6 *The role of English in research*

6.1 General perspectives

This enormous RA production raises two inter-related questions that need to be addressed in a book of this purpose and scope: first, how much of this production is in English as opposed to other languages and, second, how is it globally distributed? Then in the following section (6.2) we will consider what the answers to such questions might mean for researchers who do not command native or near-native skill in English and/or who are located in 'off network' parts of the earth.

Until recently such language issues have been surprisingly neglected in the relevant literatures. As Baldauf and Jernudd (1983a) comment:

> Although language of publication is an inescapable feature of scientific communication, it is most often treated as background noise, a variable in which neither information specialists nor scientists have shown much interest, nor is it, as far as we can tell, a problem which linguists have examined.
>
> (Baldauf and Jernudd, 1983a:97)

In an effort to provide some redress, they re-examined Wood's 1967 study of the language used in original articles abstracted by the major science abstracting and indexing services in 1965 and, as far as possible, attempted to replicate the study using 1981 data. (The 1965 and 1981 figures are not exactly comparable because the abstracting services have both modified their names and their procedures as a consequence of increasing computerization.) The advance made by English is both striking and consistent. A simplified version of their findings is given in Table 1 overleaf.

Baldauf and Jernudd were able to show that in their data the gain in English has been largely achieved at the expense of the major European languages. Japanese and Chinese both increased in percentage terms, although the latter from a minute 1965 base line. Maher (1986b), in a detailed study of the medical literature, largely reaches the same conclusion except that he notes a small decline in the Japanese percentages beginning sometime between 1970 and 1975. Even higher percentages

TABLE 1. PERCENTAGES (TO 1%) OF ABSTRACTED ENGLISH LANGUAGE RAS

	1965	1981*	Gain
Chemistry	50	67	17%
Biology	75	86	11%
Physics	73	85	12%
Medicine	51	73	22%
Maths	55	69	14%

for English have been found in more specialized sectors. Baldauf (1986), in a study of the four leading journals devoted to cross-cultural psychology published between 1978 and 1982, found an English-medium publication percentage of 97%.

The upshot of all these figures would seem to suggest that the anglophone grip on published research communications is both strong and tightening. However, this English domination may in fact not be as real as it appears; indeed, typical claims that 80% of the world's scientific production is written in English (e.g. Garfield, 1983) may be decidedly on the high side. The major difficulty is one of bias in the data bases. The Institute of Scientific Information's (ISI) largest bibliometric tool is the *Science Citation Index* (SCI), which scans about 4,000 journals. While, with an estimated world total of 70,000 science journals, considerable selectivity is clearly necessary, there is evidence (as there is in other US data bases) of a tendency to select journals emanating from the most developed countries in the northern hemisphere and publishing in English. These dual tendencies in effect constitute a kind of 'double-whammy' for non-anglophone Third World research. Indeed, the ISI itself sponsored an investigation into The Coverage of Third World Science and concluded that it may be under-representing worthwhile research from lesser developed countries (LDCs) by a factor of two (Moravcsik, 1985). A particularly striking case is that of Brazil which, according to Arvanitis and Chatelin (1988) publishes 149 scientific periodicals, only four of which are currently included in the ISI data base. Najjar (1988) claims that not even one Arab world science journal was consistently entered in the data base in the 1980s.

Nor are 'peripheral' English-speaking countries necessarily immune from deselection. Byrne (1983) in an article trenchantly entitled 'How to lose a nation's literature: database coverage of Australian research' showed that (apart from education) at most only a third of Australian journals in various social sciences and humanities were represented in international data bases.

There have been a number of attempts to get round the problem of

bias. Throgmartin (1980) used UNESCO's *International Bibliography of the Social Sciences*, fearing that an American (and hence largely anglophone) abstracting service would under-represent non-English items. His relatively small (1000–1100) samples from sociology, economics, political science and anthropology show English percentages of between 39% and 51% with consistent double-figure percentages (14–26%) for French. Analogously, Baldauf and Jernudd (1983b:247) analyzed in detail a sample from *Aquatic Sciences and Fisheries Abstracts* 'because of its international sponsorship (i.e. UNESCO and based in Zurich) and because it appeared to abstract articles on fisheries in a wide variety of languages and from a wide variety of sources'. Here, however, English continued to be by far the most important language of publication (75%) with French (5.5%) and Spanish (4%) coming a long way behind.

Arvanitis and Chatelin (1988), in an impressive recent study of tropical agriculture, used PASCAL, a large French multidisciplinary data base, in another attempt to get round the problem of English-language bias. However, like Baldauf and Jernudd (1983b) they found even here that three-quarters of the publications were in English, with French (10%), Portuguese (7%) and Spanish (5%) falling far behind.

A rather different kind of evidence comes from the meticulous bibliography of the literature on schistosomiasis, (Warren and Newill, 1967). The bibliography covers the period from 1852 to 1962, contains nearly 10,000 articles on the subject and represents a truly global search, the most significant figure being that these articles have been garnered from no less than 1738 different periodicals including those publishing in languages as peripheral as Albanian. The Warren–Newill bibliography thus provides a fully comprehensive listing without any bias either in terms of RA language or in terms of provenance. Not surprisingly, the language distribution of RAs is a little different to that which we have

TABLE 2. LANGUAGE DISTRIBUTION OF SCHISTOSOMIASIS RAS 1957–62

Language	Total	%
English	750	45.5
Portuguese	230	15
Chinese	200	12.5
Japanese	160	10
French	140	9
Spanish	40	2.5
German	30	2
Italian	20	1
Other	40	2.5
Total	1610	

come to expect. Table 2 is an approximate adaptation of part of the graphs in Warren and Goffman (1978: 34–5).

Undoubtedly, certain percentages have been skewed by the specialized topic – as they are likely to be in any specialized bibliography. For instance, the high percentages for Portuguese and Chinese are almost certainly due to the fact that schistosomiasis is endemic in parts of both Brazil and China. It may be further objected that the reason for the lower percentage of English-language articles is simply caused by a relative lack of interest in the disease in the United States, but even here American concern with tropical diseases is known to be strong in both the missionary and military fields. While the Warren–Newill bibliography is now 25 years old, it still represents the most complete bibliography that I know on any research topic. It thus represents a useful partial indicator of the role of English in an area of very broad and long-standing international concern.

We are now in a position to attempt an answer to the first question regarding the importance of English. While there is no doubt that English has become the world's predominant language of research and scholarship, the extent of that predominance may have been exaggerated. High overall percentages can indeed be extracted from the major US-located international data bases, but these data bases are themselves predisposed to English-language sources. Arvanitis and Chatelin, writing from Venezuela, draw the consequence with considerable irony: if one relies on the SCI data base, they argue 'one cannot be surprised to learn that the US produces 40% of the international production and receives 60% of the citations, or that 80% of the world scientific production is written in English' (Arvanitis and Chatelin, 1988:114). An overall figure of 80% is almost certainly too high.

More important than any overall figure is the small but growing body of evidence that the role of English is variable. We have already seen this in very high percentages for cross-cultural psychology and in much lower ones for schistosomiasis research. There may be special language-preservation efforts as in French-language sciences, or in German-language psychology (Becker, 1984). Najjar (1989), in a study of journals published in the Arab world, found Arabic predominantly in social science RAs, occurring about 25% of the time in agriculture but hardly at all in medicine. Finally, Jernudd and Baldauf (1987) contrast the strong effect of English on Scandinavian psychology with its very weak effect on articles concerning the teaching of Swedish. In the latter case, the medium of scholarly communication has predominantly been a Scandinavian language for the last 40 years (overall about 93%). In more general terms, it may be hypothesized that research fields relying on localized input (archaeology, agriculture, literature, religious studies) are more likely to resist or escape the domination of English than those that do not

(chemistry, genetics, physics etc.). A useful study might be made, for example, comparing language use patterns of research into a spottily-distributed rural disease of great antiquity like schistosomiasis and a very recent and increasingly-global urban disease like AIDS.

The final aspect of the first answer pertains to the future. At present all the available studies indicate that the predominance of English is currently growing (e.g. Table 1). However, as Maher (1986b:208) observes, 'language is maintained or declines in response to the amount of (new) information that it carries'. A time will surely come in the future, as it always has in the past, when one premier language for the exchange of knowledge and information will give way to another. Indeed a presage of this may be detected in the role of Japanese as the language of the fifth-generation computer (Grabe, 1988b).

The new field of *linguistic scientometrics* (Baldauf and Jernudd, 1986) might also be expected to provide answers to the second group of questions: where are (English-language) RAs coming from, and how many are being produced by non-native speakers of English? These are harder questions to answer than the first set; in consequence, the answers must be more partial and more tentative.

Baldauf and Jernudd (1983b) in their fisheries study, estimated, on the basis of the institutional affiliation of the first author, that four-fifths of the English-language articles originated in countries where English is either the national language or the official language. Two-thirds of the remaining 20% appeared as outcomes of international conventions or as the work of multinational organizations such as the International Whaling Commission. The remaining 6% were submitted from locations where English was a foreign language, but very few of these were outside Western Europe or Japan.

Swales (1985b) attempted to estimate the percentage of NNS or 'probably' NNS authors in a selection of RAs taken randomly from journals in health sciences and economics available in a British university library. Again, the overall percentage was around 20%. However, of the 117 locations traced for NNS or probable NNS writers, only 21 were in the Third World, and 'ten of these were from institutions in the Indian subcontinent 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' (Swales, 1985b:98).

The two studies reviewed above suggest that published research from peripheral countries (non-anglophone, LDC etc.) is minimal. Work on tropical agriculture by Arvanitis and Chatelin (1988), however, throws doubt on the general validity of such a conclusion. Their investigations into the 1983 PASCAL tropical agriculture data base suggest that perhaps up to half of the research in this field – one of course with a high

Third World priority – is being published from southern hemisphere locations, even when important anglophone contributions from Australia, New Zealand and South Africa are excluded. They further show that LDCs have evolved somewhat different national traditions for agricultural research publication. Of the Third World's agricultural research 'big three', Brazil tends towards national publication in the national language, Egypt towards national publication in English rather than Arabic, and India towards international publication in English.

Support for at least the Brazilian and Egyptian findings can be seen in more localized studies. Velho and Krige (1984) investigated the publishing activities of a group of Brazilian agricultural scientists. The group published 55 RAs in national Brazilian journals in 1982 (none on the ISI SCI list) and only seven abroad. This result is in marked contrast to other findings such as those of Rabkin and Inhaber (1979), who concluded that the Latin Americans published 74% of their work in 'advanced country' journals. However, Velho and Krige were working *within* the Brazilian context rather than relying on US data bases and almost certainly have more complete and more reliable information. Although Velho and Krige are not explicit about the matter, they certainly imply that the preference for domestic journals is related to the opportunity thus afforded to write in the native language, Portuguese, as well as to the opportunity of making a more direct appeal to the leaders of their agricultural community.

Najjar (1989) was able to corroborate the fact that Egyptian agricultural researchers have a strong tendency to publish locally but in English. In Jordan, on the other hand, the picture is more complex. At one of the two major agricultural research institutions, the University of Jordan, the trend remains to publish mainly locally, and largely – but by no means exclusively – in English. At the other institution, the University of Yarmouk, researchers tend to publish internationally. The explanation for the difference is in fact quite simple; the University of Jordan produces a suitable scholarly journal of its own, but Yarmouk does not.

The answer to the second set of questions is considerably less clear than it is to the first set of questions, partly because of the technical difficulties in getting at reliable data. However, both answers share a common danger. If we rely exclusively on the major US data bases we will fulfill the prophecy, in the first case, that English is overwhelmingly predominant and, in the second, that published research activity outside the northern hemisphere is minimal. However, when we go beyond such data we see that published research activity in peripheral contexts is not so much minimal, as largely invisible to the leading specialized discourse communities. Certainly, this seems to be the case with agricultural science in Brazil and in much of the Arab World.

Naturally enough, such conclusions raise in turn legitimate concerns

about bias and lack of equity. There already exist intractable imbalances between the north and the south in such areas as GNP, quality of health care, life expectancy, educational opportunity, level of literacy and so on. Is it inevitable that we have to add to this list disregard for valid Third World research and scholarship? And if not, what role could our own discourse communities play in making research activity around the world a more equitable field of endeavor? These issues are taken up in the next section.

6.2 Individual perspectives

One of the aims of *linguistic scientometrics* 'is to develop an understanding of how language acts as a barrier in various disciplines, and what language correction procedures are able to solve them' (Baldauf and Jernudd, 1986:388). Although there exist a few localized studies of how non-native speakers of English manage to survive in an increasingly English-dominated research world, such as St John's investigation into the behaviors of Spanish scientists at the University of Cordoba (1987), much work remains to be done. Some of the most dramatic findings are the self-reports elicited by Jernudd and Baldauf (1987) from Scandinavian psychologists. The findings are dramatic because Scandinavians are in English-language terms one of the more privileged NNS groups. The advanced Scandinavian countries have highly-developed and largely successful foreign-language educational components and their leaders have a keen sense of the need to foster polyglot citizenries in order to maintain connections with Western Europe and beyond. And yet for the psychologists, all of whom have published in English-language international journals, communication in English represents a heavy burden:

> – It is constantly depressing to be confronted by one's shortcomings in foreign language.
> – It is meaningless to publish original research in psychology in Swedish.
> – I regard the language barrier as a central problem for Norwegian researchers in my professional field.
> – One year in England/USA – even as a street sweeper – would likely mean more to a scientific career than half a million crowns in the form of a research grant.
> – It is important for those of us who are non-native speakers to create some understanding among many researchers that English is not their natural (or obvious) language of communication.
>
> (Jernudd and Baldauf, 1987:150)

There may be other burdens as well. All non-native English speakers must take additional time out of their academic and research careers if

they wish to acquire and maintain high-level English-language skills (Lewin and Jordan, 1981). Further, the general increase in research production is leading to high journal rejection rates, currently reaching 80–95% in the arts and humanities, which in turn means increasing pressure on manuscripts that betray evidence of non-standard English.

In fact, the Scandinavian psychologists are probably relatively well-placed in comparison to all those researchers in more isolated, more 'off-network' and generally less supportive research environments. In those situations, library resources may be limited, with the result that only the leading journals will be carried. One important consequence of this is that disadvantaged researchers may not even *know* of the existence of journals (particularly newish ones) which might in fact be eager to publish their contributions. Access to good English-language assistance may be limited, and advice on rhetorical differences in research writing in the mother tongue and English non-existent or even misleading. And finally, there is the question of bias against submission from 'obscure' places. In a somewhat notorious study, Peters and Ceci (1982) claimed to have established such bias even within the confines of the United States. They took 12 leading American psychological journals (all with a 'nonblind' refereeing policy) and resubmitted to each an article that each published 18–32 months previously. The only changes they made were to use fictitious author names and to substitute institutional affiliations of somewhat dubious credibility such as 'The Tri-Valley Institute for Human Growth'. Of the 12 papers, just three were recognized by the journals as having been previously published by them under other names. Of the other nine, one was accepted, but eight were rejected.

There may be other more unexpected perils facing the under-resourced researcher. Here are the comments of an editor of an international journal as reported to me:

> We get single copies of these papers from India. They are manually typed with an old ribbon on that grey recycled paper. As they won't photocopy there is I'm afraid little that we can do with them.

For want of a nail then.

The most serious attempt so far to situate the NNS researcher within a general framework is that of Jernudd and Baldauf (1987:164). As they say, 'we would argue that if one could understand the process by which the individual scientist makes choices, language correction procedures could be developed to improve communication and presumably specific language "products", which would lead in turn to an increase in information exchange which is related to the level of human resource development possible in any community'. A simplified version of their model of this process, and one also somewhat translated into the key concepts of this book, is presented overleaf as Figure 4.

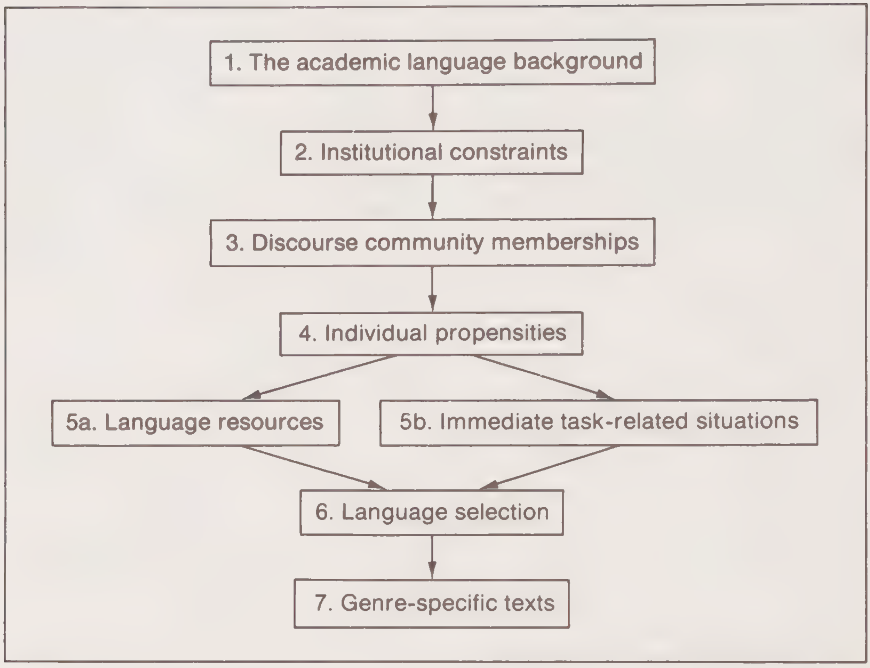

Figure 4 *Language selection in research* (adapted from Jernudd and Baldauf, 1987:172)

I shall discuss the operation of this process model by referring to the situation in the Arab World. The academic language background (Stage 1) in the Arab World reveals a not untypical sociolinguistic setting in which the main regional language, Arabic, co-exists with one or two imported and ex-colonial international languages. Arabic itself has several varieties, from pure colloquial sub-regional forms to the traditional literary and religious language of classical Arabic (El-Hassan, 1977). The variety relevant to research is MSA or Modern Standard Arabic, a modernized version of classical. Not all educated speakers of Arabic command MSA with equal facility and conviction, and the genre development of MSA as a language of research communication may not yet be complete (Najjar, 1988). By and large, English is the alternative language for research communication in areas that fell under British influence during the first half of this century (Iraq, Sudan, Saudi Arabia etc.), while French is the alternative language in areas of historic French influence (Syria, the Maghreb etc.). However, the neat distinction between French and English has broken down somewhat in the last decade or two, partly as a result of crossover patterns of individual advanced training and collaboration, but these can be handled under Stage 4.

Stage 1 shows that Arab researchers are likely to have two choices as the vehicle for their research publication: MSA, on the one hand, or English or French, on the other, (or possibly both if the researcher is Lebanese or Egyptian). These choices may be affected by institutional constraints (Stage 2). The National Research Council may have a language policy for its publications. The universities may impose constraints: at the University of Jordan promotion is not possible without at least one publication in Arabic, while at the University of Khartoum in Sudan, promotion may be blocked without publication in international refereed journals.

Stage 3 attempts to characterize the discourse communities that the individual researcher belongs to. If these are purely local, opportunities will be limited to the contacts available to that local community. However, membership in regional or more distant communities may bring valuable advantages of collaboration (including co-authorship), access to funds, and useful advice about where and how to submit a manuscript. Stage 4 then characterizes the individuality of the researcher. In the Arab World, for instance, quite a number of Arab researchers have received doctoral training in Russia or Eastern Europe and may therefore be as prepared to write in Slavic languages or in German as in English or French. Additionally, Stage 4 accounts for the important nationalistic variable, since, as Najjar (1989) shows, researchers vary considerably in the extent to which they feel a commitment to advance the cause of Modern Standard Arabic.

Stages 1 through 4 allow us to characterize the general language orientation of a researcher. Stages 5 through 7 relate to the immediate specifics of the communicative tasks in hand. One researcher might, for instance, have decided to write his or her research in Arabic but has recently recognized that additionally an abstract in English needs to be provided. Are there English-language resources (human, textual or both) in his or her environment to facilitate this task? And if there are, could they later be employed to produce an English-language paper based on another aspect of the research? Finally, the selection of a particular language as vehicle at Stage 6 then invites consideration (Stage 7) of both language-specific genre requirements (as was discussed in 3.7) and of the specific expectations of targeted journals. Thus, in these ways, linguistic choices are made and publications and presentations in particular languages constructed.

This model of language selection in research is a generalized one, and in reality a number of 'cut-outs' may operate. For example, the researcher may receive an *invitation* to participate in a communicative event with a previously established linguistic choice. More importantly perhaps, the model as it now stands fails to account directly for the variable of perceived quality. This variable is probably significant as there is strong

anecdotal evidence from around the world that many NNS researchers try to publish what they consider to be their 'best' research in a major international language, leaving their 'lesser' findings to appear in local or regional language format.

Although the process typified by the model is otiose for the native speaker of English, it definitely may not be for the non-native speaker. In particular, Stages 3 and 5a may be crucial. If these boxes are largely unfilled or infertile, then the chances of a researcher's output being reduced in both quantity and visibility are increased. As Baldauf (1986) observes, there may be a 'lost generation' of cross-cultural psychologists out there who have only a reading knowledge of the English language, the overwhelmingly dominant vehicle for information exchange in this subfield.

The geopolitical implications of being 'off-networked' are themselves not trivial. They raise serious questions about the effectiveness of the investments being made by LDCs in doctoral scholarships held by their nationals in the USA and Europe. They raise questions about the scientific, scholarly and developmental value of research scholarships and visitorships offered to LDC nationals by American, European and Russian governments and other sponsoring agencies. They raise questions about whether advanced country training is as relevant as it might be and about whether *institutional* mechanisms for initiating NNSs into appropriate discourse communities need to be strengthened. And finally they raise questions about the policies of academic English programs abroad – in particular with regard to their largely exclusive concern with undergraduate education. In consequence, we may perhaps legitimately ask whether these programs should not also be concerned with mitigating post-doctoral 're-entry shock' and with providing alternative frameworks for those not selected for advanced training outside of their local environments.

In Chapter 6 I have tried to assess our current knowledge about the role of English in contemporary research. The fact that reaching even a provisional assessment has occupied several pages should not, I believe, be seen as mere scholarly fussiness. Facts about the medium of communication, as about the medium of instruction, are *always* important because decisions to use a particular language inevitably confer advantage on some and disadvantage on others. For this reason I have been at pains to look at the available evidence from the perspectives of both the privileged and the less privileged, from the viewpoints of users of major bibliometric resources and from those of their critics, and from the stance of general survey and that of local investigation.

The role of English in research that has been depicted in this chapter offers a number of opportunities for action. One is to consider whether there is anything we can do to modify, where necessary, the attitudes of

NS scholars, researchers and students so that they are more tolerant of non-native speakers, more willing to accept them into their discourse communities and more aware of the extra burdens which they carry. A second is to resist a mindset which associates quality with location – a mindset, to take a real example, that would likely deny the possibility that a group of lecturers at the University of Tabriz in north-west Iran could have produced the most elegant ESP course design so far constructed (Bates, 1976). A third would be to ask whether we have the skills and interests to foster 'lesser' languages as vehicles for research communication, particularly by engaging in collaborative ventures aimed at developing and consolidating genres such as the research article in those languages. Fourthly, and as the most direct form of affirmative action, we may consider what roles we may play as teachers *of* and *about* research English.

As it happens, the texts cited throughout this chapter are very well suited for such an instructional purpose (Swales, 1987b). Themes such as the anglophone grip on contemporary research will be of direct relevance and concern to non-native speakers and are, at the least, educative for native speakers. The fact that the themes concern language makes them amenable to English-language instructors. The straightforward quantitative nature of many of the studies, such as tabulating locations of authors in a sample of RAs, may appeal to scientists and offers a methodology that can easily be replicated on a small-scale for class projects. Further, the methodology itself engages participants in searching and scanning processes (Bazerman, 1985) that can be valuably generalized for other purposes and tasks. Finally, the texts themselves, particularly those with a quantitative orientation, are prototypical exemplars of the RA genre. Figure 5 overleaf is part of one such text.

The text might be utilized in the following way:

a) The first objective is to get the class 'into' the text. An obvious place to start is with the interpretation of *North, Periphery* and *South*. So, blank maps of the world are distributed and individuals or pairs are asked to divide the main areas of the world up into these three categories. Discussion towards consensus follows, the instructor using a copy of the full article for reference.
b) A second priming activity is to raise consciousness about the data base. On the limited textual evidence, what is it? Where is it located? Is PASCAL a clue?
c) A preliminary writing task would be to supply the missing data-comment paragraph for Table 2. Pointers to possible items for written discussion might be: is the Portuguese/Spanish order surprising? Does PASCAL adequately cover the Far East? Mandarin, Thai, Bahasa? Paragraphs are compared – and compared to the original.

Findings

Agricultural sciences in 1983 were represented by 9398 references in the PASCAL data base, in those sections that we selected. Of these, 2040 references (21.7%) concerned a tropical environment or a tropical country. Table 1 shows the distribution of this production by areas.

TABLE 1. ORIGIN OF PUBLICATIONS ON TROPICAL AGRICULTURAL SCIENCES

North	420	21%
Periphery	536	26%
South	1042	51%
Others	42	2%
Total	2040	100%

The South produces half of the research on tropical areas – that is, 11% of the total of world research in agricultural sciences. This figure surpasses the traditional 'near to 6%' usually admitted for Third World countries. The Periphery produces about a quarter of tropical agricultural research, and the North a little less. The 'Others' represent difficult-to-identify references, anonymous items, affiliation to international institutions, and so on.

The linguistic distribution is shown in Table 2.

TABLE 2. LINGUISTIC DISTRIBUTION OF TROPICAL AGRICULTURE RESEARCH

English	1530	75%
French	204	10%
Portuguese	143	7%
Spanish	102	5%
German	20	1%
Others	41	2%
Total	2040	100%

Figure 5 *Language and location in research* (Arvanitis and Chatelin, 1988: 118–19)

d) Matters are now ready for a real task orchestrated as a class project (Hutchinson and Waters, 1987). The class divides into self-selecting fields of interest. The groups go to the library to obtain data from journals in their fields on origin of articles and language of publication. Group results are tabulated and presented orally; cross-group figures are assembled and discussed; comparisons are made with pre-

vious research; methodological procedures (and problems) are written up, as are suggestions for further research. A multi-author small research paper is constructed. NS colleagues are invited to attend a formal conference-type presentation and then to participate in a NS–NNS panel discussion on the issues of language barriers in research communication.