academic discipline?' *Journal of Pragmatics*, 36, (10), 1807–1825.
- De Cock, S. (2000), '“Repetitive phrasal chunkiness” and advanced EFL speech and writing', in C. Mair and M. Hundt (eds), *Corpus Linguistics and Linguistic Theory*. Amsterdam and New York: Rodopi, pp. 51–68.
- Duszak, A. (ed.) (1997), *Cultural Styles of Academic Discourse*, Berlin and New York: Mouton de Gruyter.
- Duszak, A. (1994), 'Academic discourse and intellectual styles'. *Journal of Pragmatics*, 21, (3), 291–313.
- Fox, G. (1998), 'Using corpus data in the classroom', in B. Tomlinson (ed.), *Materials Development in Language Teaching*. Cambridge: Cambridge University Press, pp. 25–43.
- Frankenberg-Garcia, A. (1990a), *Second Language Writing Instruction: a study of the effects of a discourse-oriented programme upon the ability of skilled writers to improve their written production*. Unpublished PhD thesis. Edinburgh University, UK.
- Frankenberg-Garcia, A. (1990b), 'Do the similarities between L1 and L2 writing processes conceal important differences?' *Edinburgh Working Papers in Applied Linguistics*, 1, 91–102.
- Galtung, J. (1981), 'Structure, culture, and intellectual style: an essay comparing Saxon, Teutonic, Gallic and Nipponic Approaches'. *Social Science Information* 20, (6), 817–856.
- Gomes, J.F. (1995), *Para a História da Educação em Portugal: Seis Estudos*. Porto: Porto Editora.
- Granger, S. and Meunier, F. (eds) (2008), *Phraseology: An Interdisciplinary Perspective*. Amsterdam and Philadelphia: John Benjamins.
- Halliday, M.A.K. and Martin, J.R. (eds) (1993), *Writing Science: Literacy and Discursive Power*. Pittsburgh and London: University of Pittsburgh Press.
- Hofland, K. and Johansson, S. (1982), *Word Frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities.
- Kachru, Y. (1987), 'Cross-cultural texts, discourse strategies and discourse interpretation', in L. Smith (ed.), pp. 87–100.
- Kaplan, R.B. (1966), 'Cultural thought patterns in inter-cultural education' in K. Croft (ed.), *Readings on English as a Second Language*, Cambridge MA: Winthrop (2nd edn. 1980), pp. 399–418.

- Leech, G., Rayson, P and Wilson A. (2001), *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman.
- Lillis, T. and Curry, M.-J. (2006), 'Professional academic writing by multilingual scholars: interactions with literacy brokers in the production of English-medium texts'. *Written Communication*, 23, (1), 3–35.
- Martin, J.R. (1998), 'Discourses of science: recontextualisation, genesis, intertextuality and hegemony', in J.R. Martin and R. Veel (eds), *Reading Science: Critical and Functional Perspectives on Discourses of Science*. London and New York: Routledge. pp. 3–14.
- Martín Martín, P. (2003), 'A genre analysis of English and Spanish research paper abstracts in experimental social sciences'. *English for Specific Purposes*, 22, (1), 25–43.
- Mauranen, A. (1993), 'Contrastive ESP rhetoric: metatext in Finnish-English economics texts'. *English for Specific Purposes*, 12, (1), 3–22.
- McKenny, J. and Bennett, K. (2009), 'Critical and corpus approaches to academic text revision: a case-study of articles by Portuguese humanities scholars'. *English Text Construction*, 2, 228–245.
- Moreno, A.I. (1997), 'Genre constraints across languages: causal metatext in Spanish and English RAs'. *English for Specific Purposes*, 16, (3), 161–179.
- Mur Dueñas, P. (2007a), '“I/We focus on . . .”': A cross-cultural analysis of self-mentions in business management research articles'. *Journal of English for Academic Purposes*, 6, (2), 143–162.
- Mur Dueñas, P. (2007b), 'Same genre, same discipline; however, there are differences: a cross-cultural analysis of logical markers in academic writing'. *ESP Across Cultures*, 4, 37–53.
- Nunes, M.F. (2002), 'Opinião pública, ciência e tecnologia em Portugal XVIII–XX'. *Cultura: Revista de História e História de Ideias*, vol. XV (2nd series), 211–223.
- Pennycook, A. (1994), *The Cultural Politics of English as an International Language*. New York: Longman.
- Phillipson, R. (1992), *Linguistic Imperialism*. Oxford: Oxford University Press.
- Rayson, P. (2003), *WMatrix: a statistical method and software tool for linguistic analysis through corpus comparison*. Unpublished PhD thesis. Lancaster University, UK.
- Scott, M. (2004), *Wordsmith Tools 4*. Oxford: Oxford University Press.
- Sinclair, J. (2004), *Trust the Text: Language, Corpus and Discourse*. London and New York: Routledge.
- Smith, L. (ed.) (1987), *Discourse Across Cultures: Strategies in World Englishes*. London: Prentice Hall.
- Tognini Bonelli, E. (2001), *Corpus Linguistics at Work*. Amsterdam and Philadelphia: John Benjamins.
- Tribble, C. (2008), 'In this paper . . . Some emerging norms in lingua franca English writing in the sciences?', in *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbon: ISLA.
- Ventola, E. and Mauranen, A. (eds) (1996), *Academic Writing: Intercultural and Textual Issues*. Amsterdam and Philadelphia: John Benjamins.
- Venuti, L. (1995), *The Translator's Invisibility: A History of Translation*. London and New York: Routledge.
- Yakhontova, T. (2006), 'Cultural and disciplinary variation in academic discourse: the issue of influencing factors'. *Journal of English for Academic Purposes*, 5, (2), 153–167.