Language Re-use among Chinese Apprentice Scientists Writing for Publication

JOHN FLOWERDEW and YONGYAN LI
University of Leeds, City University of Hong Kong

Plagiarism has been a topic of considerable discussion in Applied Linguistics. In the literature on plagiarism a distinction can be found between the taking of the ideas of others and the taking of others’ words. In this paper the focus is on the latter, which is referred to as ‘language re-use’. Specifically, the study focuses on the practices and beliefs of a group of doctoral science students at a major university in China regarding language re-use in writing for publication in English. Examples are presented illustrating the students’ strategies of language re-use in each section of the prototypical IMRD (Introduction, Method, Results, Discussion) structure of the genre of scientific research articles, along with the writers’ justifications for such writing practices. It can be seen that the students’ language re-use goes well beyond formulaic expressions and technical terminology which are characteristics of the scientific research article, yet the students believe that their textual practices do not constitute plagiarism, which, to them, primarily means the stealing of others’ work. For English for Academic Purposes (EAP) instruction targeted at novice scientists, the paper calls for a pedagogy that acknowledges and exploits the formulaic nature of scientific writing as well as discusses the relationship between ‘form’ (language) and ‘content’ (the work reported) in natural sciences.

INTRODUCTION

A scientific paper authored by a group of researchers at Peking University (Pan et al. 1994) was found to have plagiarized from a paper written by two Canadian researchers (Misra and Gedamu 1989). In Pan et al. (1994), the second paragraph (about 280 words) of the Introduction and the last paragraph (about 150 words) of the Discussion were taken almost word-for-word from Misra and Gedamu (1989). Including other plagiarized parts, almost one third of Pan et al.’s paper was identical to Misra and Gedamu (1989), whose paper was not included in Pan et al.’s references. Following the accusation, the Chinese authors of the paper in question argued that although ‘there is a significant degree of identity in the wording’, the accusation of plagiarism was not justified ‘because we have all the original data’ (as reported in X.-G. Li and Xiong 1996). After a personal investigation of the case in Beijing, the Dutch Editor in Chief of Plant Molecular Biology...
It is my finding that the similarity between both papers, dealing with the same subject, only concerns certain parts of the text. The data published in *Plant Molecular Biology* by Dr. Pan et al. are from original work carried out under the direction of Dr. Pan in Prof. Dr. Ru’s laboratory for several years. A different gene and different methods are used for plant transformation and analysis of the expression of the mouse metallothionein-I gene. It remains the case that it is not acceptable practice to copy text—not even small passages—from published materials without citation reference. This is regardless of severe English language difficulties, as was clearly the case with preparation of the manuscript by Dr. Pan et al. . . . (Schilperoort 1995: v; emphasis added)

From the above quote, it can be seen that the Editor of *PMB* was considering plagiarism from two perspectives: that of the ‘work’ itself (the ideas), on the one hand (where he found no evidence of plagiarism) and that of textual copying—what we will refer to in this paper as ‘language re-use’—(where he did find evidence of plagiarism), on the other. Although the Dutch editor may well have been understating the seriousness of the case—for it has been concluded by others that Pan *et al.*’s case did indeed involve the stealing of others’ ‘work’ (P. Li and Xue 1996; X.-G. Li and Xiong 1996)—what is of particular interest to us is the reference to Pan *et al.*’s ‘severe English language difficulties’ (as identified by the editor of *PMB*) in their misconduct and what this means for other writers, especially those whose native language is not English. Academic Science is now a truly international activity and, increasingly, scientists who use English, the *lingua franca* of International Science, use it as an additional language, that is they have had to learn it through the education system rather than as their mother tongue. In addition, as a result of the globalization of scholarship, internationally, scientists are being encouraged to publish in prestigious international journals, which are invariably published in English (e.g. Ammon 2001; Wood 2001; Tardy 2004). The Dutch editor described language re-use, even of ‘small passages’, as ‘unacceptable’, and yet, as we will show in our review of the literature and in our empirical study, such practices are common among both first and second language writers, especially those who are apprentices to the discipline.

**LITERATURE REVIEW**

**Understanding language re-use**

A considerable number of studies among literacy educationists are increasingly acknowledging what has been variously referred to as ‘textual plagiarism’ (Pecorari 2003), ‘patchwriting’ (Howard 1993, 1995), and ‘textual
borrowing’ (Currie 1998; Casanave 2004; Shi 2004), as a bona-fide writing strategy employed by writers who are learners of a target discourse. This strategy has been justified by a number of potential factors: (1) belief on the part of the plagiarizing students that a certain extent of language re-use from other texts is acceptable (i.e. not plagiarism, which students may have been severely warned against); (2) linguistic and cognitive overload in fulfilling certain writing tasks—and for students for whom English is a second language and who lack linguistic flexibility this can be especially salient; and (3) unfamiliarity with the discourse of the target disciplinary community (hence lack of confidence) (e.g. Howard 1993; Spack 1997; Currie 1998; Ivanicˇ 1998; Jones and Freeman 2003; Pecorari 2003).

Apart from these attributions, a very influential explanation in recent years in the literature when it comes to the issue of language re-use is in terms of cultural difference, and here much of the literature has had to do with Chinese students, based on researchers’ experience of working with such students either in Chinese (Matalene 1985; Deckert 1993; Pennycook 1994, 1996; Scollon 1995; Sapp 2002) or in North American contexts (Gregg 1986; Bloch, 2001). Matalene (1985), for example, attributed her Chinese students’ language re-use to conventions of Chinese traditional literacy where memorization of classic and model texts is strongly emphasized. Similarly, Gregg (1986) considered how Chinese students may tend to depend on the words of established authorities in writing. While we would agree that the cultural orientation adopted by these writers has been useful in fostering an understanding of alternative text-making practices, we concur with a number of researchers (e.g. Liu 2005; Pecorari 2003; Casanave 2004; V. Ramanathan, personal communication, December 2005) that this cultural interpretation is inadequate for explaining plagiaristic behaviours; the most obvious evidence for this is that it has been extensively shown that language re-use is found with English-L1 students (e.g. Hull and Rose 1989; Howard 1993; Prior 1998) as well as among students from a wide range of cultural backgrounds (e.g. Spack 1997; Currie 1998; Ivanicˇ 1998; Dryden 1999; Angéwil-Carter 2000; Pecorari 2003; Shi 2004).

However, closely related with the cultural interpretation yet particularly enlightening if viewed separately, is a historical perspective upon the notion of plagiarism, which holds that the overwhelming emphasis on the authorial self in the West can be traced back to the rise of the individualization of authorship in the seventeenth and eighteenth centuries (Pennycook 1996) and the rise of the Utilitarian ideology established with the Enlightenment (Scollon 1995). Considered as a historical construct, plagiarism would be susceptible to change in different historical times even in the Western context.

**Language re-use as generic practice**

A major gap in the previous approaches to the issue of language re-use, we believe, is an inadequate attention to the conventions of text/language re-use
in particular disciplinary/professional fields. To us, Jameson’s (1993) genre approach to examining text re-use in a range of genres (e.g. the news article and different types of reports in business/academic contexts) effectively emphasized an alternative perspective, showing that ‘what would constitute culpable plagiarism in one context might constitute proper use of sources in another context depending on the group whose expectations defined “misappropriation”’ (1993: 20, as cited in Angénil-Carter (2000: 30); also see Bell (1991), in the case of the production of daily news reports, and Selzer (1983), in the case of an engineer producing technical reports by incorporating the text of previous documents).

What about the text-making process of the genre of scientific research articles? A few studies that involved mature scientist writers have also produced evidence of language re-use. Rymer (1988: 234) reported an English-L1 scientist who attached much importance to writing a good starting sentence in any section of text—he might invent such sentences himself or take them from a previous publication of his or someone else’s paper. The following extract is from the think-aloud protocol of the scientist:

I have before me a paper by _____ and I’m going to snitch some of the phraseology... So I used his sentence structure, but I put my own, put it into my own words and rearranged it some.... I’ve got my start.

In another study, Dubois (1988) reported that biomedical scientists commented that directly extracting ‘common phrases’ from others’ work is normal in their writing. Studying Spanish-L1 scientists’ writing processes, St. John (1987) presented extracts of her participants’ text, with the extracts appearing in two separate multiple-authored articles but the two articles having a common author. As St. John noted, the extracts exemplified how the Spanish-L1 scientists ‘transfer [the description of similar methods] from one paper to another’ (1987: 117). St. John quoted one author involved as saying: ‘I use a previous description and just add my own bit’ (ibid.).

**Language re-use and science students**

Although there have, to our knowledge, been only two studies of the writing practices of science students, they are significant in their focus on language re-use. In one of these studies, Jones and Freeman (2003) examined the laboratory report writing of some first year physics students in an Australian university, and distinguished three levels of what they called ‘imitation’: (1) word-for-word copying; (2) calquing (‘mapping new content/lexis onto fixed structures’); and (3) paraphrasing (‘the content remains unchanged, but the lexis and structures vary’). In the other study that concerned undergraduate science students, Krishnan and Kathpalia (2002) analysed engineering undergraduates’ strategies in using source materials
for writing literature reviews in their final year projects, and found that the students practise: (1) stringing of summaries (where students juxtapose summative reviews of different works, in an incoherent form); (2) plagiphrasing (‘combin[ing] words/phrases/sentences from published works with their own’ (Krishnan and Kathpalia 2002: 193)); (3) shadowing (hiding behind a chorus of others’ voices ‘instead of providing their own claims or comments’ (2002: 194)); (4) and direct quoting (i.e. by quotation marks, which is somewhat against the norm in science/engineering disciplines).

RATIONALE FOR THE STUDY

Re-use of ‘phraseology’, ‘sentence structure’, ‘common phrases’, or ‘previous (methodological) description’ at the length of a line or so, as described in the previous section differs from the kind of extensive copying, with the intention to deceive, as exemplified by Pan et al.’s case described at the beginning of this paper. The problem is, however, that there are further variations of the type of copying, between the presumably harmless re-use of common phrases and technical terms (Swales and Feak 1994; Barks and Watts 2001) on the one hand, and the obvious theft of science on the other.

Little is known about the scenarios of source-based language re-use that science students are engaged in when writing. With regards to science students at graduate level, it seems there has been a lack of focused analysis of their language re-use except mention of the practice, or, brief sampling of textual evidence (in writing dissertations (Shaw 1991; Dong 1996) or research articles (Yakhontova 2001; Pecorari 2003; Y.-Y. Li 2005)). A further gap in the analysis performed in the above-mentioned studies is that, even where there is presentation of textual examples, there is:

1 no consideration given to the particular part of the text where the re-use occurs; and
2 no attempt to investigate the motivations for the practices, i.e. testimonial evidence on the part of the writers.

Such an analysis, focusing on in which parts of a text and for what purposes students engage in what types of language re-use would be particularly valuable if we take a genre perspective toward scientific writing. The present study therefore addresses this gap in the literature, by focusing on the practices and beliefs of a particular group of doctoral science students in a major university in China regarding language re-use in writing for publication in English.

RESEARCH GOALS

This study was part of a larger project which triangulates issues, theoretical perspectives, and methodologies in investigating the beliefs and practices
of Chinese doctoral students in a range of scientific disciplines in writing for international publication (Y.-Y. Li 2002, 2006a, 2006b, 2007). The study was conducted at a large, research-oriented comprehensive university (henceforth XU) in East China. At the university, publishing papers in English, especially in journals included in the Science Citation Index (SCI), is strongly encouraged among doctoral students in all science disciplines and it is a graduation requirement for doctoral students in basic research-oriented disciplines such as Physics, Chemistry, Astronomy, and Medicine.\(^1\) The goal of this particular aspect of the study was to investigate the student writers’ beliefs and practices regarding the use of language taken from published papers and from the work of their colleagues in writing their own research papers intended for international publication. Such a goal had emerged from the preliminary stage of work for the larger project, where the second-named author of the present article conducted extensive interviews with doctoral students in science disciplines at XU and it was found that using language from journal articles seemed to be a prevalent strategy among the students in their writing. This practice had previously also been identified by the first-named author in an earlier project investigating the scholarly writing practices of Hong Kong academics (Flowerdew 1999a, 1999b, 2000, 2005). The basis was thus provided for our focused investigation of this particular phenomenon.

METHOD

Data collection

The data employed in this report were of two types: textual data and interview data.

Textual data

The textual data to be reported on in this paper consist of excerpts reproduced from participant Ph.D. student writers’ draft papers aimed at publication in international SCI indexed scientific journals published in English, together with the corresponding model/source texts writers had used in producing their own texts. The participants’ texts were willingly provided by email by the participants either before or after the interviews to show how they utilized specialist literature in composing their own texts. Our student participants, like those in Pecorari (2003), generally did not hesitate in providing the pairs of texts and pointing out how their own texts were based on their model texts, presumably because they tended to perceive their practice as a necessary and prevalent strategy in research writing. The student-authored texts contained in the examples to be given later in the present paper are either in their published form, or as submitted to a journal, or as they were when handed in to supervisors.
Interview data

The interview data in this study were collected by the Chinese author during January 2003 to October 2005 at XU, through interviews with nine doctoral students from a range of disciplines, as indicated in Table 1.

The nine student participants in the interviews were an ‘opportunistic’ sample (Miles and Huberman 1994) in that eight of them had been in the Chinese author’s English class previously and that they were the students who quickly responded to her email requesting interview participation, while S8 was an acquaintance of the Chinese author (the interviewer). S3 (oldest in age among the nine) was at a visibly lower level in English proficiency than the other eight students, who were at a similar level of proficiency. The nine students became the participants of long-term case studies over several years, as part of the broader project already referred to. The willingness of the participants to collaborate in the research and the long-term relationship between the interviewer and the participants are likely to have increased the level of trust placed in the interviewer on the part of the participants in discussing an issue that is potentially sensitive.

The interviews, which were all conducted in Modern Standard Chinese, lasted about an hour, with all of the participants being interviewed several times as part of the broader project. The exploration of how these participants adopted/re-used language from other work thus tended to be mixed in with other issues, such as discussion of their writing strategies during the paper writing process and their paper submission experience. This ‘indirect’ approach to the issue is viewed as a positive feature of the

Table 1: Participants involved in the interviews in this study: their year of Ph.D. study and previous first-author English publication experience at the time of data collection

<table>
<thead>
<tr>
<th>Disciplines</th>
<th>Year 1 English papers published/accepted</th>
<th>Year 2 English papers published/accepted</th>
<th>Year 3 English papers published/accepted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics</td>
<td>S1 0</td>
<td>S2 2</td>
<td>S4 11 or 12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S3 2</td>
<td></td>
</tr>
<tr>
<td>Chemistry</td>
<td>S5 2</td>
<td>S6 2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S7 2</td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td></td>
<td>S8 1</td>
<td></td>
</tr>
<tr>
<td>Astronomy</td>
<td>S9 0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
study, insofar as it is likely to have helped allay any mistrust of the interviewer on the part of the participants, as well as providing a contextualized view into the participants’ language re-use in action. The interviews can be described as ‘reflective’ (Hammersley and Atkinson 1995), that is the interviewer did not start with a set of specific questions, but rather a number of broad issues, these evolving as more interviews took place. While providing some structure in terms of what was or was not relevant and eliciting clarification where ambiguity occurred (Spradley 1979), the aim was to allow the participants a lot of freedom in directing the discussion in directions which they thought were relevant, and to minimize the researcher interviewer’s influence on the interviewees’ comments and observations.

During the interviews, discussion on language re-use was sometimes based on the participants’ own texts, an exception being an interview with S9, which included a discussion of a text by a fellow student, because S9 happened to be reading it, along with a source text, at the time). This approach has been referred to as a discourse-based interview (Odell et al. 1983; Hyland 2000). The aim of the approach is to elicit information concerning the learners’ tacit knowledge about the rhetorical context of their texts, that is why they write things the way they do, especially, in our case, their use of copying. Special care was taken by the interviewer not to connect the students’ writing practices with the term chao xi (the equivalent term in Chinese for ‘plagiarism’; for discussion of the term chao xi, see Liu 2005). Instead, the students tended to be asked questions such as: ‘What do you think of the similarity between the two?’ and ‘What extent of copying is acceptable to you?’ In addition to the students, two post-doctoral fellows and eight Chinese professors were also interviewed, as well as a number of international journal editors. For reasons of space, however, this data is referred to only very briefly in this paper. Apart from the interviews, on many occasions, especially during the writing up of this study, follow-up discussions were conducted and clarifications were made via email or in person.

**Data analysis**

A comparative reading approach toward the textual data was adopted. This consisted of student text and corresponding source text(s) (all provided by the students themselves, as noted above). The text-analytical framework can be described as genre based. Given our familiarity with the extensive research into the genre of the scientific research article, with its rigid structure of the IMRD and its standardized associated communicative moves and functions (see Swales 2004 for a recent overview of this research), we found the overlappings between student texts and source texts as well as the functions of such overlapping passages relatively easy to identify.
While it has been argued in the literature that interview responses should be viewed as co-constructed through the interview interaction (Silverman 2001; Gubrium and Holstein 2002), we would nevertheless concur with Hyland (2000: 144)—who relied on discourse-based interviews in researching expert disciplinary writing—that ‘it [interviewing] seems to be the most effective way of bringing the insider’s perspective to the analysis, taking us nearer to a description of cultural practices in terms of its members’ understandings’. We do not take our participants’ statements to be either true or false reports, but rather as displays of their perspectives on the issue at hand (Silverman 2001: 112). While there may be more ambitious ways of analysing our interview data, such as ethnomethodology or various forms of discourse analysis, our emphasis is on the descriptive triangulation of our interview data with our textual data (student text and source texts), focusing on the issue of language re-use. We believe a number of factors have further enhanced the reliability of our interview data. These factors would include: our long-term liaison with all the participant students (with the present study being part of a larger project extending over a number of years, as mentioned above), the trust that the student participants placed in the interviewer (as mentioned above), and interviews being conducted many times with each participant and their beliefs being consistently reiterated by themselves over time.

In our analysis, the interview data were transcribed and translated into English. The analysis, which considered both the Chinese and English versions of the transcripts, followed an inductive approach which involved a cyclical process of data collection, analysis, creation of hypotheses, and testing of hypotheses in further data collection. The perspective was an emic one, utilizing as much as possible the language the respondents themselves used. As already mentioned, this is why, for example, care was taken not to use the word chao xi (plagiarism) unless it was used by one of the participants. In analysing the data, the aim was to look for both commonalities and differences between the participants in terms of their beliefs and practices.

The analysis and interpretation of the interview and the textual data were conducted by both researchers working independently and together. This provided an element of investigator triangulation to the analysis (Patton 1987), with the Chinese researcher bringing ‘familiarity’ with the situation and the non-Chinese researcher maintaining an element of ‘strangeness’ (Coordinating Committee 2004).

FINDINGS

Language re-use as a developmental strategy

The science doctoral students, newly into their doctoral programmes at the university in question, generally have a strong desire to improve
their English. They see the shortcomings of their previous learning in English (characterized by an emphasis on passing various tests) and recognize that this neither enabled them to speak English with any fluency, nor to write English presentably. Not surprisingly, trying to meet the graduation requirement of publishing in an English-Language SCI journal definitely poses a big challenge for the students, who have, in general, had little or no experience of writing a research paper in English. Neither do they have an English course in the university that focuses on research paper writing. Given this situation, students find their greatest help to be what Casanave and Vandrick (2003) have referred to as the ‘textual mentor’, that is the published journal articles in their own fields.

For our student participants, learning to use language from other texts can take two primary forms: (1) keeping notes of useful expressions, either in traditional notebook form or in word processing files and (2) using a selection of articles and taking useful language as they write.

The following is an example of the former. It is taken from an Microsoft® Word document file (S5, Chemistry). It can be seen from this example that S5 extracted whole sentences from the literature, highlighting in bold particular words and expressions and giving Chinese translations to them.

Excerpts from S5’s computerized record of useful expressions

The film thickness (ranging between 20–1000 nm) is controllable by changing either the potential or the duration of its application. (膜厚)

The size of the branches in copper and tin deposits are of the order of hundreds of nanometers. (為…數量級)

In contrast, in the case of tin deposition, hydrogen bubbles are generated mostly from the copper substrate due to a higher hydrogen overvoltage on tin. (相反, 對於而言, 由於 因為)

The present review focuses entirely on silica-based materials… In this article, In this paper, Here we report…(本文…)

Nevertheless, silica presents attractive properties (absorption capacity, acid/base chemistry, thermal stability) which could be advantageously exploited, for example in the accumulation of electroactive analytes before their electrochemical detection. (具有, 表現出)

Shown in Figure 4a is the TEM image of the branch, revealing that the copper branches consist of copper particles of several tens of nanometers. (顯示, 揭示, 表現出)

This figure reveals that the electrophoretic field extends quite far into the waste reservoir. (顯示, 揭示, 表現出)
Figure 2 depicts the fluorescence spectra obtained from NR and NR-Si in bulk ethanol solution. (描述)

But, as with other students who had used this method at sometime, S5 found it hard to stick to the habit for long, because persisting in taking notes while reading is quite ‘troublesome’, as he put it.

A prevalent scenario involving almost all of the participants in this study is having a collection of papers (often the key references for the paper under construction) available while drafting their own. The papers are either in print form (downloaded from the Internet), or some, in electronic version (having been downloaded from the Internet and saved on the computer). Whether in print or electronic form, the papers tend to bear markings at places of potentially ‘useful’ expressions and sentences. Having a set of references at hand while writing, as S5 put it, really facilitates composing process:

Now, I take ten references when I write, when I come to a point, not knowing how to express myself. I’ll look up the papers; it’s possible that one sentence is adopted. I’ve only replaced certain words with what I need, but the sentence pattern may be the same.

Reporting on the same practice, one of S5’s lab-mates, S6, said:

The first draft—we have to build up a framework—make it look like a paper; if you write it all by yourself, you definitely cannot produce a paper if you all do it by yourself…we may take from different sources, synthesize them, and express our own meaning. If we see a particularly good expression, we’ll definitely use it.

What is not clear here, though, is what sorts of ‘good expressions’ the students take from source texts.

Another lab-mate, S7, agreed, but emphasized that as she revised repeatedly, her own language would gradually replace others’ words:

The first draft—based on other papers—build up the framework; starting from the second draft, replace gradually. Now if you ask me to tell which sentence is from which paper, it’s definitely hard. It’s not possible to find in my paper a whole sentence, a large passage the same as other’s. After being revised dozens of times—it has become your own language.

Notably, not every student would readily assert that his/her paper (either in the draft stage or even occasionally in the published form) does not contain a single copied sentence; our later examples in the following section show that copying to the extent of ‘patches’ is perhaps not rare among our student participants.

It can be added that apart from referring to a collection of articles while writing, some students favour accumulation of ‘materials’ before writing.
For example, S4 (Physics) and S8 (Medicine) habitually copy and paste potentially useful expressions and passages (the latter, in the case of S8, often containing research findings to be cited) into their paper draft, at the end of each of the IMRD sections. They said this is to prepare *su cai* [materials] for writing.

In the interviews, many students referred to the linguistically formulaic and fixed structure of research papers. A number of students explicitly pointed out how such features of their disciplinary papers had made their language copying both necessary and possible. S2 (Physics), for example, observed,

‘Scientific English’ has its own features and regularities. We can obtain a feel for ‘scientific English’ through literature reading in the research-preparation stage, to fulfill a transition from ‘not familiar’ to ‘a bit familiar’; and with ‘mindful’ attention to the special expressions in the field and *zhai chao* [extracting and copying], we can greatly improve the speed of scientific paper writing.

Similarly, S4 (Physics) said,

I think this [language copying] is certainly natural, because language is imitation. Scientific writing needs imitation even more. Why, because scientific writing . . . its form is fixed, so even more possibility for imitation.

As we will show in the following section, with individual sections of IMRD, the students have further justifications for their language re-use.

**Students’ practices of language re-use**

In the following, we will provide examples to show the nature of copying and adaptation among the participants in this study, along with participants’ justifications for such practices. In accordance with our data analysis, we will arrange our examples according to the four sections of the IMRD structure of typical scientific articles: the introduction, method, results, and discussion sections, and delineate further ‘subcategories’ as necessary. This manner of presentation, we believe, both summarizes our participants’ case effectively and enables us to highlight a genre-based dimension in examining the students’ textual practices.

**Introductions and discussions**

We will deal with introductions and discussions together, as in both sections (especially the introductions), it is essential for an author to map out intertextual relations (Berkenkotter and Huckin 1995), that is refer to the literature, thereby opening up the opportunity for novice writers to copy
language from other texts. We found it useful to distinguish two categories of language re-use in introductions and discussions: referring to specific findings in previous work (where the source is cited); and stating ‘shared’ knowledge (where the source tends not to be cited).

Referring to specific findings in previous work

In Example 1, S1 (Physics), was reiterating previous work with citations in the introduction, but reproducing stretches of sentences from the two sources as well:

Example 1

S1’s text: (as in the version handed in to the supervisor)

Formation of oxynitride alloys and the Burstein-Moss effect with high charge carrier concentrations may be responsible for sizable changes in the bandgap. [6] ...[48 words] [7]. J. Wu et al. persisted that the electron concentration dependence of the optical absorption edge energy was fully accounted for by the Burstein–Moss shift, O and H impurities couldn’t fully account for the free electron concentration in the films. [8]

Source text [6]: (in a review article)

The Burstein-Moss effect in polycrystalline samples with high charge carrier concentrations may also be responsible for sizable changes in the bandgap.

Source text [8]: (in the Abstract of a research article)

The electron concentration dependence of the optical absorption edge energy is fully accounted for by the Burstein–Moss shift. Results of secondary ion mass spectrometry measurements indicate that O and H impurities cannot fully account for the free electron concentration in the films.

Example 2 is taken from the discussion section of a biomedical paper, S8’s (Medicine) paper in English:

Example 2

S8’s text: (as in the version handed in to her supervisor)

Apoptotic cell was first described by Kerr Wylie and Currie in 1972 [19,20] as controlled cell death, with characterized morphology as nuclear and cytoplasmic condensations and the later formation of numerous apoptotic bodies. Biochemical features were now added as exposure of phosphatidylserine (PS) on the outer leaflet of the plasma membrane and the internucleosomal cleavage of DNA, resulting in the observed DNA laddering [20,21].
Two passages from the source article [21]

The term apoptosis was first used to describe controlled cell deletion by Kerr, Wylie and Currie in 1972 (REFS 5,6) [sic]. It was characterized morphologically on the basis of nuclear and cytoplasmic condensations and the breaking up of the cell into numerous membrane-bound vesicles, known as apoptotic bodies. To these morphological criteria have now been added biochemical features of programmed cell death, which include exposure of phosphatidylserine on the outer leaflet of the plasma membrane, proteolytic cleavage of many intracellular molecules (for example, poly-ADP-ribose polymerase, PARP) and the internucleosomal cleavage of DNA, resulting in the observed DNA laddering in late-stage apoptotic cells.

Commenting on the similarity between her text and the source, S8 explained that her version was not chao xi (plagiarism), because she had given the source of the cited material, indicating that ‘the work’ in the sentence was not her own, but someone else’s:

As long as you give the source—showing it’s not your work—it’s OK even if you copy a paragraph—sometimes you modify more, other times less—depending on your circumstances. The key is you give source and show it’s others’ work or results, not yours.

S8’s ‘you can copy a paragraph as long as you give the source’ perhaps cannot be taken as what she meant literally, for at least in her text there was no evidence of ‘copying a paragraph’. However, based on our interview data, it seems that S8, as well as quite a few other students, did believe that in writing introductions and discussions it is acceptable (i.e. not chao xi) to copy some stretches from sources, in particular, as long as the sources are cited. However, when presenting some ‘shared’ knowledge, rather than referring to specific findings of particular studies, it may not be necessary to give the source.

Presenting ‘shared’ knowledge

In an alternative case of language re-use in introductions and discussions the sources are often not cited. This frequently happens where the source-absent elements are more or less ‘shared’ knowledge, or general commentary remarks concerning a subject area, rather than specific findings retraceable to particular studies. In Example 3, S3 (Physics) reproduced a whole sentence from a source text:

Example 3

S3’s text (in published form)

Indeed, the cyanide bridge has been shown to be a good mediation for the electronic interaction between two parmagnetic metal ions.
Example 4, featuring the practice of a fellow student of S9’s (Astronomy), is another example of source-absent language re-use in stating ‘shared’ knowledge.

Example 4

The fellow student’s text: the beginning part of the introduction (in published form, 2004)

High mass X-ray binaries (HMXBs) can be roughly divided into Be/X-ray and supergiant binary systems. The X-ray source often is a pulsar, and is powered by accretion of material provided by the optical companion star. In the supergiant systems, this can be achieved through incipient Roche-lobe overflow or stellar wind accretion, while in the Be/X-ray systems, generally, only the latter process occurs because of...

The beginning part of a source (1989)

Massive X-ray binaries containing a neutron star can roughly be divided into two groups: those containing a supergiant (e.g. Vela X-1), and those containing a Be star (such as X Per). The X-ray source often is a pulsar, and is powered by material provided by the optical star. In supergiant systems, this can be through Roche-lobe overflow or stellar wind accretion, while in Be/X-ray systems only the latter process occurs, since...

As a reader of the two texts, S9 felt the two passages were similar—with the similar background information they introduced. But he believed that it is sometimes unavoidable to have similar background introductions:

In the background part of the Introduction, the background necessarily connects to some similar matters. If you’re in the same research field, then the background you want to describe does not have a big difference. And especially when you’re not proposing an entirely new model, then you have to refer to other’s work, because some background you can’t avoid introducing.

More specifically, on the one hand, S9 believed one should avoid using identical language, that is direct copying:

When you write a paper, you can’t copy from another paper; this gives others a bad feeling, a little suspicious of chao xi. Even though copying language [i.e. as opposed to copying research ideas] is not too serious, it’s after all not too good. If it can be avoided, one should try to avoid it.

On the other hand, S9 also believed his fellow student did not commit chao xi, for the fellow student had tried to avoid being ‘identical’ to,
or reproducing, a whole passage word-for-word from the model text, but had used his own words too.

**Methods sections**

We noted in our introduction that St. John (1987) has reported Spanish scientists using a previous methodological description within their home research group. Our students seem to be practising this to a much larger degree than the Spanish scientists (as seen by the textual extracts provided by St. John). S5 (Chemistry), in Example 5, modified a segment of the method section in a published paper first-authored by a lab-mate, without citing the latter:

Example 5

*Electrophoresis procedures’ in S5’s paper (as in his submitted version)*

The running buffer was directly introduced into the reservoirs and flushed through channels under vacuum for several minutes. Then voltages were applied to the separation and injection channels respectively for several minutes until the separation and injection currents leveled off. The buffer in sample reservoir was then replaced by sample solution. The sample was injected into the separation channel using a cross injector.

*Electrophoresis procedures’ in a published paper first-authored by S5’s labmate:

Before separation experiments the running buffer was introduced directly into the reservoirs and then flushed through the channels under vacuum for several minutes until no bubbles remained in the microchannels. Potentials were then applied to the separation and injection channels for several minutes until the separation and injection currents leveled off. The buffer in sample reservoir was then exchanged with sample solution. Injection of sample into the separation channel was achieved by use of a cross injector.

S5 observed that since he and his lab-mate followed similar experimental procedures to those contained in the segment, and since papers published by the home research group were their ‘own work’, it is both legitimate and necessary to refer to the lab-mate’s paper in his own writing. There is no need to cite the lab-mate’s paper for this purpose, because obviously it was not reiterating any findings. Besides, after all, he claimed, the experimental procedures section is not as important as other parts of a paper (see Swales 1990 on this):

In general, only those who want to do work on the basis of your work will give attention to your experimental procedures. There is
no need for much ‘originality’ in writing this part; it is the most ‘plain’ part of a paper, although an indispensable part. This is also a feeling I got from reading and from my submission experience. Referees questioned my Introduction and Results and Discussions [rather than the Experimental section].

A case similar to S5’s is found with S8 (Medicine), who also modified a segment in the Materials and Methods of a paper first-authored by a lab-mate (and likewise, saw no need to cite the latter text). S8’s justification of her copying was:

Our lab has this set of established, successful experimental methods, he [the lab-mate] wrote it into his paper; when I arrived, the lab taught me this set of methods, then I did my experiments. But he has expressed it into words, so when I wrote mine, I referred to his words, with some modification according to my experiments. Even if I hadn’t used his words, we have generally the same methods.

To both S5 and S8 then, their copying from the paper of their home research group was justified and perhaps necessary. Neither felt the need to cite their lab-mates’ articles in writing the particular methods sections, although the lab-mates’ articles are cited elsewhere in their papers.

Results sections

In the results section of scientific articles, there are many tables and figures. Students commonly make use of source texts for the ‘metalanguage’ for labelling, referring to, as well as interpreting, such images. The following examples are from S4 (Physics):

Example 6 (caption to a figure)

S4’s text:

FIG. 4. Permittivity ($\varepsilon_{33}$) vs ac field $E_{AC}$ for different frequencies. (Dash lines are fitted data and symbols indicate experimental data points.)

Source text:

FIG. 7. Permittivity ($\varepsilon_{33}$) vs ac field $E_0$ for different frequencies. (Solid lines are fitted data and symbols indicate experimental data points.)

Example 7 (referring to figure)

S4’s text:

Fig.3(b) shows the applied voltage ($V_{DC}$) dependence of dielectric constant ($\varepsilon$) and loss ($\tan\delta$), respectively, at oscillating voltage ($V_{AC}$) 0.5 V.
Fig. 2 shows applied voltages dependence of dielectric constant \( (\varepsilon) \) and loss \( (\tan \delta) \), respectively, at 0.05 MHz and oscillating voltage 50 mV.

Example 8 (interpreting a figure)

S4’s text: (as in his manuscript)

Figure 2 shows the P–V loops with \( V_{\text{max}} = 20 \) V, where \( V_{\text{max}} \) is the amplitude of the voltage pulses for P–V loops test.

Source text:

Figure 1 shows the P-V loops with \( V_{\text{AMP}} = 519 \) V, where \( V_{\text{AMP}} \) is the amplitude of the voltage pulses for P-V loops test.

S4 explained that such language adoption from the literature was based on the similarity of ‘content’ and that the adoption ensures the correctness of his writing:

It’s mainly according to content. Things like this, the content is similar, adopt it, no problem, and I won’t make a mistake. Although there’s still some difference [pointing to ‘0.5v’ and ‘50mv’ as shown in Example 7], ‘0.5v’ and ‘50mv’ are of a-thousand times different, so you make some change.

S9 (Astronomy), who likes to write into his notebook some expressions from journal articles, recorded two expressions useful for interpreting figures: ‘In figure 1, the observed torque reversal event is reproduced by…’, and ‘The fit shown in Figure 1 corresponds to…’. He apparently appreciates the ‘variety’ of the expressions:

In my paper, I select the first expression and I think the second expression is also a nice way to describe the observation and maybe if you need two different expressions in one paper in order to make others feel variational [sic] as they see the same meaning in many places of a paper while with different expressions. (Email in English)

Example 9, considering its proportion of language re-use—which goes beyond what might be described as ‘metalanguage’—seems somewhat unusual:

Example 9 (interpreting a figure)

S3’s text: (in published version)

The Mössbauer spectrum shows that a doublet and a singlet encounter at room temperature… The singlet appeared because the low spin \( \text{Fe}^{II} \) \( (S = 0) \) ion is situated in a symmetrical octahedral CN ligand coordination sphere. No electrical field gradient (EFG)
exists and hence no QS. However, for the low spin F=III \((S = \frac{1}{2})\) ion, the valence electron contribution to the EFG is non-zero.\[^{[11]}\] Thus a doublet is observed.

Source text: (passages from the source article as indicated by S3’s reference 11)

The ambient temperature $^{57}$Fe Mössbauer spectra of Fe-Cu cyanides annealed at different temperatures were shown in Fig.5. Two doublets and one singlet were encountered. . . . The singlet arose because the low spin Fe$^{II}$ \((S = 0)\) ion is situated in a symmetrical octahedral CN ligand coordination sphere. No electrical field gradient (EFG) exist [sic] and hence no QS \[^{[11–13,14]}\]. Whereas for low spin Fe$^{II}$ \((S = \frac{1}{2})\) ion, due to the half-spin on the cation, the valence electron contribution to the EFG is non-zero, so a doublet is observed.

S3 felt his adoption was fine, since he was reporting some more or less standardized results, in the sense that, as he put it:

> the results are widely accepted—Mössbauer is a tool—that is to say, if there’s such phenomena, then there should be the results; the phenomena and results are uncontroversial and universally accepted.

Insofar as ‘if there’s such phenomena, then there should be the results’, this justification for re-using someone else’s description of results resembles the claim of S5 and S8 (see under the ‘Methods sections’) that their experimental procedures were somewhat routinized or were similar to those of their labmates and it is thus legitimate and necessary to re-use the latter’s description.

**DISCUSSION**

Our findings have shown that student re-use language taken from other sources in all of the sections of the prototypical IMRD paper. This usage varies from short phrases to stretches of sentences in a row. Our participants defend their practices of language re-use with various arguments. While we cannot make any generalizations, we can summarize these arguments as follows. In the introduction and discussion sections one argument in support of copying is that citation of the source mitigates any copying that might occur. Similarly, another argument is that it is acceptable to copy whole sentences in introductions and discussions if they refer to ideas that are common knowledge within the discipline. As regards the methods section, it is necessary to refer to procedures that have been used by others. Given that the experimental procedure is likely to be the same or very similar to what has been done before within the same research group, sharing the same language is not seen as a problem. In addition, copying is not a serious problem in this section, according to another argument, because this is
not an important section and it will receive little attention. Regarding the results section, language is re-used for labelling and describing figures and charts (thereby ensuring accurate language) and for reporting results of uncontroversial and relatively routinized phenomena.

In what follows we will interpret our students’ (novice scientists’) practices and beliefs by first acknowledging the novice’s perspective, which then leads to a discussion of the formualicity of scientific writing and originality in science.

Language re-use: from the students’ perspective

Based on our interaction over several years with students such as those reported in this study, we should still note, if cautiously, that language re-use such as we have described above is likely to be prevalent among Chinese doctoral science students, even though our examples were collected from a small opportunistic sample. Given the prevalence, we find it relevant to draw attention to what Currie (1998:2) has referred to as a ‘tension’ that complicates the discussion of the issue of plagiarism. The tension is that students are both expected to ‘assume the appropriate discourses’ of the disciplinary communities into which they are being socialized and yet to show ‘the required mastery in their own words’ so as not to commit plagiarism. It seems that the participants in this study are subject to this ‘tension’.

Our sample of students, except S4, were still writing their first paper(s) at the time of data collection (see Table 1). Given our sample, it is hard to make any general/clear-cut statements correlating the participants’ year of study and publication experience with their practices/beliefs in language re-use. Nevertheless, it is possible to at least speculate that as they write more papers and become more ‘fluent’ in expressing themselves, they are likely to rely less on others’ texts for language re-use. S8 said, for example, in her second paper, that she already felt she had more ‘fluency’, relying less on the wording of source texts and expressing herself more confidently in her own words. Indeed, if language re-use is an intermediary stage in mastering the target discourse, as argued in the literature, the language re-use of our participants can also be seen as their means of developing an appropriate disciplinary discourse.

Formulaicity in scientific writing: separation of form and content

Returning to the anecdote with which we began this paper, as we noted earlier, the Dutch editor, in a verdict over a question of originality in scientific writing, was assuming a separation between work/content versus language/form. It would seem that both Pan et al. and our students share this assumption: as long as the ‘work’ is their own, the language can be
borrowed, to various extents and for various reasons, from one text into another.

Such a view of the separation of work/content on the one hand and language/form on the other, might indeed be justified to a greater extent in scientific writing than writing in the humanities, say (where constructivism holds sway), by the fact that scientific writing is given to formulaicity. This formulaicity applies all the way up from the word to the phrase and from the individual communicative moves or functions to the overall rhetorical structure. ‘Prototypical syndromes of features’ (Halliday 1993: 54) and ‘IMRD’ (Swales 1990) are descriptions that capture such formulaicity. Given its formulaicity, and at the same time its complexity, scientific English is likely to be both susceptible to imitation/recycling and difficult to master, as our student participants were very much aware. Interestingly, contemporary pedagogy in language education fully acknowledges and exploits the formulaicity of language at these levels: re-use of formulaic structures at the syntactic level and formulaic chunks at the lexical level are basic learning strategies upheld by corpus-based pedagogy, while the formulaicity of various texts at the rhetorical level has been the foundation of genre-based pedagogy.

Referring back to the practices and beliefs of our student participants, if there is no problem re-using formulaic forms of language (as in our Examples 6, 7, and 8), questions nevertheless remain. What about, for example, re-using someone else’s description of background information in writing the introduction (Examples 3 and 4), or long stretches of wording when composing citations (our Examples 1 and 2), or experimental description in the methods section (Example 5), or the reporting of routinized results (Example 9), even if such re-use is not word-for-word copying but with some editing/adjustment. To our student participants (and to Pan et al.) these actions were more or less non-problematic, because the ‘work’ was their own.

Originality in science: more on separation of form and content

In answering the above questions, it is useful to first turn to the attitude of expert members of the scientific community, the novice scientists’ supervisors for example. As reported in Pecorari (2006) and as we learned in our research, the supervisors clearly considered the students’ language re-use (exemplified in this paper) as unacceptable. In other words, these mature scientists (NESs or NNESs themselves) generally hold far more strict criteria over acceptable/unacceptable language re-use in scientific writing. Yet on the other hand, the supervisors we talked to both emphasized the necessity for student writers to learn to write by ‘imitating’ others’ writing, and acknowledged that for the students, ‘imitating’ to an extent that is ‘unacceptable’ may easily occur. One supervisor (a professor of Physics)
figuratively described the situation: ‘It is hard to walk by the river without getting wet’, by which we understand that, if one is to avoid unacceptable language re-use, finding the right balance of language re-use is a difficult task.

Nevertheless, perhaps ‘getting wet’ (i.e. committing textual plagiarism) and ‘originality’ in ‘science’ itself can indeed be separately viewed, even though this does not mean that textual plagiarism is justified/legitimate, ‘as long as the work is our own’. From the fact that some of our ‘textual plagiarism’ examples were taken from the students’ published texts, and indeed one post-doctoral fellow of Chemistry we talked to suggested that it was not rare for him to come across two highly-similar passages in reading journal articles authored by scientists of various L1-background. We may thus venture to suggest that the primary concern of the scientific community in evaluating an article is the ‘originality’ of its ‘science’, rather than its language. Cases of ‘plagiarism’ in science, which have been sensationaly reported (cf. Wang 1999), involved the stealing of other’s ‘work’ (science); while texts containing ‘plagiarized’ passages (e.g. a paper that contains a passage on background information borrowed from another text, with some editing) which do not affect the ‘originality’ of the ‘work’ reported are not the focus of attention—although the post-doctoral fellow of Chemistry referred to above further suggested that the scenario tends to be found in average-level journals (often the target journals of the students) but is rarely identified in prestigious journals.

This view of the separation of work/content on the one hand and language/form on the may be different to the relation of language and fact in the humanities, and where the two go hand in hand, and where language constructs reality (e.g. Bazerman 1994).

CONCLUSION

Given the discussion above in terms of formulaicity and originality in scientific writing, it might be suggested that applying the criteria for ‘plagiarism’, as practised in the humanities, to the natural sciences can be problematic. This, of course, poses a challenge to EAP pedagogy targeted at helping novice scientists writing for publication in English, especially given that EAP practitioners tend to have a background in the humanities/social sciences rather than in the natural sciences. As EAP teachers, while we can help the students to develop their language proficiency in order to avoid ‘textual plagiarism’ by gradually relying less on others’ wording in expressing themselves (e.g. Leki and Carson 1997; Liu 2005), the present study implies that we should go further. On the one hand, we should acknowledge with the students the formulaicity of scientific writing so as to help them to exploit it in their own work, by using a corpus-based pedagogy for example. On the other hand, we should discuss the relationship between ‘form’ (language) and ‘content’ (the work reported)
in the natural sciences (as contrasted with that in the humanities, and preferably with the participation of the students’ disciplinary supervisors), as well as the potential room thus allowed for language re-use in scientific writing, that is ‘walking by the river without getting wet ‘in the words of the physics professor quoted earlier.

Coulthard (2004), in a paper published in this journal, while accepting the significance of Sinclair’s (1991) idiom-principle (i.e. that language is to a large extent formulaic), demonstrated through some informal experiments that the longer a string of words the less likely it is to have been used previously. This would seem to be an important principle for EAP teachers to apply. It can be translated into simple advice along the lines that shorter chunks of re-used language are more likely to be acceptable than longer ones. This also can be demonstrated by means of corpus-based pedagogy. Coulthard’s insight also fits, and indeed may justify, many of the examples of language re-use in this paper, those which are at the level of the phrase.

To summarize and point the way forward for further research, in this study, we examined our sample of Chinese novice scientists’ practices and beliefs in terms of language re-use in the individual sections of the IMRD structure of the genre of scientific articles, leading to the prominence of two issues which we found crucial to understanding language re-use in scientific writing: the formulaicity of scientific writing, on the one hand, and originality in science, on the other. From here, for future research we think it would be worthwhile to proceed along at least two lines. First, to track on a longitudinal dimension novice scientist writers’ development in language re-use over time, to see what changes occur as they increasingly acculturate into their target specialist community as research writers. Second, on a contrastive dimension, to investigate the practices and beliefs of (novice or mature) scholars across disciplines in alternative ESL/EFL contexts, and even those of native English speaking scholars.

Final version received November 2006

ACKNOWLEDGEMENTS

We would like to gratefully acknowledge Margaret Cargill and Vai Ramanathan for valuable comments and suggestions, and the anonymous reviewers and the editors of Applied Linguistics for insightful critiques on an earlier version of this paper.

NOTES

1 The publication requirement varies somewhat across disciplines. For doctoral students in Physics and Chemistry at XU for example, it is understood that the minimum is three first-authored papers in English accepted for publication by SCI journals (preferably overseas journals, rather than those domestic journals indexed by SCI, which are generally much less prestigious than overseas journals).
‘First-author’ means a student writer being listed as the first author of a published/accepted paper. However, the students’ published/accepted manuscripts normally bear the contribution of their supervisors to various degrees in terms of language correction as well as adjustment of content (Burrough-Boenisch, 2003). Even for S4, who described himself as having published ‘five or six papers of relatively high quality, and eleven or twelve papers in total’ during his candidature, his supervisor actually devoted much time to amending his English and presentation, which was crucial to the successful publication of his papers.

In this example, we should mention that apart from S8’s reference [21] being the source article, her references [19] and [20] were taken from the source article, i.e. corresponding to the source article’s references [6] and [5] respectively. This way of adopting references from source articles is apparently common among student writers (e.g. Pecorari 2006), or even, to various extents, among more mature writers. A scenario of some resemblance is recounted by Canagarajah (2003) talking of his experience of lacking access to needed important works from Sri Lanka, but resorting to citing the works anyway, having seen them cited in the accessible articles.

REFERENCES


Dong, Y. R. 1996. ‘Learning how to use citations for knowledge transformation: Non-native


Dubois, B. L. 1988. ‘Citation in biomedical journal articles,’ English for Specific Purposes 7: 181–93.


Li, Y.-Y. 2007. ‘Apprentice scholarly writing in a community of practice: An ‘intraview’ of an NNES graduate student writing a research article,’ TESOL Quarterly 41: 55–79.


