ASp
la revue du GERAS

11-14 | 1996
Actes du 17e colloque du GERAS

French researchers publishing in English

An analysis of a corpus of first drafts

Susan Birch-Bécaas

OpenEdition
Journals

This text was automatically generated on 30 April 2019.

French researchers publishing in English

An analysis of a corpus of first drafts

Susan Birch-Bécaas

1. Introduction

1   The pivotal role of the English language on the international stage is nowhere more evident than in the scientific research paper. It is of no surprise then that the anglicist has found himself drawn into the world of scientific communication and the analysis of the *specific purposes* of research articles and the exchanges that take place within this discourse community. In this way, it has been possible to describe the rhetorical patterns and linguistic regularities which underlie this discourse (Swales 1990, Hopkins & Dudley-Evans 1988) while others have focused on the sociological parameters which play a part in shaping and constraining the scientist's construction of knowledge claims (Latour & Woolgar 1988; Myers 1985). The objective in NNS language instruction has been to heighten the learner's rhetorical awareness of the genre and reinforce the language items which are used to realise the rhetoric (Weissberg & Buker 1990; Swales & Feak 1994).

2   In recent years, there has been increased collaboration between anglicists and scientists in the various universities and research institutions across France as francophone researchers are faced with the preponderant use of English as the language of publication and consequently feel they are at a linguistic and cultural disadvantage when drafting articles. For some years now, the D.L.V.P., Bordeaux 2 has offered a revision service to researchers preparing papers in English. Many researchers, sensitised to the dangers of submitting an unrevised paper to the journal or leaving it in the hands of a translator, prefer this option. During the sessions of re-reading and reformulation of scientific papers, it becomes apparent that the same sort of modifications are required from author to author and it seems that this may hold for researchers working in different domains

and with varying levels of language skills. This observation led us to suppose the existence of "a common-core of unknown or poorly acquired elements" (Cooke 1993: 470).

3   For this reason, the present study is based not on published articles but on a corpus of first drafts, written directly in English by the researchers and then corrected in our department by an anglicist. This would enable us to analyse the errors which recur in the writings of French scientists and the reformulations deemed necessary by the corrector in order that the paper be published. A definition of these problem areas would, we hoped, go some way to increasing the autonomy of the researcher in his writing and lead to a better quality draft. As Claude Sionis (1995: 112) points out:

> Only when full attention to the linguistic and stylistic requirements of the research article genre are fully identified *before* putting pen to paper or mouse to mouse pad will the EST teacher's role change from practising *post-mortems* to helping to deliver healthy scientific discourse.

2. Materials

4   As the corpus was compiled several criteria were taken into account:

5   **Homogeneity** - the corpus is composed of texts from just one specific genre, the research article. In this context the writer has a clear communicative purpose, to convey new information. A high degree of precision is required in the writing and the message can only be decoded by peers within the closed community.

6   **First drafts** - this initial phase of the article gives us an insight into the mental processes of the researcher and enables us to focus on the points at which there is a breakdown in communication.

7   **Researchers** - the corpus includes university, CNRS and INSERM researchers but not doctoral students as we wished to remain homogenous as far as the level of scientific competence was concerned and we also felt that the latter might experience additional organisational problems in the drafting of their articles. The researchers all had considerable experience of publishing both in French and English. Indeed analysis of the composing process revealed that they have very little problem in dealing with documentation in English and their English has in fact been learnt in this way by their " reading for professional purposes ". They thus have little difficulty in organising their work and following the conventions of the research article. The difference between anglophone and non-anglophone authors thus becomes more evident on the grammatical and lexical level.

8   **Availability** - some researchers had conscientiously archived all the stages of their publications, often with the intention of looking over the corrections before putting pen to paper for their latest article. Many though had only kept the final version or the off-prints.

9   **Structure** - the articles all follow the IMRAD[1] model (Introduction, Material and Methods, Results and Discussion). This was important as we wished to analyse the errors occurring in the different sections of the article and to establish the areas of discourse which are most problematic for French writers. We also tried to ensure that the articles were of a similar length.

10   **Disciplines** - all the researchers in the study work in the fields of medicine and the life sciences. The corpus falls naturally into 3 subdivisions, biochemistry (17 articles),

cancerology (10 articles) and a third group (13 articles) including more varied disciplines such as cardiology, psychology and oenology. The aim was to include the greatest number of disciplines and authors possible in order to take these factors into account. The corpus finally consisted of 40 articles, representing 24 authors.

11   **Period** - the corpus includes articles written between 1982 and 1993. We did not include any articles written before this date to make our results as relevant as possible in the constantly evolving world of publication.

12   It seemed likely that some of the problem areas would be inherent difficulties of the English language but that the communicative function of the discourse would heavily influence the structures used and thus the errors that occur, making them specific to scientific English, the genre of the research article and its subsections. It would thus be necessary to analyse not only the forms occurring but the functions of those forms. Kennedy and Bolitho (1984) suggested that scientific English has the same structures as general English but in different proportions and Halliday and Martin (1993: 4) define it as "a form of English in which certain words, and more significantly certain grammatical constructions, stand out as more highly favoured, while others recede and become less highly favoured than in other varieties of the language". Pierre Lerat (1995: 21) states that the frequency of certain forms is determined by the context of the utterance and the role of the language as a vector of knowledge. He gives the following definition:

> Une langue spécialisée ne se réduit pas à une terminologie: elle utilise des dénominations spécialisées (les termes), y compris des symboles non-linguistiques, dans des énoncés mobilisant les ressources ordinaires d'une langue donnée. On peut donc la définir comme l'usage d'une langue naturelle pour rendre compte techniquement de connaissances spécialisées.

3. Methods

13   The corpus was examined manually. A computerised analysis was impossible as the poor quality of certain documents and the hand-written corrections meant that the articles could not be scanned. Our first approach was to record any omissions, substitutions and errors on a grid but it soon became apparent that the errors were of little use to us without the full context and the function of the form in that particular case. Furthermore, the grid meant that we would be dealing with predetermined error types and that the categories were thus limited in advance. We therefore came to the same conclusion as Norrish (1983: 81) "the drawback of the first approach is that the issue is prejudged; errors will be found to fill categories, and the investigation takes on a certain circularity, since errors can be sorted out only in terms of predetermined error types". We thus decided to let the errors form the categories by recording them in a data base on Filemaker Pro, Macintosh (cf. figure 1).

14   Each card represents an error and its correction. The cards can then be grouped, counted and sorted according to different criteria: author, discipline or section of the article. The full context of each entry can be consulted in the corpus as the entry is identified by the number of the article and the page number. Each error and its correction were attributed to a certain use of the language and added to an existing category or a new category was created. Forty-two different categories were formed and these were then grouped into larger denominations or 'families' such as determination, tense, cohesion, and word order and finally three global categories, grammar, lexis and spelling. The figure given on the

card indicates the number of errors found for that particular category. For the 40 articles of the corpus we completed 2,928 cards.

15 From this base we were able to calculate the frequency of each category by sorting the cards, selecting each category and noting the number of cards found. These calculations were then repeated for each category in each of the 40 articles taken separately in order to compare the performance of each researcher with the results for the corpus as a whole. Certain patterns were found indicating that some errors do occur systematically in the drafts of these researchers. The author must use the language to achieve precise goals but the francophone researcher has not always mastered the specific lexico-grammatical elements which ensure this pragmatic success.

Figure 1

4. Results

4.1. Global categories

16 Focusing first of all on the global categories (cf. figure 2) we see that grammar represents 75% of the errors and corrections. 2189 of the 2928 cards concern grammatical problems. Lexis represents 20%, or 583 occurrences which were mainly problems of collocation, inappropriate lexical choices, L1 interference and use of compounds. Finally there were 156 spelling errors in the corpus, 5% of the total.
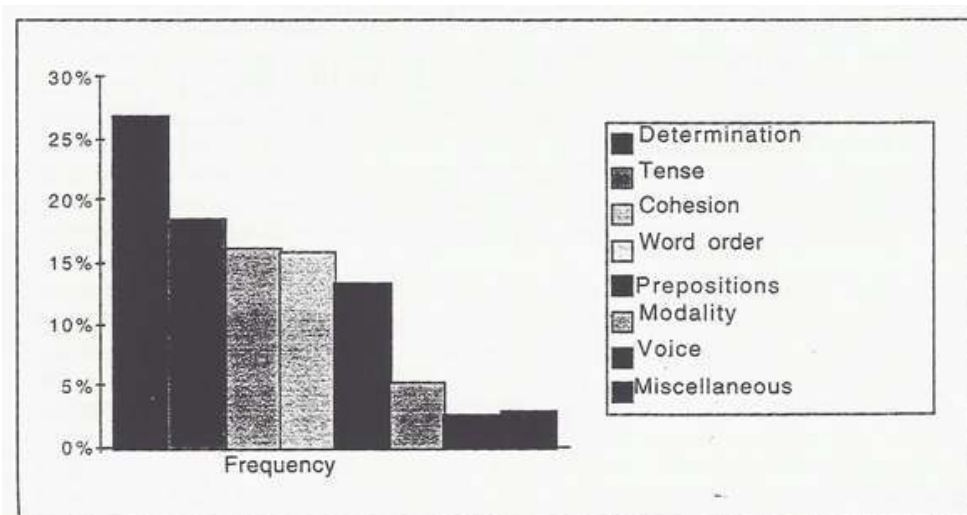
Figure 2
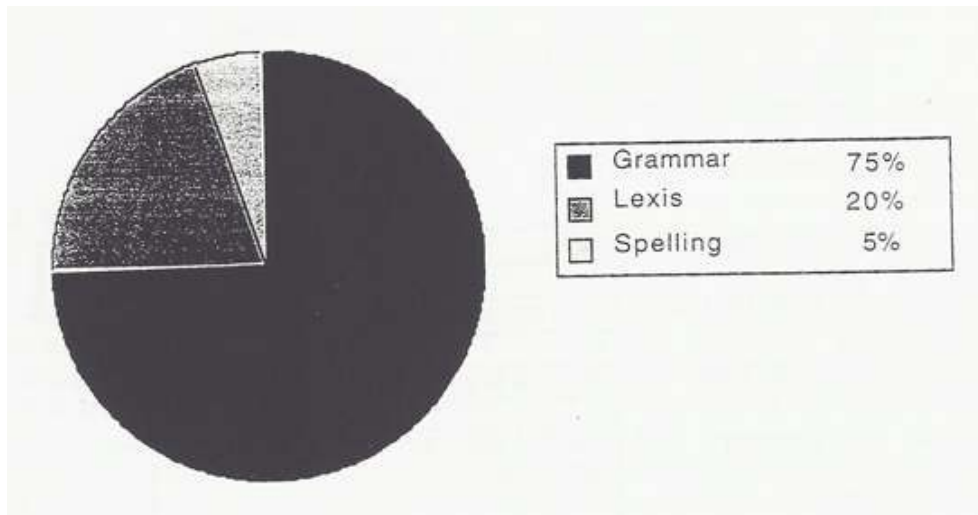
4.2. Families

17 Grammar can be broken down into determination, tense, cohesion, word order, prepositions, modality, voice and miscellaneous (problems of concord, the use of time expressions such as 'for, since and during') and these *families* are present in the proportions shown (cf. figure 3).

Figure 3

18 What are of interest are not the grammatical elements themselves but the reasons why they are used and to what effect. The NNS author must learn to manipulate grammar and select appropriate tense, voice and modality to achieve the desired effect. Lackstrom *et al.* (1972: 14) emphasise the importance of the choices made by the writer, "We are drawn from purely grammatical relationships to the attitudes and intentions of the writer and to the position of the sentence under discussion in its rhetorical relationships to the rest of the paragraph." It is necessary to analyse, not the language in its abstraction but its realisation, the specific features in their context. Textual and contextual factors, together
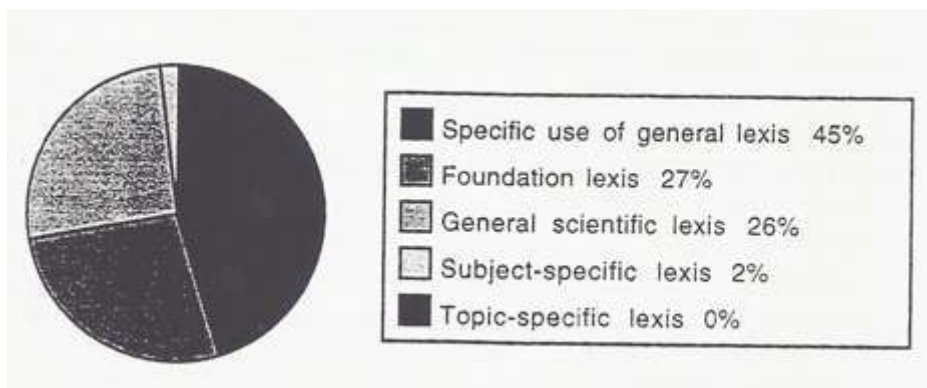
with the writer's awareness of the reader will influence the linguistic choices to be made and ultimately the way the content is interpreted.

19 **Determination** - the use and non-use of the article is one of the most difficult elements for NNSs to grasp. The research article contains many generic statements but in our corpus the definite article is often used where an anglophone researcher would use the zero article to refer to concepts, notions, abstract nouns, names of chemical compounds and techniques and references to figures, tables and illustrations. The scientific writer cannot always fall back on grammatical rules though as the article will often be omitted by NS researchers whose aim is to express their ideas in a minimum number of words. The French researcher must also master the uses of the definite article in contextual, situational and anaphoric references.

20 The misuse of **tenses** discovered in the corpus further illustrates the fact that we are not only dealing with grammatical relationships but the intentions of the writer. Thus the choice of tense may be determined by the degree of generality that the author attributes to his utterance. In the Introduction, the author should use the present simple to introduce general truths and make omnitemporal assertions. He will then need to justify his research project. However, the francophone author may have difficulty in situating his work amongst existing studies and nuancing his review of the literature. He must make a distinction between those studies of continuing relevance (he should use the present perfect) and other studies further removed (he should use the preterit.) The French researcher also tends to confuse the preterit and the present perfect for the description of procedures. Results should be presented in the simple past if the author is referring solely to his experiment and the case in hand but the NNS author must learn to shift to the present tense if he wishes to comment upon his observations or to emphasise the implications of his results. It is to be noted that many of the linguistic choices of the French researchers are not ungrammatical but they are inappropriate in the context and do not always translate the writer's intention.

21 It is also necessary to consider the structuring of the discourse with the families **cohesion** and **word order**. The French researcher must help his reader to interpret his ideas coherently by providing cues and markers to co-ordinate and connect statements. The functions frequently necessary and those not always mastered by the NNS are those of comparison, contrast, concession, cause and effect and the avoidance of repetition and ambiguity with demonstrative and anaphoric reference. As Sionis (1995: 111) points out, these effects can only be achieved by a mastery of the language itself and not the specialised elements:

> What ML (mathematical language) apparently cannot do adequately (for publication) are introductions and transitions, and also the expressing of notions like co-ordination, subordination, causality, restrictive hypothesising, the presentation of the chronological unfolding of a process, the definition of the physical environment of an experiment or process (diagrams and other visual aids are used in this instance).

22 The French researcher must also be aware of strategies that he can use to highlight certain aspects of the discourse. He will thus be able to draw attention to new information as opposed to presupposed knowledge and choose how to relate different elements of his discourse to advance his argument. He will also wish to incorporate a mass of information in as concise a text as possible. This leads to dense lexis and complex compound constructions. Halliday and Martin (1993: 119) term this "expert grammar" as it can only be decoded by members of the discourse community. For the NNS however,

the use and over-use of nouns as pre-modifiers is problematic. The use of adverbs of frequency and possibility is common and their position in the sentence is another source of confusion for the researchers. Other syntactic problems arise as the author attempts a concise style and compact prose.

23 **Prepositions** are problematic as the author must situate elements in time and space. He must also indicate movements, fluctuations and changes in state.

24 **Modality** can be used by the author to indicate his attitude towards the propositional content and the validation of relationships. The author must manipulate the language to express what is conceivable not impossible or logically deducible. The French author must fully realise the scope of his utterances, and be able to modulate the strength of his assertions. Through the conjectures and speculations his reputation will be at risk. As Latour and Woolgar (1988: 249) explain the aim of the research team is to turn hypotheses into facts, "*L'activité scientifique est faite de la défense de points de vue d'abord fictionnels qui sont parfois transformés en objets stabilisés*". If necessary the author will distance himself from the knowledge claims. To this effect he may use modal auxiliaries or compound hedges incorporating adverbs, lexical verbs and approximators. For Kourilova (1993: 16), the possible imprecision or overt directness of a NNS writer may have a cultural explanation, "this failure to internalise the English system of modality may however be rooted not only in language problems but also in different approaches to treating the truth".

25 **Voice** – the author may wish to move from the active to the passive or vice versa to highlight certain ideas or to distance himself from the prose.

26 The **miscellaneous** category is comprised of problems of concord subject— predicate and time expressions.

27 **Lexis** – this family subdivides into five categories but is dominated by what we have termed 'inappropriate lexical choices'. French authors are able to reuse words and expressions assimilated from their bibliography and their reading of source texts but collocations are difficult to assimilate. False cognates are tell-tale signs of their resorting to L1.

28 Looking at the category 'inappropriate lexical choice' (cf. figure 4), we can see there is no perceived difficulty with the specialised vocabulary but rather the specific use of general vocabulary in scientific writing such as verbs used for description, classification, analysis and methodology and the foundation lexis.

Figure 4

Specific use of general lexis 45%
Foundation lexis 27%
General scientific lexis 26%
Subject-specific lexis 2%
Topic-specific lexis 0%

29   Finally **spelling** errors were divided into two categories, those where there are rules to follow such as the formation of adverbs and the doubling of consonants and other morphological errors.

4.3. Individual categories

30   Table 1 presents the errors and corrections in their individual categories from the most frequent to the least represented. This gives us a more detailed breakdown of the problems mentioned above.

Table 1

Category	Description	Nb. occurrences	Frequency
LEXa	lnappropriate lexical choice	442	15.1%
PRE	Prepositions	286	9.8%
DETa	Determination - zero article	280	9.6%
DETb	Determination - definite article	230	7.9%
ORDa	The noun group, compounds	120	4.1%
ORDb	Position of the adverb	113	3.9%
ORDc	Sentence structure	93	3.2%
TPSa	Preterit - to describe a procedure	92	3.1%
MODa	Modality - auxiliaries	92	3.1
TPSb	Present - to comment on results	83	2.8%
ORTa	Spelling - rules to follow	83	2.8%
DETc	Determination – lndefinite article	74	2.5%
ORTb	Spelling - morphology	73	2.5%
TPSc	Preterit - reporting results	70	2.4%
LEXb	Problems of number	67	2.3%
COHa	Coordinating conjunctions	59	2.0%
COHb	Comparisons	58	2.0%
COHc	Demonstrative reference	56	1.9%
COHd	Adversative conjunctions	53	1.8%

TPSd	Present perfect - past studies	50	1.7%
LEXc	Forming nouns, compounds	49	1.7%
DIVa °k	Concord - subject-predicate	47	1.6%
VOia	Active/Passive	45	1.5%
TPSe	Present - general truths	38	1.3%
COHe	Causality	33	1.1%
COHf	Relatives	30	1.0%
TPSf	Tense formation	25	0.9%
LEXd	Errors with determinants	22	0.8%
COHj	Ellipsis	19	0.6%
ORDd	Verbal constructions	19	0.6%
MODb	Hedges	18	0.6%
TPSg	Preterit – past studies	17	0.6%
COHh	Conjunctions of time	15	0.5%
TPSh	Present – refs. to figures, etc.	14	0.5%
TPSi	Past Perfect	13	0.4%
DIVb	For/since/during	13	0.4%
COHi	There as subject	10	0.3%
VOIb	Passive/Active	8	0.3%
COHj	Conjunctions of addition	7	0.2%
COHk	Conjunctions of organisation	5	0.2%
COHl	Personal reference	4	0.1%
LEXe	Lexico-grammatical errors	3	0.1%
TOTAL		**2,988**	**100%**

31   The first category does not concern grammar but lexis, 'inappropriate lexical choice' accounts for 15.1% of the errors or 442 corrections made in the texts. This is followed by the substitution of one preposition for another 9.8%, the use of the zero article 9.6% and the definite article 7.9%. Then, we note categories concerning word order such as the

noun group, position of adverbs, sentence structure and these are followed by problems of tense choice, modals, spelling and so on.

4.4. The sections of the article

32 The next stage of the study was to calculate the frequency of the different mistakes in the different sections of the article to see to what extent the varying communicative purpose would influence linguistic requirements. *Material and Methods* was found to contain the least number of mistakes as it consists for the most part of a simple account and description of procedures in the past tense. The *Introduction* and *Discussion* were apparently more difficult to write as these sections of the text are more subjective, there is less support from model texts and the author must enter into the discourse to interpret his observations and convince the reader.

Table 2

Categories	Corpus	Abstract	Introduction	M & M	Results	Discussion
GRAMMAR	**75%**	**75%**	**74%**	**70%**	**77%**	**75%**
Determination	20%	21%	19%	25%	22%	17%
Tense	14%	16%	19%	10%	13%	13%
Cohesion	12%	10%	13%	8%	13%	13%
Word order	12%	10%	10%	10%	12%	13%
Prepositions	10%	12%	9%	10%	10%	9%
Modality	4%	3%	3%	1%	3%	6%
Voice	2%	2%	1%	2%	2%	2%
Miscellaneous	2%	2%	1%	4%	1%	2%
LEXIS	**20%**	**17%**	**21%**	**22%**	**18%**	**20%**
SPELLING	**5%**	**8%**	**4%**	**8%**	**5%**	**5%**

33 In Table 2 we can see that the *Abstract* follows the pattern of the corpus in general. However, in the *Introduction,* tense becomes more problematic than determination. The researcher has to situate his work with in the body of literature, refer to previous studies and incorporate generalisations of the domain. In *Material and Methods*, grammatical errors in general are less prevalent but determination poses problems as there are numerous references to methods, processes, compounds. Tenses appear to be easier to handle as it is a descriptive account, detailing the experiment which has been performed. The discourse may be easier to structure and we note fewer problems with cohesive devices. The increase in the miscellaneous category may be due to problems with 'for, since and during' to express the duration of experiments. The *Results* section follows the general tendencies of the corpus. However, in the *Discussion* there is an increase in errors of modality as the author must modulate the strength of his assertions. It would seem then that the linguistic choices of the author are closely linked to the rhetorical function of the discourse as certain elements of the language are used more frequently in certain parts of the discourse and certain errors recur in certain sections of the article.

5. Conclusion

34 The frequency counts are of interest but we must also take into account the relative gravity of the errors. Certain errors may lead to a breakdown in comprehension, others confuse the reader and some simply irritate the reader or referees. For this reason it is

interesting to study such articles from the point of view of the reviewer (Crosnier 1994; Birch 1996).

35   This study focuses on a corpus in medicine and the life sciences, however it would be interesting to explore other domains and the drafts of other groups of researchers with a different L1 see if there are common mistakes. This would make it possible to go some way towards explaining the origin of these errors.

36   As we stated at the beginning of this study, publishing in English is a key element in a French researcher's career. Paradoxically though, there are few manuals or 'tools', the odd handbook and style guide excepted, which cater for the specific needs of fully-fledged researchers as opposed to undergraduate science students.[2] The results of this study have, we believe, their pedagogical implications as they highlight problematic linguistic features in the context of he research article. S. P. Corder (1974: 48) defines the role of the anglicist as follows, "We are concerned with teaching him not just to produce grammatically acceptable strings of words but also to use language to some purpose, to communicate and be communicated to, to assure certain rules". Thus the anglicist should raise the researcher's awareness of the different rhetorical moves of the research article and the anglo-saxon patterns of argumentation and sensitise him to the linguistic devices used to establish these moves. To conclude we can only agree with Bhatia (1995: 165) who states that "Genre-based grammatical explanations raise learners' awareness of the rationale of the text-genre that they are required to read and write".

BIBLIOGRAPHY

Bhatia, Vijay K. 1995. "Applied genre analysis and ESP". *TESOL France Journal* 2/2, Functional Approaches to Written Text. Classroom Applications, 161-179.

Birch, Susan. 1996. "Les besoins linguistiques des scientifiques français publiant en anglais : analyse d'un corpus de premières rédactions et de leur correction". Thèse de doctorat de l'Université Bordeaux 2, sous la direction de M. Perrin.

Cooke, Ray. 1993. "Learning to publish in English: How can French researchers bridge the gap?". *ASp* 1, 463-474,

Corder, S. Pit. 1973. *Introducing Applied Linguistics.* Harmondsworth, Eng.: Penguin.

Crosnier, Elisabeth. 1994. "Enquête sur l'évaluation de l'anglais par les anglophones dans les publications des scientifiques français". *ASp* 3, 41-55.

Halliday, M.A.K. and J.R. Martin. 1993. *Writing Science. Literacy and discursive power*. London: Falmer Press.

Hopkins, Anthony and Tony Dudley-Evans. 1988. "A Genre-based investigation of the discussion sections in articles and dissertations". *English for Specific Purposes* 7, 113-121.

Kennedy, Chris and Rod Bolitho. 1994. *English for Specific Purposes.* London: Macmillan.

Kourilova, Magda. 1993. "Epistemic modality in written scientific discourse". *UNESCO-ALSED LSP Newsletter* 15/2, 4-18.

Lackstrom, John, Larry Selinker and Louis Trimble. 1972. "Grammar and technical English". *English Language Teaching Forum 10/5, 3-14.*

Latour, Bruno and Steve Woolgar. 1988. *La vie de laboratoire : la production des faits scientifiques.* Paris: Éditions de la Découverte.

Lerat, Pierre. 1995. *Les langues specialises.* Paris: Presses Universitaires de France.

Myers, Greg. 1985. "Text as knowledge claims: The social construction of two biology articles". *Social Studies of Science* 15, 593-630.

Norrish, J. 1983. *Language Learners and their Errors.* London: Macmillan.

Sionis, Claude. 1995. "Communication Strategies in the Writing of Scientific Research Articles by Non-native Users of English". *English for Specific Purposes* 14/2, 99-113.

Swales, John and Christine Feak. 1994. *Academic Writing for Graduate Students: Essential Tasks and Skills.* Ann Arbor: University of Michigan Press.

Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings.* Cambridge: Cambridge University Press.

Weissberg, R and S. Buker. 1990. *Writing Up Research. Experimental Research Report Writing for Students of English.* Englewood Cliffs: Prentice Hall Regents.

NOTES

1. Structure defined by the American National Standards Institute in 1979.
2. The team of Soula *et al.*, INSA Toulouse, are currently working on a multimedia project to aid researchers in their preparation of oral presentations for scientific congresses in English.

ABSTRACTS

This study focuses on a corpus of research articles written by French native speaker researchers. Their first drafts, written directly in English, were then corrected by a revision service. An error analysis of the corpus highlights the problem areas for French scientists and enables the author to pinpoint, in the corrections, appropriate language forms and the rhetoric that they are used to express. In this paper the author presents the recurrent errors and corrections and their distribution across the different sections of the research article.

Cette étude porte sur un corpus de premiers jets rédigés directement en anglais par des chercheurs francophones, et corrigés ensuite dans un service de révision par un correcteur angliciste. Une analyse des erreurs commises par les chercheurs et des corrections effectuées par l'angliciste vise à identifier les éléments qui font problème. La fréquence d'occurrence et la dispersion des erreurs est analysée et commentée.

INDEX

Mots-clés: article de revue scientifique, chercheur scientifique, erreur et correction, première rédaction
Keywords: error and correction, first draft, researcher, scientific journal article

AUTHOR

SUSAN BIRCH-BÉCAAS

Susan Birch-Bécaas enseigne notamment aux étudiants de médecine d'odontologie et de pharmacie au DLVP de l'Université Victor Segalen Bordeaux 2. Elle participe également au fonctionnement du SAPAG : Service d'apprentissage et de perfectionnement en autonomie guidée. susan.birch@u-bordeaux2.fr