



## **ANALYZING PRELIMINARY INSIGHTS ON AUTHORS' FIRST LANGUAGE AND ENGLISH LANGUAGE PAPERS**

**<sup>1</sup>DHARMARAJULA NAGARANI,<sup>2</sup>VELPULA VENKATESWARLU,<sup>3</sup>B. DIVYA VANI**

**<sup>1,2</sup>Assistant Professor**

**Department of H&S Engineering**

**Abdul Kalam Institute of Technological Sciences, Kothagudem, Telangana**

Page | 776

### **Abstract:**

The paper argues for greater language sensitivity in scientometric research, in addition to discussing Baldauf and Jenrudd, 1, – ~3, and raising issues of North/South and English/non-English imbalance in research communication. The procedure for identifying authors in monthly publications on textual evidence as native or non-native English speakers is then suggested. A preliminary application review and report are given for 623 articles. The NNS rate in Economics was 23%, which was half that of Health Science. Additionally, the dearth of documents from the Third World served to support Baldauf and Jernudd. Scientometries are suggested as a potential supplement to Research English training in order to address purported imbalances..

**DOI Number:10.4847/NQ21291**

**NeroQuantology 2021;19(11);Page:776-**

**784**

### **1. Introduction**

The literature in the overlapping fields of Scientometrics, Research Communication Studies and the Sociology of Science would appear to have paid little attention to the linguistic dimension in at least two important ways. In one way, these fields may have been disadvantaged by lack of contact with research in discourse analysis and in the applied field of English for Specific Purposes ESP (the provision of specialised English language programmes for various professional groups, particularly science students and researchers). In the other, there has been some neglect of the language of publication as an important variable in scientific communication. Discourse analysis and ESP have potentially useful insights and techniques to contribute in such areas as the rhetorical organisation of scientific text, 1 i 2 the elucidation of meaning in scientific papers,3, 4 and in pertinent specific areas such as the textual patterning of citations, s,6 And indeed there is some sign that the value of discursal approaches is gaining recognition, particularly because of its focus on "describing how scientists' accounts are organised to portray their actions and beliefs in contextually appropriate ways". 7 The relationship between the language chosen for publication or presentation and the writer or speaker's proficiency in that language and the further relationship between language choice and its visibility, audience-size and prestige are equally deserving of more consideration. For instance, these relationships are nowhere referred to Bazerman's extensive and recent 142-citation review of the literature despite the fact that his section headings include topics such as The Writing Process, Textual Form, the Dissemination Process and Audience Response. 8 Blickenstaff and Moravcsik,9 in their





investigation of the professional profile of participants in an international meeting, did not include a questionnaire item about the language proficiencies and preferences of those participants. Schubert et al.<sup>10</sup> have recently analysed the proceedings of more than 500 international scientific meetings and concluded: "the distribution of the participants of international scientific meetings depends of (sic) the geographical location of the host country, and in addition, the similarity of efforts for scientific development (e.g. in the developing countries) the organisational structure ('open' or 'closed' nature) of the scientific communities, the economic situation (travelling expenses can influence the participation rate) and in some cases, political consideration may also have an important role." A sociolinguistic orientation would certainly wish to raise the issue of a possible relationship between the language or languages formally accepted for particular international meetings and the distribution of participants both as auditors and speakers. However, two valuable recent papers by BaMauf and Jernudd<sup>11,12</sup> do take up the question of language use patterns in the periodical literature and thus their findings need to be discussed in a little more detail. Having established that English was the dominant language in the subject literatures they had scanned, Baldauf and Jernudd (henceforth B & J) then examined language use patterns in 884 abstracts from the fisheries literature. Because of the wider implications of their conclusions they are iterated here: "First, over eighty per cent of the articles analysed form what might be called the national literature tradition where authors write in the national or official language of the country in which they reside... The second pattern involved the use of English as a publication medium in countries, mainly in Europe, where English is a foreign language. About sixty five percent of these publications were, published in international forums, leaving about fifty articles in which a potential language-national mismatch existed. Thus English language dominance appears to be the result of the large number of scientists doing research in English-speaking countries, English national language countries, or for international organisations or forums. Only a small proportion of authors appeared to be potentially involved in a situation where language selection was a deliberate variable in the communication process. ~ 3.

If these findings were to be verified they are decidedly uncomfortable. They would seem to imply that research is the preserve of countries where English is either the national language or the official language, of countries with an international language of scholarship such as Japan and the major nations on the continent of Europe, or of those individuals who go to international meetings. B & J's figure of only 40 English-national mismatch papers out of 884 indicates, for example, very low levels of research in the so-called lesser developed countries of the Arab World, Latin America, South-East Asia and perhaps Francophone Africa.<sup>14</sup> Such a conclusion is *prima facie* unlikely, given the almost universal consensus that publication in respectable refereed journals should be an important criterion for promotion in academic and research fields. An alternative hypothesis would therefore be that research in non-anglophone LDCs is indeed being done, but little of it is finding its way into the journals that come to the attention of major (and highly efficient) abstracting services. Either way, we seem to be faced with a further instance of North - South imbalance presumably due to an interlocking multiplicity of factors among which we could propose additional language hurdles facing non-native speakers of English (NNS), editorial gatekeeping bias under influence of the "Matthew effect", as and a generally less supportive research environment. And either way,





we are faced with serious questions about the effectiveness of the massive investment by hard-pressed LDCs in doctoral scholarships held by their nationals in the USA and Europe, and about the long-term scientific and developmental value of the large number of research scholarships offered to LDC nationals by American, British, Russian and European governments and other agencies. In view of these wider implications it would seem appropriate to examine the methods used by B & J, in particular to ascertain whether or not their mismatch coding may have underestimated the proportion of papers written by NNS of English and/or papers written in non-anglophone environments. The first point to note is that their data was in the form of abstracts (all in English) from ASFA. The relevant linguistic variables provided in AFSA are (1) the original language of the abstracted article, (2) the country for the original abstract, the address of the first author, (3) language(s) used for the original abstract, and less certainly, (4) number of authors, (5) location of study. B & J compared in particular (1) and (2) i.e. first author-residence matching and mismatching, and found a strong relationship between language and residence of first author, to the extent that there were few instances of potentially linguistic mismatches from the expected pattern of communication. They observe that "further specification of these discrepancies would require surveying individual authors to ascertain their motivations for publishing in non-native language publications")<sup>6</sup> It is here suggested that abstracts provide rather less textual evidence than the original articles as to whether the author(s) of articles written in English are native or quasi-native users of that language or non-native speakers. It is therefore the first aim of this paper to propose a provisional set of criteria for identifying the NS/NNS status of authors based on the evidence discoverable in their publications. If such a procedure could be validated it would operate as a useful intermediate strategy between the rapid but possibly unreliable scanning of abstracts and the laborious but certainly revealing method of communicating directly with individual authors. The second aim is to apply the procedure to a small sample of the periodical literature and to compare, insofar as this is possible, the findings reached with those found by B & J using their abstract-analysis technique. The final purpose is to indicate potentially fruitful areas of collaboration between Scientometrics and English for Specific Purposes.

## 2. Procedures and results

Towards the identification of NS/NNS status

Cursory examination of articles showed that there were pieces of evidence frequently present but of limited use and other pieces of evidence that might occur only in a small percentage of papers but which were strongly indicative of status. Of the first kind is the last name of the author, which is always given but is of limited identificatory value, particularly of course, in the USA where ethnic origins are so diverse. Whilst it is true that an Anglo-Saxon last name will, with some degree of probability, indicate an English NS, the reverse is certainly not true (although anglicizations of non-English names can be useful i.e. Hartman for Hartmann). On the other hand, a footnote explaining that author has had his paper translated into English or has relied upon an NS for correction is strong evidence of NNS status, but occurs infrequently. After some experimentation, we have relied on the following six categories of potential evidence and given each one a reliability rating of 1 to 3 (R, RR and RRR), the six evidence-types being listed in order of decreasing frequency of occurrence:





1. R (one point) Last name (including Anglicization etc.),
2. RRR (3 points) Institutional affiliation (assumed to be permanent unless otherwise indicated),
3. R (1 point) All citations to English language publications,
4. RR (2 points) First name of author (John, Jean, Juan, Giovanni, Johann, etc.; surprisingly useful evidence when available, especially useful for identifying the language status of married women),
5. RR (2 points) All self citations to English language publications,
6. RRR (3 points) NS/NSS Status-relevant footnotes or endnotes (Translation; acknowledgements for linguistic help; visiting affiliation and being on leave away from permanent base; we also think that source of funding falls into this category especially if the grant is being provided by an agency such as the Ministry of Defence). Two constructed cases follow to show the procedure was applied:

<i>Case I</i>	<i>Score</i>	<i>Case II</i>	<i>Score</i>
1. Jones	+ 1	Ibrahim	- 1
2. Cambridge U.	+ 3	U. of Jordan	- 3
3. All citations in English	+ 1	Some citations in Arabic	- 1
4. George	+ 2	Mohammed	- 2
5. All self-citations in English	+ 2	All self-citations in Arabic	- 2
6. Grant from UK Ministry of Defence	+ 3	Acknowledgement of English help	- 3
Total	+12		-12

In reality, 12 scores are quite rare either because of the absence of information, particularly in categories 4 - 6, or because of conflicting and hence subtractive evidence. Nevertheless, plus or minus scores of 8 or 9 are common and 5 - 7 scores the most frequent. In the end, we adopted a five rank system on a somewhat arbitrary basis; further research via author contact may indicate that the characterizations of the scoring bands may need some adjustment. The system is as follows:

- A. +5 to +12 - native speaker of English,
- B. +2 to + 4 - probably NS,
- C. +1 to - 1 - uncertain status,
- D. -2 to - 4 - probably NNS,
- E. -5 to - 12 - NNS.

In the common case of multiple authorship each author was scored and the average taken for the paper, part numbers being ignored. This procedure is undoubtedly crude and later refinements should be possible.

#### Pilot application of the procedure

The stock of current unbound journals in the University of Aston Library was examined in two randomly selected areas: part of Health Sciences (Dewey 610-615) and part of Economics





(Dewey 330). However, in the two areas, only journals of international aspiration, with a provenance from Britain or the USA and with a monolingual English language editorial policy were selected for scanning. In the pilot stage, it was thought that this decision would obviate the complexities that might arise from English language publications emanating from countries where English was not the national or official language. It turned out that late '83 and early '84 copies of twenty journals were sampled, fourteen from the Health Sciences and six for

Economics. The total number of papers scored was 623 (434 in D610/615 and 189 in D330), giving an average for each journal of a little over 30 articles. The results for the Health Sciences are given in Table 1; the Roman numerals refer to journal titles which are listed in the Appendix. The results show that about 72% of the authors appeared to be native speakers of

English, about 23% non-native speakers and no clear evidence either way in about 5% of the sample. They also show wide variation in percentages between journals (NNS authors from 4% to 40%) but each journal sample is so small that this finding cannot

Table 1  
 NS/NNS status in a sample of current Health Science journals;  
 figures refer to numbers in each category

Journal	Provenance	A(NS)	B(? NS)	C(? )	D(? NNS)	E(NNS)	Total
I	GB	15	2	0	1	1	19
II	US	23	8	1	4	7	43
III	US	23	1	0	3	5	32
IV	GB	11	1	3	3	6	24
V	GB	17	1	0	0	3	21
VI	US	22	1	1	0	1	25
VII	GB	24	3	2	2	15	46
VIII	US	4	7	0	1	2	14
IX	US	22	1	1	3	5	32
X	US	24	1	4	11	8	48
XI	US	16	9	2	6	1	34
XII	GB	20	4	0	3	3	30
XIII	US	27	4	1	1	2	35
XIV	US	16	6	4	0	5	31
Total		264	49	19	38	64	434
%		60.8	11.3	4.6	8.7	14.5	







**Table 2**  
**NS/NNS status in a sample of current Economic journals**

Journal	Provenance	A	B	C	D	E	Total
XV	US	53	3	0	0	2	58
XVI	US	9	3	1	0	3	16
XVII	US	24	3	1	0	1	29
XVIII	GB	20	2	0	0	1	23
XIX	GB	20	3	3	2	4	32
XX	US	24	2	1	2	2	31
<b>Total</b>		<b>150</b>	<b>16</b>	<b>6</b>	<b>4</b>	<b>13</b>	<b>189</b>
<b>%</b>		<b>79.3</b>	<b>8.5</b>	<b>3.2</b>	<b>2.1</b>	<b>6.9</b>	

**Table 3**  
**A comparison of NS/NSS percentage in British and US journals**

Journal	A&B	C	D&E
Total Health Science	72.1	4.6	23.2
US Health Science	73.1	4.8	22.1
GB Health Science	70.0	3.6	26.4
Total Economics	87.8	3.2	9.0
US Economics	90.3	2.2	7.5
GB Economics	81.8	5.4	12.7

be considered other than marginally indicative of a large range of interjournal NS/NNS authorship patterns. The results for the Economics journal are strikingly different, as can be seen from Table 2. The Economics data shows a much increased tendency towards domination of the small sample of the journal literature by native speakers of English (88% as opposed to 72%). This is not the place to speculate about the reasons for this disparity; rather, it would be more useful to consider three other issues: the possibility of different NS/NNS ratios in British and American journals, possible differences between single or multiple-authored papers, and the location of putative NNS authors. The general trend on the first issue is self-evident, as can be seen from Table 3.

In both cases there is a relatively small shift to increased NS percentages in American journals, a movement that would be fairly predictable in view of the much larger native American academic constituency. Secondly, the data reveals a light trend towards rather more NNS writers in coauthored papers than in single-authored ones. In the Economics journals (where there was a small majority of single authored papers) 10.5% of the co-written papers were by NNS whereas NNS wrote only 8.6% of the individual papers. Similarly in the Health Sciences (where only 13% of the papers were single-authored) NNS multiple-authored papers amounted to 24.1% of the sample and NNS single-authored papers 17.6%. The final issue to be reported on relates to the location of authors categorised by the scoring procedure as being non-native speakers of English or probably non-native speakers of English. Of the 117





locations traced, 42 were in Western Europe, 22 in the Nordic countries, 12 in Japan, and 11 in Eastern Europe or Russia, and 9 in North America. There were 21 Third World locations, but ten of these were from institutions in the Indian sub-continent with its strong tradition of using English as the language of scholarship, and five from Israel, where the data is particularly suspect because of the large amount of US - Israeli academic traffic. Thus, there remain but five papers of clear NNS Third World provenance, four from Latin America and one from Korea.

Evaluation of the procedure The reliability of the procedure was assessed in the following way. The two research assistants selected the same 100 articles from the ESP literature and scored them independently. These were then coded into the A-E scale and an inter-scorer reliability of 90% was found - in only one case (A/C) did the discrepancies go beyond adjacent categories. The present writer (who has long familiarity with the ESP field) then examined the list of authors of the 100 pre-selected papers to see whether he knew their NS/NNS status. In 75 cases he identified native speakers, in 20 instances non-native speakers and in five cases he did not know. This personal knowledge was then correlated with the results from the text-scrutiny procedure, as shown in Table 4 (the double letters referring to the two scorers). The results look encouraging, although the procedure will certainly place some NS with permanent residence in non-anglophone countries in the C category rather than in A or B.

**Table 4**  
 Reliability of scoring procedure in cases of known NS/NNS status

Scores	NS	NNS
AA	58	0
AB	5	0
(AC)	1	0
BB	4	0
BC	2	0
CC	4	0
CD	1	0
DD	0	4
DE	0	1
EF	0	15
<b>Total</b>	<b>75</b>	<b>20</b>

### 3. Discussion and Conclusion

No direct comparison between the results presented here and those found in Baldauf and Jernaud's Refs. 11–13 is intended.owing in part to the fact that the samples and their classifications differ, but also because the two research' objectives—while connected—differ significantly in a number of key areas. B & J were primarily interested in language selection—international, regional, or national—while disseminating research findings. This paper, on the other hand, has focused on the use of English for research communication and has made an effort to determine the proportion of papers published by non-native speakers within that linguistic framework. Second, B & J calculated the proportion of accepted articles that were written by authors from non-English speaking nations as opposed to NNS authors.





In this sample of 144 publications, they discovered that 509 publications in the English language might be assigned to non-English speaking nations. Compared to the overall average of 19% for NNS writers in the current sample, this 28% rate is much higher. Nevertheless, B & J's percentage drops to little under 8% if we adhere to both B & J and the current study design and exclude publications for international organizations or in English-language magazines published outside of the English-speaking globe, which is significantly less than the 19% stated above. To determine the extent to which this discrepancy can be attributed to small sample sizes, sampling from other study regions, variations in the method used for the investigation, or minor variations in the objective of the investigation, more research would be required. However, the current study does provide significant indirect support for B & J's conclusion that relatively few Third World papers—whether or whether they are produced by native English speakers—are published in US and British journals in the regions they investigated. Finally, this final section revisits some of the more general implications for linguistic scientometric research that were discussed in the Introduction. There was undoubtedly mention of the disheartening lack of attention that scientists studying communication in science have shown to the linguistic issue. The seeming lack of concern is not disheartening per se, but rather concerning the chance that scientometrics has missed to truly and practically advance the study of English and other Languages for Specific Purposes. English for international conferences, cross-cultural scientific rhetorics, assisting NNS in reading scientific publications, and, if requested, assisting NNS in writing papers and abstracts in English are all areas of great importance to ESP. Specialist research support in language use patterns, editorial gatekeeping policies regarding NNS and Third World submissions, surveys of publication careers of researchers returning to LDCs after advanced training in Western or Eastern countries, and studies of international conferences that take sociolinguistic parameters into account would all be extremely beneficial to the ESP profession. Data on these subjects could therefore be used in two ways: first, as input for designing courses and syllabuses; second, as statistical data that can be included to reports and project proposals, and utilized in discussions with funding agencies and institutional authorities.

## References

1. E. TARONE, S. DWYER, S. GILLETTE, V. ICKE, On the use of the passive in two astrophysics journal papers. *The ESPJournal*, 1 (1981) 123.
2. H. WIDDOWSON, *Explorations in Applied Linguistics*, OUP, London, 1979, p. 51-61 and passim.
3. L. SELINKER, On the use of informants in discourse analysis and Language for Specialised Purposes, *IRAL*, 27 (1979) 189.
4. T. HUCKIN, L. OLSEN, On the use of informants in LSP discourse analysis, *Reading for Professional Purposes*, Heinemann, London, 1984, p. 120-129.
5. J. SWALES, *Aspects of Article Introductions*, LSU Research Monograph, Birmingham, 1981.







6. J. SWALES, Developing materials for writing scholarly introductions, in: Case Studies and ELT, R. JORDAN (Ed.), Collins, London, 1983, p. 188-200.
7. G. NIGEL GILBERT, M. MULKAY, Opening Pandora's Box, CUP, Cambridge, 1984, p. 14.
8. C. BAZERMAN, Scientific writing as a social act: A review of the literature of the sociology of Science, in: New Essays in Technical and Scientific Communication: Research, Theory and Practice, P. V. ANDERSON et al. (Eds), Baywood, Farmingdale, N. Y., 1983, p. 156-184.
9. J. BLICKENSTAFF, M. J. MORAVCSIK, The profile of an international meeting, *Scientometrics*, 5 (1983) 143.
10. A. SCHUBERT, S. ZSINDELY, T. BRAUN, Seientometric analysis of attendance at international scientific meetings, *Scientometrics*, 5 (1983) 177.
11. R. B. BALDAUF, JR, B. H. JERNUDD, Language use patterns in the fisheries periodical literature, *Scientometrics*, 5 (1983) 245.
12. R. B. BALDAUF, JR, B. H. JERNUDD, Language of publications as a variable in scientific communications, *Australian Review of Applied Linguistics*, 6 (1983) 97.
13. R. B. BALDAUF, JR, B. H. JERNUDD, Language of publications as a variable in scientific communications, *Australian Review of Applied Linguistics*, 6 (1983) 105.
14. c. f. C. H. DAVIS, Institutional sectors of 'mainstream' science production in Subsahara Africa, 1970-1979: A quantitative analysis, *Scientometrics*, 5 (1983) i63.
15. R. MERTON, The Matthew effect in science, *Science*, 159 (1968) 56.
16. c. f. Ref. 11 p. 254.

