REPORT

Trajectories of knowledge production: English medium academic writing for national, transnational and international journals

1. Background

The dominance of English as the global language of scholarly publishing in almost every academic field is well documented. More than 90% of the social science journals indexed by the Institute for Scientific Information (Thompson Corp., 2008) and 80% of the 59,591 scholarly periodicals indexed by Ulrich's Periodical Directory (2007) are published in English. English might therefore be viewed as an academic lingua franca (EALF) increasing opportunities for global participation in academic knowledge making and exchange across national borders. However, findings from our previous research (Curry and Lillis, 2004, 2006a, b) indicated that the global status of English is creating significant obstacles for scholars researching and writing from non-Anglophone contexts and is impacting on written academic knowledge production and dissemination in unacknowledged ways: in particular, that the different contexts of English medium publication (national or international) influence both the knowledge that is produced and the ways in which it is circulated.

Our project investigated the production and publication of English medium academic journal articles in one disciplinary area, psychology. Using a text-oriented ethnographic approach (see Lillis and Curry 2006a) we tracked how articles by non-Anglophone scholars from four European countries – Hungary, Slovakia, Spain and Portugal – were produced, from early drafts through successive stages towards publication. Using a corpus of 1.5 million words, we compared the reporting of knowledge by non-Anglophone and Anglophone psychologists in two distinct contexts of publication: English medium national journals, that is English medium journals published in Hungary, Slovakia, Spain and Portugal (henceforward EMN) and English medium international journals, that is journals usually classified as ‘international’ because of their high status or high ‘impact factor’ (henceforward EMI).

The goal of the project and the longitudinal study of which it forms a part, is:

1) to identify the specific ways in which English is (or is not) functioning as an academic lingua franca and influencing written academic knowledge production and exchange in the twenty first century;

2) to make findings available to relevant user groups, such as editorial boards and international research bodies and professional groups (translators, proofreaders, teachers) who work to support non-Anglophone scholars.
2. Objectives

The research questions we sought to address were as follows:

1. What are the similarities and differences in the English medium academic journal articles produced by non-Anglophone psychology scholars in Hungary, Slovakia, Spain and Portugal for the three contexts of publication: EMN, EMT (English medium transnational, henceforward EMN) and EMI?

2. How do specific practices surrounding text production impact on trajectories towards publication in each of the three contexts of publication: EMN, EMT and EMI?

3. Is it possible to identify and categorise specific patterns of similarities and differences between academic texts produced for the three contexts of publication – EMN, EMT and EMI – and if so, what are these?

4. How robust is the distinction between the three proposed English medium contexts of journal publication – EMN, EMT and EMI?

5. What benefits does a combined ethnographic, text and corpus methodology offer to researchers exploring academic writing and publication practices?

We modified these research questions in light of a key methodological decision in building the corpus, TAK_Psych (dataset 2); this decision was to replace the EMT category with an EMI 2 category for a number of reasons. Firstly, EMT was found to be a highly relational category which could be more accurately explored through qualitative case study methods (see 4.5). Secondly, we considered it important to build a reference subcorpus which could be used as a point of comparison between texts produced by non-Anglophone and Anglophone scholars. The TAK_Psych corpus is thus constituted by 3 main subcorpora: EMN and EMI 1 (both of which consist of articles written by scholars from the 4 national sites); and EMI 2 (which consists of articles written by scholars affiliated to Anglophone institutions).

Using text, ethnographic and corpus data (see 3) and under each of the original research questions, we focused on the following specific issues that emerged during our research and are discussed in section 4 of the report:

1. The dominance of English as the medium of Psychology journals.
2. Citation practices in EMN and EMI journal articles by non-Anglophone and Anglophone scholars.
3. The politics of style – sentence length.
4. The validity of the categories EMN, EMT, EMI.
5. Issues surrounding combining text, ethnographic and corpus methodologies.
3. Methods

3.1 Data Collection

In order to carry out our study, we collected three main data sets:

**Dataset 1 Text Histories** comprises 61 “text histories” from 21 psychology scholars. Text history data comprises drafts and published versions of texts, interviews with scholars about the production of these texts, correspondence between ‘brokers’ (i.e. reviewers, editors, colleagues, translators) and scholars, as well as institutional documentation about the function of academic publications within institutional rewards systems. (See Annex 2).

**Dataset 2 TAKPsych Corpus** comprises a corpus of 1.5 million words of English medium psychology articles published between 2000–2007 made up of 3 subcorpora: 1) EMN 1- English Medium National articles by authors affiliated to non-Anglophone institutions (in Slovakia, Hungary, Spain and Portugal); 2) EMI 1- English Medium International articles (from journals listed in the Social Science Citation Index, henceforward SSCI) by authors affiliated to non-Anglophone institutions (in Slovakia, Hungary, Spain and Portugal); 3) EMI 2 - English Medium International articles (from journals listed in the SSCI) by authors affiliated to Anglophone ‘centre’ institutions (predominantly in the UK and US). Articles were selected from experimental and non-experimental paradigms so as to represent the broad spectrum of psychology. (See Annex 3).

**Dataset 3 Ethnographic data** comprises field notes taken by two researchers in each of the 4 sites, departmental and national documentation on systems of research evaluation, and data on the material conditions in which scholars work. (See Annex 4).

In addition, we carried out extensive internet searches to find examples of English medium journal publications in the field of psychology and searches of journal indices such as the SSCI.

3.2 Data Analysis

A key aim of the project was to integrate qualitative and quantitative approaches to data analysis using methodological tools associated with corpus analysis, text analysis and ethnography. The careful tracking and analysis of specific text histories (dataset 1) using a previously developed heuristic alongside ethnographic data (Lillis and Curry 2006a) enabled us to develop rich descriptions of the processes and practices involved in writing for English medium publications, including the decisions scholars make, the texts they produce, the difficulties they face, how editors and reviewers respond to articles submitted for publication and how scholars respond to these. Corpus analysis, including key word searches and concordances using Wordsmith 4.0 software (Scott 2004), enabled us to explore whether key issues arising in the text histories were evident in a larger sample of psychology articles and to explore patterns of difference and similarity in articles published in EMN and EMI journals.
(see Annex 5 for example of searches on ‘Hungarian’; see Annex 6 for example of searches on 19*/20* to extract instances of citations). We also commissioned the development of software, *Countword*, aimed at investigating a specific stylistic feature in the published articles, that of sentence length (see Annex 7; see 4.3). Together these approaches enabled us to build a rich picture of the use of English for academic publication in the disciplinary area of psychology.

4. Results

4.1 The dominance of English as the medium of Psychology journals

The dominance of English as the written medium of academic journals in the field of psychology is evident both from an analysis of the high status index, the SSCI, and from the growth of English medium journals produced in contexts where English is not the main or official means of communication.

The SSCI index, managed by Thomson Scientific, is widely used as a normative approach to measuring the status and importance of journals, through calculations of a journal’s ‘impact factor’ (see 4.2). The journals listed in the index are therefore both formally, through the notion of ‘impact factor and informally through the status of this index, accorded prestige (the index can be accessed at http://www.thomsonreuters.com).

A search of the index (August 2008) showed that of the 459 psychology journals listed, 36 are in languages other than English. In relation to the 4 specific sites of our study, the SSCI indexes no Hungarian or Portuguese medium psychology journals, 1 in Slovak and 3 in Spanish (out of a total of 5 Spanish medium psychology journals listed in the SSCI, 3 are published in Spain and 2 in Latin America). Secondly, not only is the SSCI heavily skewed towards English medium journal publications but to English medium Anglophone centre journals, predominantly from the US (for centre/periphery see Wallerstein 1991). Thus in the field of psychology, of 423 English medium psychology journals listed in the SSCI, 394 are from US and UK based publishing contexts.

There is evidence of the growth of English medium ‘national’ journals in the field of psychology, a phenomenon long since noted in the field of natural and medical sciences (Ammon 2001; Swales 1997; Salager-Meyer 1997). These are journals produced and published in non-English dominant contexts. In our study, a total of 16 English medium psychology journals were identified in the contexts of Spain, Slovakia, Hungary and Portugal. Some of these were journals established from the outset as English medium, such as the Hungarian EMN *Journal of Evolutionary Psychology*; whilst others were long standing journals which appear to be switching to English medium (or English alongside another language), such as the Spanish journal *Psicologica* (see 8). English medium ‘national’ journals are notably absent from the SSCI. Of the 16 EMN journals we identified, only 3 are listed in the SSCI.

Ethnographic accounts indicate that scholars are aware of the usually lower status accorded to EMN rather than EMI (SSCI indexed journals; see 4.4) particularly in national and international systems of research evaluation. However, scholars gave a number of reasons for continuing to publish in these journals:
a) they are committed to publishing in nationally produced journals (whether in English or in national languages);

b) they want to make research more widely available (‘few people understand Hungarian’);

c) they want to support the development of national journals which are more internationally accessible because of the linguistic medium, English, but can also be shaped by national academic interests and concerns;

d) they are keen to encourage students and colleagues to participate in international debate within their fields, by writing and reading papers in English.

Some scholars choose to publish in national journals in national languages, whereas some scholars choose to publish in EMN journals; of the 21 scholars whose text histories were studied in this phase of the project, 8 had published in EMN journals, as well as EMI journals; in contrast, 13 had published in local national languages when publishing nationally, rather than in English medium national journals, in addition to EMI journals (see Annex 2). Preliminary analysis indicates that EMN journals may offer publishing opportunities for research which is inter- or trans-disciplinary in nature which are not available within the more intradisciplinary approach often adopted in EMI journals (see 8).

A key issue with regard to the position of English as an academic lingua franca (EALF) is that there are two potentially contradictory forces at work; that is, that there is strong evidence of scholars’ interest and effort towards using English as an EALF in order to share and build knowledge across national borders; yet at the same time there is a highly stratified system of journal evaluation in place which privileges English medium texts from Anglophone centre contexts (EMI) rather than EMN contexts. Furthermore, preliminary investigation into the circulation and readership of EMN journals suggests that these journals occupy a fragile position and may at times be failing to reach both national and international communities (see 8).

4.2 Citation practices in EMN and EMI journal articles by non-Anglophone and Anglophone scholars

A key issue that emerged in the course of the study was that of citation practices in EMN and EMI journals, by Anglophone and non-Anglophone scholars. Given the centrality of citation to knowledge building and to knowledge evaluation, we considered this an important issue to pursue. By ‘citations’ we mean both final bibliographical references and in-text citations and by ‘practices’ we mean both how scholars use citations and how these are received by reviewers and editors. Citations are a conventional feature of academic writing with a significant number of studies carried out from two key disciplinary perspectives: the first from an applied linguistics/discourse perspective and the second from a bibliometric perspective. Applied linguistics/discourse research on citations has focused on a range of linguistic and rhetorical features including reporting clauses (Charles 2006), evaluation (Thompson and Yiyun 1991) and rhetorical functions (Harwood 2008). In contrast, bibliometric studies focus on the ways in which citations are quantified in order to
evaluate and measure knowledge production and dissemination, notably through the notion of ‘impact factor’. In such studies, the interest is less on the rhetorical function of citations (as in applied linguistics research) and more on documenting (and sometimes critically exploring) the way that citations are used to measure and value journals (see for example Buela-Casal et al. 2006; Bornmann and Hans-Dieter 2008). In our analysis we used approaches from both applied linguistics and bibliometric studies (as discussed by Swales 1986).

In order to investigate citation practices we analysed data from the three datasets; one text history which we viewed as a ‘telling case’ (Mitchell 1984); bibliographic references from 240 articles in the TAK Psych corpus; in-text citations in the Portuguese EMN subcorpus; ethnographic accounts drawn from scholars about their views on citation practices. Key findings can be summarised as follows.

Corpus analysis shows that:

1. English medium citations dominate, both as bibliographical references and as in-text citations.
2. Articles written by non-Anglophone scholars writing in EMI journals include fewer non-English medium citations than in EMN journals.
3. Anglophone scholars rarely cite works in languages other than English.
4. Self-citation is higher in articles written by Anglophone scholars than by non-Anglophone scholars.

Text history and ethnographic data shows that there is evidence that:

5. Some reviewers from EMI journals view citations in languages other than English negatively.
6. Non-Anglophone scholars are highly aware of the contested nature of citation practices, particularly the tension between making local/national research visible and getting their articles published in EMI journals where reference to ‘local/national’ research and publications may not be valued.

The conclusion we reached from this analysis is that whilst the disciplinarity of a community is emphasized in existing studies on citations (see for example Hyland 1999), the importance of geolinguistic dimensions to the citation practices of disciplinary communities tend to be backgrounded, particularly where the same linguistic medium –English here – is the focus. Focusing solely on English medium or English as an academic lingua franca as if this were one linguistic community signals a homogeneity within disciplinary knowledge making which underestimates the range of practices and evaluation systems in play. Of particular concern is the negative reaction against non-English medium citations as this has implications for which knowledge and research is being circulated in a global context (see Nominated Output 1).

4.3 The politics of style – sentence length in English medium publications
The issue of what counts as an acceptable or appropriate ‘style’ of English for securing the publication of articles in high status Anglophone centre journals is one recurring theme in cyclical talk with scholars. Scholars express concerns about the English they use, in different ways, sometimes indirectly through phrases such as my ‘poor English’, ‘brutal English’, ‘my Spanish English’, sometimes through reference to ‘style’, and sometimes through reference to specific features – the feature most commonly mentioned by participants across all four national contexts as being problematic is the ‘long sentence’. Whereas long sentences are a key feature of academic writing in many languages and often index scholarly activity and erudition (see for example discussion in Bennett 2007), scholars indicate that the ‘long sentence’ is viewed negatively by reviewers or editors involved in EMI journals. In order to explore this particular issue of style, we analysed data from the three datasets; two text histories; a comparative analysis of sentence length in the EMN subcorpora and the EMI 2 subcorpus; ethnographic accounts drawn from scholars about their views on citation practices. Key findings were as follows:

1. There is evidence of reviewers explicitly mentioning sentence length as a problem with the English ‘style’ of non-Anglophone scholars’ texts.

2. Statistical analysis of sentence length across the subcorpora indicates that there is no simple or consistent difference in sentence length between EMN articles and EMI 2 (Anglophone scholars) articles. Sentence length was found to be longer in the Portuguese EMN articles; shorter in Hungarian EMN articles; and there were no significant differences in sentence length between EMI 2 (Anglophone writers) articles and the EMN Spanish and Slovak articles.

3. There are indications that negative comments about sentence length by reviewers, and concern about sentence length by non-Anglophone writers, points to a more complex unease or dissatisfaction with linguistic and rhetorical features which are marked by reviewers as ‘non-native’ in some way.

In conclusion, style is a significant dimension to writers’ concerns about their English medium texts and reviewers’ response to texts. Whilst ‘sentence length’ is mentioned explicitly by both writers and reviewers as one of the key problem areas of style, it seems that ‘sentence length’ may be being used as a ‘catch-all’ way of labelling some more subtle workings of the text and responses to the text. In order to go some way towards capturing such subtlety, we have found the use of two notions, drawn from interactional sociolinguistics and linguistic ethnography particularly useful: indexicality and orientation. (See Nominated Output 2).

4.4 The categories EMN, EMT, EMI

The working hypothesis that framed this study was that there are three significantly distinct contexts of English medium journal publications: English medium national (EMN), for example an English medium Slovak journal which publishes primarily nationally generated papers; English medium transnational (EMT), for example an English medium Slovak or Spanish journal which publishes papers from a range of
national contexts; and English medium international (EMI), that is high impact factor journals based in Anglophone contexts, notably the US. The three categories – EMN, EMT and EMI – were are used to signal the geopolitical dimension of journal publishing contexts. A key goal was to explore the robustness of this distinction empirically and theoretically.

The study demonstrated that the category and the label EMN have a strong empirical referential basis from both an ethnographic and a corpus perspective. Thus all scholars identify national journals in English medium (and national languages) and make a clear distinction between what they are publishing nationally and internationally and why, often related to evaluation systems (see 4.1). EMN was therefore relatively straightforwardly operationalised for the corpus building as: English medium journals published in national contexts (Slovakia, Hungary, Spain and Portugal) by authors affiliated to non-Anglophone institutions located in these contexts. In contrast, the EMI category is a more fluid and complex label and category in that emic accounts by scholars indicate that ‘international’ is used with a range of meanings relating to publication: in another country; in the US; in another language; in a ‘high status’ journal (variously defined); in one of the high status indexes, notably the SSCI. In order to build a corpus of EMI articles, we recognised that a (small) corpus could not reflect the full complexity signalled by emic accounts but that a corpus could be constructed which connected with these accounts and therefore have strong face validity; that is, ‘international’ as indexing significant prestige or ‘impact factor’, and Anglophone centre contexts (notably the US). The EMI category was therefore operationalised in 2 subcorpora: EMI 1 English Medium journal articles listed in the SSCI written by authors affiliated to non-Anglophone context institutions (Slovakia, Hungary, Spain and Portugal); EMI 2 English Medium journal articles listed in the SSCI written by authors affiliated to Anglophone centre institutions (UK, US).

The study demonstrated that the third hypothesised category, English medium transnational (EMT) has a strong face validity from an emic perspective. Although unlike ‘national’ and ‘international’, ‘transnational’ is not a label normally used by scholars, the category itself is evident in references made to articles which fall some way between EMN and EMI in terms of geolinguistic description and status; an example of an EMT publication is an article written by a Spanish scholar and published in an English medium Greek journal which does not have high prestige or is not indexed in the SSCI. We decided that whilst this category was robust from an emic perspective, because of the strongly relational rather than referential dimension it was not possible to operationalise this category for the purposes of corpus building within the scope of the present study. As outlined in 2, for the purposes of the corpus aspect of the study, we therefore replaced EMT with a second EMI category, EMI 2, which were English Medium journal articles listed in the SSCI written by authors affiliated to Anglophone centre institutions (UK, US).

Empirically all three categories have a strong basis although EMN and EMI categories can more easily be used referentially than EMT and therefore operationalised for corpus building. Epistemologically all three categories have been used for analytic purposes, albeit with different emphases in datasets 1 and 2; in dataset 2, EMN and EMI have been used only referentially. In dataset 1 (Text Histories) the three categories are used both referentially and heuristically, that is as
both categorising journals in particular ways at the same time as raising questions about those same categories. This heuristic approach to the categories is important because it enables us to reflect something of the dynamism in journal status which might otherwise be made invisible. Thus for example, for the purposes of dataset 2, we categorised the Slovak English medium journal *Human Affairs* as EMN. But for the purposes of dataset 1 this journal is described as EMN/EMT/EMI in recognition of the journal’s goal (demonstrated in ethnographic accounts) and its gradual shift towards EMI status (see Annex 2). We see this heuristic/referential dimension as a useful tension in combining text, ethnographic and corpus dimensions which we plan to report in a future publication (see 6).

4.5 Combining text, ethnographic and corpus methodologies

A methodological question that the project set out to address was the value of combining text, corpus and ethnographic approaches. Ethnography and corpus are often posited as distinct approaches in tension (see discussions of advantages and limitations to each approach in Hunston 2002; and Stubbs 2002). In broad terms, we view ethnography as the overarching epistemology which involves two key underpinning positions: that all human actions need to be understood as situated and contextualized; that any research involves a commitment to comprehensiveness and complexity (after Blommaert 2008, p.13; see also Nominated Output 2). The building and use of electronic corpora is viewed as one of the many methodologies available to us to explore the phenomenon in question – the impact of the growing dominance of English as a global medium of academic writing on the lives and practices of multilingual scholars as well as on knowledge making practices more generally. Within this broad frame of reference, specific objects of study emerge through ethnographic immersion, that is through sustained contact with participants in order to explore what is significant to them and their goals of writing for publication. Such objects or issues can then be explored through a range of methodologies, including corpus analysis which, to date, is illustrated in our study of citations and one specific aspect of style, sentence length (see Output 1; see 4.2, 4.3 and 6). Within this framing of ethnography as the overarching epistemology, rich detail from text histories and ethnographic accounts can be juxtaposed with patterns across larger data sets in order to reach conclusions about a specific phenomenon. Furthermore, the use of corpus analysis as well as ethnographic case study allows for a productive bi-directionality between different types of datasets; thus findings and issues arising from text histories can be explored in corpus data which in turn throws up further questions to raise in the analysis of text histories.

5. Activities

- Participation in conferences, listed in *ESRC Society Today* form.
- Invited plenary and public talks listed in ‘Outputs’ and *ESRC Society Today* form.
- Principal investigator consultancy on mentoring scholars towards publication with international journal, *Compare* listed in ‘Impacts’.
- Invitation to present paper and contribute to two panel discussions at professional conference for translators and proofreaders, *MET*, Croatia 2008 listed in ‘Impacts’.
6. **Outputs**

- **Article:** Lillis, T., Hewings, A. and Vladimirou, D. (Submitted for review). The geolinguistics of English as an academic lingua franca: Citation practices across English medium national and English medium international journals. *International Journal of Applied Linguistics*.
- **Article:** Lillis, T., Hewings, A. and Vladimirou, D. (In preparation) Combining ethnographic and corpus approaches in the study of academic writing. For the *Journal of English for Academic Purposes*.
- **Article:** Hewings, A. and Lillis, T. (In preparation) What’s the problem with long sentences? A corpus based analysis of sentence length in English medium national and English medium international psychology journals. For the *Journal of English for Specific Purposes*.
- **Web site:** [http://creet.open.ac.uk/projects/paw](http://creet.open.ac.uk/projects/paw) (Annex 9)
- **Leaflet:** (Annex 8)

7. **Impacts**

1. A programme of mentoring with an international journal. Drawing on findings from the study Theresa Lillis has worked as a consultant with one international journal (*Compare: a journal of comparative education*) to develop a mentoring programme aimed at supporting periphery scholars, through initial drafting towards publication. Plans are underway to produce in conjunction with BAICE/Compare [http://www.baice.ac.uk](http://www.baice.ac.uk) a booklet with full details of the programme in the hope that other editorial boards may consider offering similar programmes.

2. The president of the professional organization MET, *Mediterranean Editors and Translators National Conference* contacted the PI to discuss the study and
to invite the PI to present a paper and take part in two panel discussions outlining the relevance of the study to professional translators and authors’ editors. Mediterranean Editors and Translators National Conference, 2008
http://www.metmeetings.org


8. Future Research Priorities

8.1 The data generated from the research is considerable and the following areas of analysis have still to be fully exploited: 1) Analysis of the rhetorical functions of explicitly naming the national context of research through lexical items relating to or cohering with nationality, national context, such as Hungarian, UK, US through concordance searches and qualitative text history analysis (see Annex 5); 2) Qualitative analysis to explore in detail the publishing practices of scholars for at least three contexts- EMN, EMT and EMI; 3) Analysis of datasets 1 and 2 to explore preliminary findings which suggest that in EMN journals, psychology is construed as a broad inter- and trans-disciplinary epistemological enterprise, as compared with the much narrower and intradisciplinary approach adopted in EMI journals.

8.2 Further relevant research involving additional data collection but building on findings from this study are as follows: 1) Historical analysis of the linguistic medium of a sample of psychology journals from different national contexts. Whilst not the focus of this study there is evidence of a move from national languages to bilingual/English dominant journals. The investigation would centre on identifying changing rhetorical practices over a 20 year period thus tracking the changing types of knowledges that circulate over time. 2) Investigation into the circulation and readership of EMN journals given that they seem to occupy a fragile position and may potentially be failing to reach both national and international communities.

9. Ethics and confidentiality

Research was conducted in accordance with the ethical codes of practice of the British Association for Applied Linguistics, the British Educational Research Association, EU guidelines and the Open University Code of Good Practice in Research:

http://www.baal.org.uk/about_goodpractice_full.pdf
http://www.bera.ac.uk/publications-guides.php
http://ec.europa.eu/research/science-society/page_en.cfm?id=3198

Careful attention has been paid throughout the research process and reporting of findings to ensure the anonymity of participants as in previous research.
Permission was sought and secured from journals to include articles in the electronic corpus for research purposes. One publisher gave permission for a three year period of use and asked that a further request for renewal be made at the end of that period.

Annexes 2 and 4 are considered strictly confidential and we request that they should not be made publically available. Although names have been anonymised, some information is included (such as details of specialist subfield and the journal names in which articles have been published) which could potentially threaten the anonymity of scholars’ identity, given the small number of people involved in some specialist fields of inquiry in each country.
Final Report  List of Annexes

Annex 1  References
Annex 2  Confidential  Text Histories overview
Annex 3  The TAKPsych corpus
Annex 4  Confidential  Ethnographic data
Annex 5  Hungarian concordance searches example
Annex 6  Citation concordance searches example
Annex 7  Countword
Annex 8  Leaflet
Annex 9  Project Web Pages