Help on the Spot:
Online Assistance for
Writing Scientific English

Ray Cooke & Sue Birch-Becaas
Université Victor Segalen Bordeaux 2
France

In the ESP literature, much attention has been focused on the problems of NNSE (non-native speakers of English) researchers attempting to publish their results in English (St John 1987, Swales 1990, Shaw 1991, Gosden 1995). These researchers often face many setbacks compared with NS (native speaker) writers (Flowerdew 1999). It takes them longer to draft articles and they require assistance from peers, supervisors, Anglophone correctors and even translators. Yet publication in English is of paramount importance to these researchers as far as career evolution is concerned. It is well established that the research article (RA) is the most important genre for these researchers, and the aim of most writing courses is to give assistance by raising the writers' awareness of the rhetorical and linguistic characteristics of the genre which have been described by many past studies (Hopkins and Dudley-Evans 1988, Swales 1990, Gosden 1992, Salager-Meyer 1992, Hyland 1994). The problems of novice writers and doctoral students have also been well documented (Belcher 1994, Blakeslee 1997, Flowerdew 2000, Li 2006). Studies have illustrated how they gradually acquire genre knowledge and become more proficient writers. In France, the norm is for doctoral students to publish at least two articles during their studies. There is therefore a lot at stake in being accepted for publication and gradually becoming integrated into the discourse community. It therefore seems that for NNSE researchers writing up research is a difficult task, especially if we take the example of 'off-network scholars' who are hoping to be published in the more prestigious journals (Belcher 2007).

There seem, however, to be few tools available to help learners improve their writing. The style guides, handbooks and instructions to authors[1] tend to focus mainly on advice on presentation and journal requirements but rarely develop questions of rhetoric or language. With the exceptions of Swales & Feak (1994, 2000) and Weissberg and Buker (1990), most writing textbooks seem to focus on tasks for undergraduate students. Furthermore, published EAP materials have been criticised on several counts. Harwood found, "a lack of fit between how academic writers write and what the textbooks teach about writing" (2005: 150). Hyland (1994) also pointed out that materials are often based on inappropriate sources such as popular or simplified accounts. Misak et al. (2005) report how, as editors of the Croatian Medical Journal, they assist authors with manuscript preparation but it is unlikely that the editors and reviewers of other journals would spend as much time and make so much effort in helping NNSE with their manuscripts. Therefore it seems to us that there is a need for a simple, accessible tool which can be used autonomously by learners to aid them with the drafting of manuscripts.

Recent research has reported on writing courses and the use of corpora in the classroom (Yoon & Hirvela 2004, Flowerdew 2005, Lee & Swales 2006, Charles 2007). The use of concordancing tools has been shown to be very useful for students on the lexico-grammatical level. However Lee & Swales (2006) warn against 'concordancing burnout' and Flowerdew (2005) underlines the limitations of a purely bottom-up approach where learners may become lost within a mass of information which lacks its communicative context. It is therefore important to go beyond word lists and frequency counts and to study texts first of all on a rhetorical level before analysing how these rhetorical moves are expressed. Flowerdew advocates combining concordancing with a top-down approach which would also highlight the rhetorical aspects of a corpus "where the analyst is probably also the compiler and does have familiarity with the wider socio-cultural context in which the text was created, or else had access to specialist informants in the area (Flowerdew 2005: 260).

Our objective here was to compile a corpus of NNSE drafts[2] which would form the basis of a tool that could be used autonomously by both novice and more experienced NNSE writers. Our experience of re-reading, editing and revising articles with the researchers at our university has shown that even experienced writers make typical language errors on the sentence level. Writing courses for doctoral students and interviews with the students and their supervisors have shown that novice writers also require help to organise their text and to imitate the typical moves, learning how to refer to the literature, to emphasise the importance of their study and not just to report

but also interpret their findings. As the corpus is analysed manually, the interesting occurrences can be studied in their complete context, guiding the learners and drawing their attention to rhetorical and linguistic features and the conventions of research writing. The corpus thus consists of drafts written by successful NNSE writers. There is first of all an emphasis on the writing process as the errors, reformulations and improvements made to the text can be studied. The drafts are then analysed on several different levels. The highlighting of the move structure enables novices to examine information structure. Subsequently, the focus is on various surface level features which are problematic for both novice and more experienced writers: typical expressions, tense usage, grammatical points, links and vocabulary.

The tool, which is called TYOS®[3], also contains a concordancer so that users can visualise how a particular word occurs in the corpus. For example, they may know the word "evidence" but are unsure of which verbs accompany it. By consulting the concordancer, it becomes clear that "provide" accompanies "evidence" and not "give", which would tend to be used in legal English, for example.

As the reader will have understood, TYOS® is a model-based tool. One limitation of model-based language learning is that it may limit the writer's inventiveness by keeping him firmly within the bounds of tried and tested manuscripts and texts. Scientific communication, however, is a domain in which there is little scope for creative writing, and in which consciously imitating a tried and tested model is actually desirable, even if repetitive. Demonstrating one's knowledge of accurately formulated scientific English to one's peers - in the peer review process – is possible only if one is aware of what constitutes "good scientific English". Yet experts disagree about what this really is. Perhaps the only valid definition is the level of expression that allows writing to be accepted for publication, provided the science itself is of sufficient quality. As mentioned above, all scientific journals have "Instructions for Authors" that may specify the style to be used and refer authors to style manuals. The problem with these, however, is that the information they contain is not immediately exploitable for drafting manuscripts – which is why TYOS® has been designed on the basis of a model that users may recognize and imitate.

Finally, TYOS® contains a "learning box" whose objective is to allow users to store newly acquired information in a space of their own and according to a framework specially designed for students of scientific writing. In this respect, TYOS® is aimed not only at novice and fully fledged researchers but also at ESP students in general. Furthermore, it can be used as a teaching

tool, not only to teach writing but also to help learners improve their reading skills, since it demonstrates how sentences are broken down into units.

The idea for TYOS® came about several years ago. One of us (RC) realised that manuscript track changes, i.e. the demonstration of writing errors, could become useful pedagogical input for learning, provided they were processed and displayed in an efficient, user-friendly mode. The challenge, therefore, was to find a suitable corpus and to process it didactically. The corpus posed no problem in that locally-based researchers and students in the health and life sciences had requested editing assistance from our team for many years, so their imperfect initial drafts formed part of our database. They were therefore requested to provide written informed consent for these drafts to be used as input for a new pedagogical tool whose objectives were fully explained to them. The response was unanimously favourable. The next step was to decide which texts would constitute the corpus. For this, we applied two criteria: local representativeness, since we felt that all local research specialties should be represented if possible; and genre variants, since our database had accrued over the years to include a wide range of genres used academically. This led to the inclusion of drafts in the fields of medicine, biology, biochemistry, wine science, dentistry, psychology and pharmacology. The genres we included were full research articles, abstracts, case reports, cover letters, letters of request, letters of complaint, replies to reviewers and remarks as reviewer.

On the other hand, the challenge was to establish a common processing framework for each draft and to decide what sort of information future users should have available. Furthermore, we decided that information of any kind should be given in as simple and user-friendly manner as possible, hence ruling out the use of complex descriptions of language points. Since Birch's doctoral thesis (Birch 1996) had identified a typology of typical errors committed by Francophone researchers when writing research articles, we decided upon five forms of analysis: use of tense, discourse linkers, interesting grammar points, vocabulary points and useful phrases and expressions. Full research articles were also analysed for their discourse moves. The content of each form was then established so that the draft mark-up phase could be conducted harmoniously by two people working separately.[4] Once this phase was complete, each file had to be converted into HTML[5] mark-up language so that it could be incorporated into a program specially designed for the purpose by an IT colleague. The interface was then completed, the scroll-down menus built and TYOS® generally became a usable web-based tool.

1. The tool

On opening TYOS®, the user goes into the menu and chooses the genre he wishes to consult. He also opens his word-processing application and opens a new file. If the genre he wishes to consult is "Case Reports", he can choose to browse several case reports or just focus on one. Each text is processed didactically in such a way that the user has access to several versions of the same text. All versions are processed to appear as one. A floating toolbar is used to display the versions available and the user simply moves to and fro between them. Once he finds an occurrence, phrase or sentence that he thinks he can reutilise in his own manuscript, he selects it and drops it into the word processing window that he has already opened.

To make a simple analogy, TYOS® helps the user to put cement between his bricks: to put structure into content that he already has. For example, scientist X studies the helical structure of yeast rotors. TYOS® will show him "The present article investigates..." and several alternatives for introducing the research subject. He can then create the sentence he wants to write. At no time will the user be copying unauthorised material because TYOS® does not contain any finally published versions of manuscripts, only initial drafts for which an authorisation for use has been granted by the authors.

The following section describes the versions available. All the didactic processing visible here was done with WORD®. However, TYOS® contains its own proprietary display system designed by its developers. The example used here comes from the "cover letters" category.

1.1 The initial draft

The initial draft version shows what the author wrote originally together with the corrections, editing and reformulations of the NS corrector. The learner can thus study typical errors, how the draft was improved, the increased clarity and the reinforcement or weakening of certain statements, for example. The focus is thus on the writing process. The other versions of the text are more 'product-orientated', with a highlighting system designed to raise awareness of certain features. Figures 1 and 2 show the initial draft and the corrector's changes.

Dear Dr AAAA

Please find attached our manuscript entitled *"Inter-expert agreement of seven operational criteria in causality assessment of adverse drug reactions."*
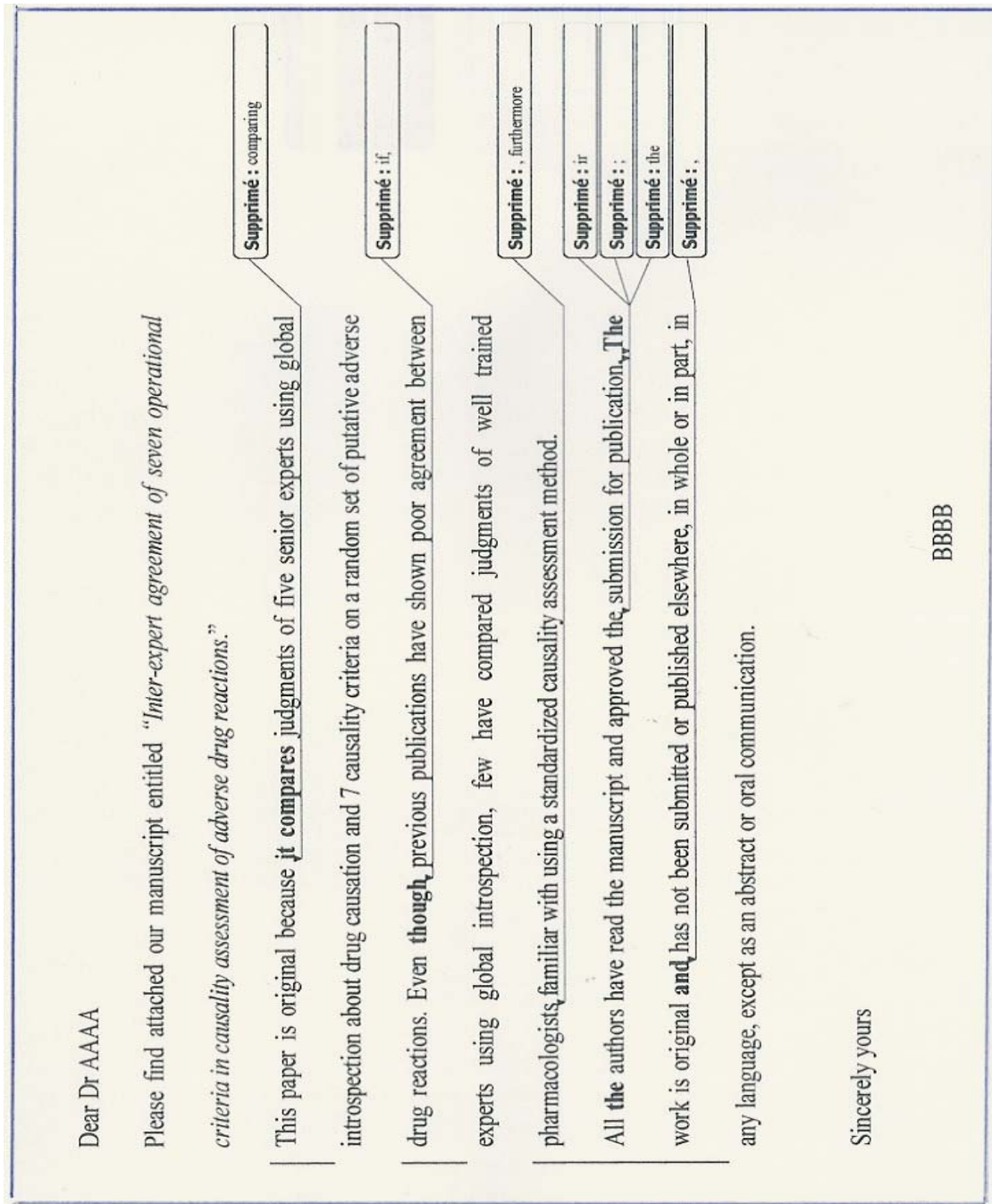
This paper is original because comparing judgments of five senior experts using global introspection about drug causation and 7 causality criteria on a random set of putative adverse drug reactions. Even if, previous publications have shown poor agreement between experts using global introspection, few have compared judgments of well trained pharmacologists, furthermore familiar with using a standardized causality assessment method.

All authors have read the manuscript and approved their submission for publication; the work is original, has not been submitted or published elsewhere, in whole or in part, in any language, except as an abstract or oral communication.

Sincerely yours

BBBB

Figure 1: Initial draft

Dear Dr AAAA

Please find attached our manuscript entitled *"Inter-expert agreement of seven operational criteria in causality assessment of adverse drug reactions."*

This paper is original because **it compares** judgments of five senior experts using global introspection about drug causation and 7 causality criteria on a random set of putative adverse drug reactions. Even **though** previous publications have shown poor agreement between experts using global introspection, few have compared judgments of well trained pharmacologists familiar with using a standardized causality assessment method.

All **the** authors have read the manuscript and approved the submission for publication. The work is original **and** has not been submitted or published elsewhere, in whole or in part, in any language, except as an abstract or oral communication.

Sincerely yours

BBBB

Supprimé : comparing

Supprimé : if,

Supprimé : , furthermore

Supprimé : ir

Supprimé : ;

Supprimé : the

Supprimé : ,

Figure 2: Initial draft with corrections

1.2 Discourse moves (only concerns research articles)

On this version of the text, attention is drawn to the organisational patterns, the author's intention and the rhetorical function of the sentences. As mentioned above, in our experience, fully-fledged researchers may be implicitly aware of the typical moves of science writing, while doctoral students and novice writers may need to imitate the structure and moves in their own writing. We have found that although this sort of genre knowledge is often acquired as they gain more experience and by interacting and writing collaboratively with their supervisors and team members, it would seem that there is a need for the rhetorical aspects of the discourse to be explained more explicitly. It appears that the most difficult rhetorical aspects of the discourse for doctoral students to learn are how to review the literature, how to underline the importance and interest of their study and how to express opinions on their findings (Gosden 1995, Hyland 2004, Birch 2008).

1.3 Typical expressions

A third version of the text (Figure 3) highlights typical expressions used. Our interviews with specialist informants and previous research (Shaw 1991) have shown that writers often use the articles of their bibliography intuitively to help them to pick out useful expressions for their drafts. This version formalises that process by highlighting reusable portions of text that serve a particular purpose. Users can integrate these portions into their own writing and are thus introduced to a set of expressions and linguistic devices for expressing a wide range of needs, such as indicating a knowledge gap for their community of researchers or (as above) for claiming originality. These expressions may be one of the factors that distinguish more experienced, mature writing from student writing and identify it as being part of the devices customarily used by a discourse community. As Hyland claims: "it is often a failure to use native-like formulaic sequences which identifies students as outsiders and there is a general consensus that formulaic sequences are difficult for L2 learners to acquire" (2007: 4). In science, it is essential to be seen as fully belonging to your community. By showing them the linguistic conventions germane to their socio-professional groups, TYOS® helps users to do that.

Dear Dr AAAA

Please find attached our manuscript entitled *"Inter-expert agreement of seven operational criteria in causality assessment of adverse drug reactions."*

This paper is original because it compares judgments of five senior experts using global introspection about drug causation and seven causality criteria on a random set of putative adverse drug reactions. Even though previous publications have shown poor agreement between experts using global introspection, few have compared judgments of well-trained pharmacologists familiar with using a standardized causality assessment method.

All the authors have read the manuscript and approved the submission for publication. The work is original and has not been submitted or published elsewhere, in whole or in part, in any language, except as an abstract or oral communication.

Sincerely yours

BBBB

Figure 3: Typical expressions (yellow highlighting)

1.4 Grammar

Through the re-reading and editing of drafts, Cooke became aware of a "common core of unknown or poorly acquired elements" (Cooke 1993: 470). These typical sentence-level errors were highlighted in a corpus of first drafts (Birch 1996)[6]. In this version of the text (Figure 4), the common core of problematic grammatical elements is highlighted such as the use of the article, noun groups, irregular plurals and prepositions. Of course, exactly what constitutes a "problematic grammatical element" is a matter of debate, other than defining it as any item recurring with noticeable frequency in a

corpus of initial drafts produced by a non-Anglophone language group such as French or Chinese native speakers. The objective of TYOS®, however, is to draw the user's attention, at some indefinable but foreseeable point in time, to a set of knowledge that he is likely to need to acquire in order to become a proficient communicator in scientific English.
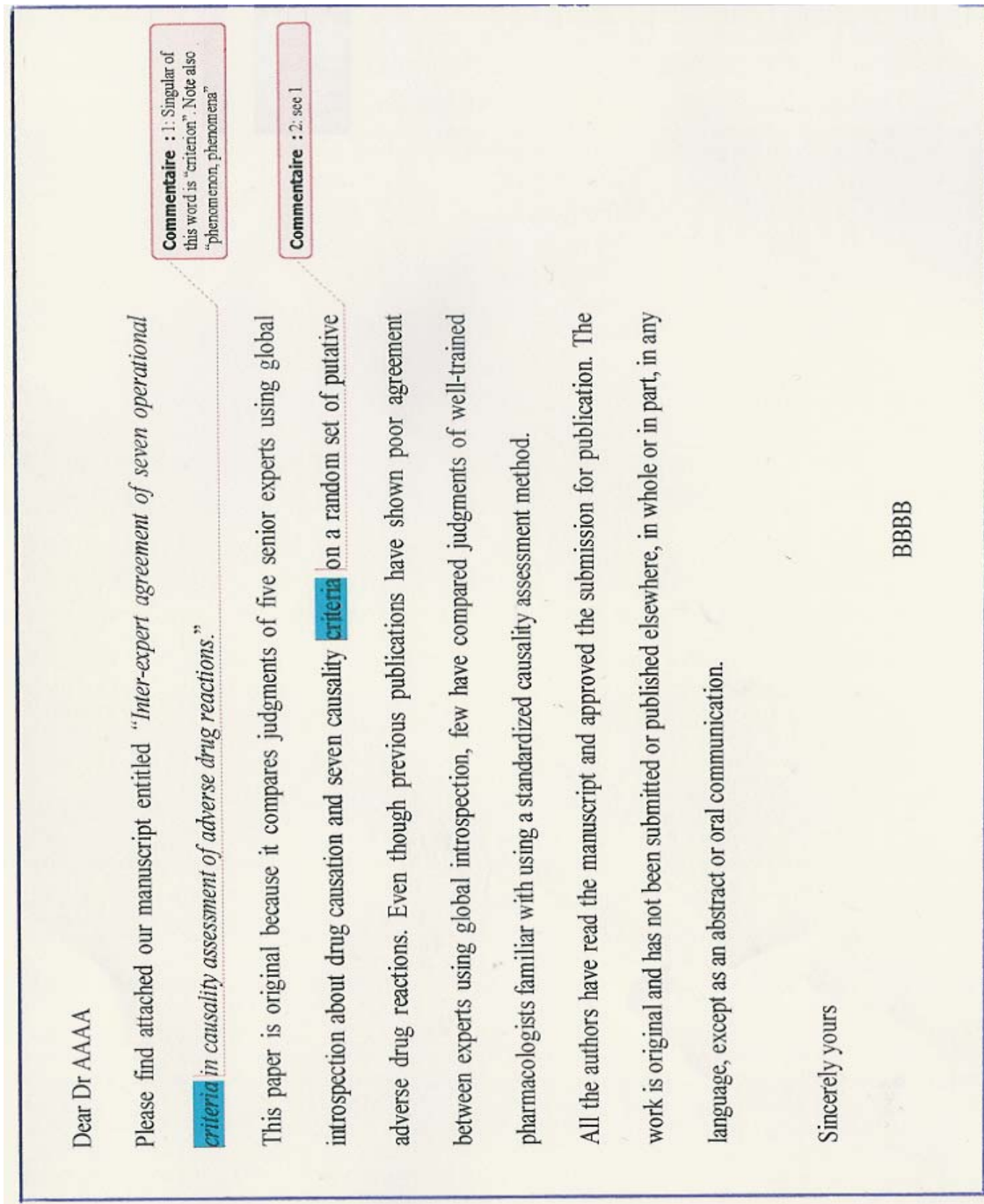
Dear Dr AAAA

Please find attached our manuscript entitled *"Inter-expert agreement of seven operational criteria in causality assessment of adverse drug reactions."*

Commentaire : 1: Singular of this word is "criterion". Note also "phenomenon, phenomena"

This paper is original because it compares judgments of five senior experts using global introspection about drug causation and seven causality criteria on a random set of putative adverse drug reactions. Even though previous publications have shown poor agreement between experts using global introspection, few have compared judgments of well-trained pharmacologists familiar with using a standardized causality assessment method.

Commentaire : 2: see 1

All the authors have read the manuscript and approved the submission for publication. The work is original and has not been submitted or published elsewhere, in whole or in part, in any language, except as an abstract or oral communication.

Sincerely yours

BBBB

Figure 4: Problematic grammatical elements (blue highlighting)

1.5 Use of tenses

Another version of the text (Figure 5) highlights tense usage. This has been shown to be a problem for NNSE writers, especially in the introduction and discussion sections of the RA. The French authors we work with have a tendency to describe their study in the present perfect rather than giving their methodology and reporting their results in the past simple. They also need to distinguish between specific references to past studies in the past simple and more general references to the literature in the present perfect. Appropriate use of tense is therefore another important aspect showing other members of a scientific discourse community that an author fully "belongs" to it.

Figure 5: Tense usage (green highlighting)

1.6 Linkers

It is important to highlight linkers and cohesive devices, which are important signals for the reader. Hyland (2004) has underlined the importance of mastering these forms of metadiscourse. Note that in this version (Figure 6), the French language is used for some explanations. Subsequent versions of explanations regarding linkers will have recourse to other widely used languages. While TYOS® has been designed with a Francophone audience in mind and by using a corpus of drafts produced by Francophones, it could in principle be used by anybody who is interested in improving their scientific English.

Figure 6: Discourse linkers (mauve highlighting)

1.7 Vocabulary

The focus here (Figure 7) is on semi-technical vocabulary and not the specialised disciplinary terminology that students of science and researchers are already likely to know. Emphasis is placed, therefore, on points of language likely to be of interest for a large number of users, irrespective of their speciality or native language.

Dear Dr AAAA

Please find attached our manuscript entitled "Inter-expert agreement of seven operational criteria in causality assessment of adverse drug reactions."

This paper is original because it compares judgments of five senior experts using global introspection about drug causation and seven causality criteria on a random set of putative adverse drug reactions. Even though previous publications have shown poor agreement between experts using global introspection, few have compared judgments of well-trained pharmacologists familiar with using a standardized causality assessment method.

All the authors have read the manuscript and approved the submission for publication. The work is original and has not been submitted or published elsewhere, in whole or in part, in any language, except as an abstract or oral communication.

Sincerely yours

BBBB

Commentaire :1: This word is being used more and more in electronic mail, simply because there is no longer any physical act of enclosure as there used to be when putting something into an envelope.

Commentaire :2: Note spelling of this word.

Commentaire :3: The French equivalent would be "mauvais". However, we avoid using the word "bad" in scientific English (which really has a derogatory sense) and prefer the word "poor".

Figure 7: Vocabulary points (grey highlighting)

108

Although public launch of TYOS® is imminent, a beta prototype has already been tested with a group of potential users composed of fully fledged researchers and doctoral students. This validation process, which will be fully described in a forthcoming article, clearly showed to what extent the subjects were satisfied by having pertinent information about writing scientific English immediately available at their fingertips. As they browsed, the information they encountered met visibly with their approval since they were already familiar with it, or it met their need in the particular writing task assigned to them. The fact that information is gleaned in an entirely random manner, and not according to a linear progression based on a schema contained in a manual, would seem to provide learners/users with a form of freedom to direct their learning and using. To provide scope for guided learning, however, learners/users could perhaps move back and fro between TYOS® and a set of exercises on the basis of directive links. Such an evolution is already being examined by the development team, as are other developments that will be announced in the coming year, such as classroom suggestions for ESP teachers. Furthermore, TYOS® circumvents a long-standing debate in ESP, i.e., in which order are items to be presented to the learner? The very notion of browsing and discovering means that a set of information is perused and implemented, if the user so decides. In this way, the same information is seen by different users but in no predictable sequence, only in response to a need. Learning is likely facilitated in this way, and future research will focus on how learners learn with TYOS® and how efficient this process is.

In conclusion, we hope that researchers and ESP students nearing the end of their university education will from now on be somewhat more autonomous in their writing and learning of scientific English than before. While TYOS® will never allow its users to produce the "perfect draft" nor iron out all the grammatical, structural and stylistic errors that non-native speakers' drafts contain, it does provide insights into the micro and macro aspects of writing that may have escaped their attention before. In this sense, it is a complementary tool that could be used alongside the proprietary spellcheckers and style correction tools that are now used worldwide.

References

Belcher D. 1994. The Apprenticeship Approach to Advanced Academic Literacy: Graduate Students and Their Mentors. *English for Specific Purposes* 13/1, 23-34.

Belcher D. 2007. "Seeking acceptance in an English-only research world". *Journal of Second Language Writing* 16, 1-22.

Berkenkotter C., T.N. Huckin & J. Ackerman. 1991. Social Context and Socially Constructed Texts: The Initiation of a Graduate Student into a Writing Research Community. *In* C. Bazerman & J. Paradis (eds.) *Textual Dynamics of the Professions. Historical and Contemporary Studies of Writing in Professional Communities.* University of Wisconsin Press: Madison Wi, 191-215.

Birch, S. 1996 « Les besoins linguistiques des scientifiques français publiant en anglais : analyse d'un corpus de premières rédactions et de leur correction. » Unpublished PhD dissertation, Victor Segalen University, Bordeaux 2.

Birch-Bécaas, S. 2008 "Growing in Maturity: From first draft to Final Version, a French Doctoral Students's Research Writing", *Les Cahiers de l'Apliut-* Vol XXVII N°1, 100-111.

Blakeslee A. M. 1997. Activity, Context, Interaction, and Authority: Learning to Write Scientific Papers In Situ. *Journal of Business and Technical Communication* 11/2, 125-169.

Charles, M. 2007 "Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions". JEAP, 6, 289-302.

Cooke, R. 1993 "Learning to Publish in English: How can French Researchers Bridge the Gap?" *ASp*, 1, 463-474.

Flowerdew J. 1999. "Problems in Writing for Scholarly Publication in English: The case of Hong Kong" *Journal of Second Language Writing* 8/3, 243-264.

Flowerdew J. 2000. Discourse Community, Legitimate Peripheral Participation, and the Nonnative-English-Speaking Scholar. *TESOL Quarterly* 34/1, 127-150.

Flowerdew, L. 2005 "An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies." ESP, 24, 321-332.

Gosden, H. 1992 « Discourse Functions of Marked Theme in Scientific Research Articles", *ESP*, 11, 207-224.

Gosden H. 1995. Success in Research Article Writing and Revision: A Social-Constructionist Perspective. *English for Specific Purposes* 14/1, 37-57.

Harwood, N. 2005 "What do we want EAP teaching materials for?" JEAP, 4, 149-161.

Hopkins, A., & T. Dudley-Evans 1988 "A Genre-based Investigation of the Discussion Sections in Articles and Dissertations", *ESP*, 7, 113-121.

Hyland, K. 1994 "Hedging in Academic Writing and EAP textbooks", *ESP*, 13, 3 239-256.

Hyland, K. 2004, "Disciplinary interactions: metadiscourse in L2 postgraduate writing." *Journal of Second Language Writing*, 13, 133-151.

Hyland, K. 2007 "As can be seen: Lexical bundles and disciplinary variation." *English for Specific Purposes*, doi:10.1016/j.esp.2007.06.001.

Lee, D. & J. Swales., 2006. "A Corpus-based EAP Course for NNS Doctoral Students: Moving from available corpora to self-compiled corpora". *English for Specific Purposes,* 25, 56-75.

Li Y. 2006b. "A doctoral student of physics writing for publication: A sociopolitically-orientated case study". *English for Specific Purposes* 25, 456-478.

Misak, A., M. Marusic & A. Marusic 2005 "Manuscript editing as a way of teaching academic writing: Experience from a small scientific journal". JSLW, 14, 122-131.

Salager-Meyer, F. 1992 « A Text-type and Move Analysis Study of Verb Tense and Modality Distribution in Medical English Abstracts », *ESP*, 11, 93-113.

Shaw P., 1991. Science Research Students' Composing Processes, *English for Specific Purposes 10, 189-206.*

St John, M. 1987. "Writing Processes of Spanish Scientists Publishing in English", *ESP,* 6, 113-120.

Yoon, H., & A. Hirvela, 2004 "ESL student attitudes toward corpus use in L2 writing" JSLW, 13, 257-283.

***

---

[1] A good example is the journal *Neurology*, which has the highest impact factor in its field. The following is taken from the Instructions to Authors.
"Competition for the limited space in *Neurology* is intense, and well-written papers have the best chance of being accepted. Be certain your words express your ideas and message. Write simply and concisely, adhering to Billings' rules: "(1) Have something to say; (2) Say it; (3) Stop as soon as you have said it." Otherwise, the scientific value of your manuscript may be obscured. The editor's office and publisher will not rewrite poorly written manuscripts. Those not fluent in English should seek help from a colleague or a professional author's editor who does this for a fee."
http://www.neurology.org/misc/sugg.htm

[2] The corpus contains not only research articles but also case reports, abstracts, replies to reviewers, letters as reviewers, and letters of submission, recommendation, request and complaint.

[3] Patent pending. Public launch is planned for January 2009. The website of TYOS® is www.tyos.org

[4] The whole design and mark-up process was conducted during vacations and weekends over a two-year period.

[5] Hypertext Mark-up Language: a mark-up language used to structure text and multimedia documents and to set up hypertext links between documents.

[6] Birch's doctoral thesis (1996) focusing on 40 first drafts of scientific articles written by Francophone biologists and biochemists revealed the between-subject and between-researcher recurrence of language errors.