WRITING FOR THE SCIENTIFIC COMMUNITY: THE CHALLENGE OF BEING ORIGINAL UNDER CONSTRAINT.

Vilson J. Leffa, UCPEL

The objective of this paper is to analyze the problems found by graduate students when they write their theses. I am using the American terminology, in which "thesis" is used for the work done at the MA level and "dissertation" is reserved for the doctoral level. This distinction is important because I am interested here in the acculturation process, the baptism of fire, so to speak, that the neophyte writer has to go through to be accepted by the academic community. I envision the MA thesis as an entrance ticket to this community.

In terms of theoretical framework, I will be using, as a background, the concept of discourse community, as proposed by Swales(1990) to which I will add the concept of language as appropriation, as proposed by Bahktin (1993). In terms of corpus I will utilize examples from MA students in applied linguistics and MS students in computational linguistics. One difference between these two groups is that the MA students come from a language teaching background and are expected to be more proficient in the use of language, both English and Portuguese, than the MS group, who come from the Computer Science Department. While most of my examples come from MA theses, I will also use some examples from articles.   In Computer Sciences it is typical for students to produce an article in English based on their theses, which is usually written in Portuguese.

I will try to describe briefly what an academic discourse community is, the obstacles the students have to overcome to be accepted by the community, and, finally, some tentative solutions that, in my opinion, have produced promising results.

For a description of discourse community, I would like to utilize the model proposed by Swales (1990), and complement it with examples from the selected corpus. According to Swales (1990, p. 24-27), a discourse community can be defined by the following six characteristics:

1. *A discourse community has a broadly agreed set of common public goals.*

According to Swales, the basic criterion used to identify a discourse community is not its object of study but its goals. Considering a topic such as language, for example, we know that it can be approached from different points of view, each one implying a different discourse community. If our goal is to describe language in terms of its form, we belong to one community; if we want to describe it in terms of its functions, we belong to another; and so on to always different communities, according to our interests: first language acquisition, phonology, language varieties, language learning, the teaching of writing, reading comprehension, etc.

The problem faced by the student here is that these goals lead to different rules and conventions, so that what is accepted in one community may be totally rejected in another. Eclecticism, to the extent that it may incorporate goals from one discourse community into another, is typically seen as intellectual impurity.

2. *A discourse community has mechanisms of intercommunication among its members*.

Belonging to a discourse community typically involves interacting with other members through intercommunication mechanisms such as meetings, correspondence, newsletters, including telecommunication means such as e-mails, news groups and homepages. In our area, these mechanisms are not locally restricted but globalized, and access to them is a basic prerequisite for the thesis writer.

While the use of electronic means undoubtedly facilitates access to the community it may also represent an extra burden for the student. Internet, more than ever before, has made it obligatory to know one foreign language. MA students who come from a Portuguese teaching background, cannot always be expected to be fluent in the language that has become the lingua franca on the world wide web. Instrumentally, they are at an enormous disadvantage with other students, lame ducks of MA programs, having to compensate their lack of competence in the foreign language through other means, sometimes even paying for translations, or restricting their interests to areas where knowledge of the lingua franca may not be so crucial.

> 3.   *A discourse community uses its participatory mechanisms primarily to provide information and feedback.*

The good news here is that academic communities are interested in disseminating what they do, not only to their own members, but also to other interested people through mechanisms that are open to the public such as conferences, newsletters, journals. Anybody can freely walk into a library and look up any of these sources and even participate in the community as long as he or she has the necessary qualifications — it is typical of academic journals, for example, to offer instructions to the reader on how to contribute to them.

Associations who were traditionally reserved in the admittance of new members, who could only join through the recommendation of effective members, are now opening up. With the advent of Internet, the goals and preoccupations of a discourse community are disseminated as never before, making it much easier for the novice to join these communities. A further consequence of Internet is that many of the conventions that underlie the interaction between members are more explicitly shown, mainly in discussion groups. Many of the hidden assumptions, that are not otherwise open to the neophyte writer, may be exposed on a news thread, although it should be recognized that even here many secrets are still kept from the members of other communities. When we belong to a discourse community we do not realize how strongly we keep the members of other communities from getting access to some of our information. I felt this as I tried to enter the discourse community of computational linguistics. My first impression was that the members of that community talked a lot about something but did not show you the real thing. I found that there are two reasons for not allowing you to have access to it. First the real thing may not be worth showing. Second, the real thing is too valuable and the author does not want to give it away. Reports of important discoveries, especially when they lead to the development of products, are not published with the details that allow you to reproduce the experiment or the product itself.

> The struggle of the student writer is not the struggle to bring out that which is within: it is the struggle to carry out those ritual activities that grant one entrance into a closed society. Or as Foucault would have it, 'The discourse of struggle does not

oppose what is unconscious, it opposes what is secret.' (Bartholomae, 1983, p. 300).

Also, everything that is published in academic journals assumes shared knowledge on the part of the reader, that is, knowledge that is not always possessed by the thesis writer and that is not always easily acquired when specific knowledge is involved, as is the case in the writing of a thesis. It is kind of a vicious circle: you can't read in the area because you don't have the shared knowledge assumed by the writer, and you don't have the shared knowledge because you can't read in the area. Of course the solution is the use of introductory texts, but since we are dealing with specific areas of knowledge, these introductory texts are not always available.

> 4.   *A discourse community utilizes and hence possesses one or more genres in the communicative furtherance of its aims.*

A discourse community usually has a variety of interest groups. In our community of university teachers of English, for example, we have groups interested in ESP, textual production, reading comprehension, computer-assisted language teaching, and so on. Some of these groups can be even subdivided into smaller groups. Within computer-assisted language teaching, for example, we can have subgroups interested in word processing, internet pen friends, or concordancing techniques in language analysis.

> 5.   *In addition to owning genres, a discourse community has acquired some specific lexis.*

The use of specific lexis not only characterizes a given discourse community but also isolates it from the reach of other communities, blocking entrance to those who are not familiar with its jargon. According to Swales (1990, p. 26-27), if an outsider attends a meeting of a discourse community and understands every word, the group has probably not formed a discourse community yet. People outside the language teaching profession have no idea of what we mean by acronyms such as ESP, EAP or EFL. Even within our profession, some of our members may have some difficulty in identifying, for example, what the subgroup concerned with computers in language teaching means by acronyms such as CALL, FTP or NLP.

6. *A discourse community has a threshold level of members with a suitable degree of relevant content and discoursal expertise.*

A discourse community is also characterized by having both newcomers and old timers. I will be using here, for "newcomer", the term "novice", and for
"old timer", the term "specialist". For novices to be regarded as community members they have to demonstrate a minimum level of expertise in the area; the kind of knowledge that is necessary to write an MA thesis, for example. Specialists should be able to demonstrate competence in performing specific tasks such as planning and teaching courses, coordinating events, and advising students in the writing of their theses.
We can define an academic discourse community as a group of producers and consumers of text, who find ways of interacting with each other as efficiently as possible, through a set of conventions, which includes an agreed upon terminology. The conventions basically define things that can and cannot be done, according to a given medium

(opinions can be more overtly expressed in an electronic mail message than in a journal article, for example), and according to the roles played by the participants (a novice cannot act like a specialist). The acculturation process into a discourse community may be a very difficult enterprise. It involves, as we have seen, different kinds of problems: acquiring competence in a foreign language, usually seen as a prerequisite; gaining familiarity with the terminology favored by the community; appropriating the shared knowledge assumed by the members of the community in the content area; incorporating the specific conventions which determine the discourse used in the community, identifying the goals of the community, which may be more theoretically or more practically oriented.

I would like to comment briefly on some of these problems, starting with the issue of linguistic competence. A good command of the language is obviously a prerequisite that has to be obtained long before the writing of the thesis starts. This condition, however, cannot always be assumed specially in graduate programs not specifically devoted to the study of language, as is the case in the computer science department. The good news is that linguistic problems, to the extent that they involve only lexical choice and syntactic construction, are easy to solve, as long as they are not embedded in other more complex problems. The following sentence taken from an early draft written by an MS student can be used as example of a problem that is purely of a linguistic nature and extremely easy to solve:

> The aim of this study is to analyze the way how lexical ambiguity is treated in its syntactic and semantic aspects in the Portuguese.

The problems are easy to correct because they affect only the surface level of discourse. They introduce some noise in communication but the text usually can be understood without difficulty. I found these linguistic problems to be very frequent in published materials, especially in the field of computational linguistics: the following are just some examples produced by native speakers of Japanese writing in English:

Word in English is applied as label to identify Universal Word (UNL, p. 35).

Conventional English-to-Japanese machine translation (MT) systems which are rule-based approaches, [*sic*] are difficult to translate certain types of Associated Press (AP) wire service news stories [*sic*], such as economics and sports, because these topics include many fixed expressions (such as compound words or collocations) which are difficult to be processed by conventional syntactic analysis and/or word selection methods (Katoh and Aizava, 1994, p. 28).

Different language may have more detailed scheme to express aspectual information of an event (UNL, p. 44).

It seems that linguistic problems are not only easily solved but also more easily tolerated, as long as they do not produce too much noise in the communication. Problems are more serious at the

textual level because they affect meaning production and readers tend to get extremely frustrated when they cannot make meaning of a text or feel that they are misled by the writer who force them to follow the wrong cues. This can be demonstrated in the following paragraph:

> It is important to stress that morphological attributes are word endings that serve the purpose of indicating gender, number and person. Therefore in English there are morphological attributes.

There is no space here to transcribe the whole text, but there is nothing in it, before or after the quoted paragraph, that justifies the importance of morphological attributes as indicators of gender, number or person. The conclusion that in English there should be morphological attributes because they indicate gender, number and person is something that baffles the reader. The problem here is not lack of competence in English, but lack of textual competence in general; even if translated into another language, the paragraph would still be problematic.

Problems at the conceptual level may be the most serious of all because they reflect incompetence in the foundational knowledge of a discipline. The following is an example taken from a student who is unable to distinguish between a variable and the value attributed to it, which are basic concepts in any science:

> When I read that Bill bought a car from John, in order to understand the sentence, I have to activate the business transaction schema with at least three variables: Bill, the car and John.

This student unfortunately ignores what both schemata and variables are. Bill, car and John are not variables - such as buyer, merchandise and seller - which, by definition, vary in each instatiation of the schema, but constant values, which do not vary, because they constitute the example itself. Problems with lack of competence in the language, textual development and knowledge of content should not exist of course at the stage the thesis starts to be written, but they do exist and advisors have to deal with them. There are other more serious problems, involving specifically the roles that neophytes are allowed to play in the discourse communities.

My opinion is that students producing MA theses work under more constraints than specialists. They have to know much more than they are allowed to write, because they have to know not only what they are writing but also what they are not allowed to write as novices. This is a challenge for the students because they read what the specialists write but cannot always write like them.

Adapting from Brazil's (1995) terminology, I would like to suggest that specialists are allowed to use a *proclaiming* tone, which emphasizes the importance of what they write, as something that is new, unknown to the reader. Novices, on the other hand, have to use a *referring* tone, emphasizing what is already known to the reader. In each of these tones, a different language is used. The following example, taken at random from one of Chomsky's writings, could be use as an example:

> I think we can also perceive at least the outlines of certain still more general principles, which we might think of as "guidelines," in the sense that they are too vaguely formulated

to merit the term "principles of UG." (Chomsky, 1995, p. 130).

Chomsky, as an specialist in the community, is allowed the privilege of introducing subjective markers in his text such as "I think", a procedure that is typically not allowed for a novice. Chomsky is an acknowledged figure in the community and in his text the subjective markers will be interpreted as an indication of hard-earned reflection, based on a long history of study and research. In a novice's writing, they mean the opposite. The reader will interpret it either as a confession of incompetence or a sign of pretentiousness. The novice does not have the acknowledged historical background of the specialist to provide credibility to what is said.

> Academic writing is at base a ritual activity. And although the details of this social ritual are subject to different discipline requirements in practice, most, if not all, research students encounter difficulties in performing this ritual, and they do so because they are still neophytes, members in training, not yet initiates and certain not masters of the "closed society" (Craswell, xerox, p. 7).

Learning the conventions of the discourse community, not only the general conventions that apply to all members but also the specific conventions that apply to novices is the task of the student who is embarking on the enterprise of writing a thesis. This learning has to go below the surface level of passive proposition recognition and reach the discoursal level where meaning is produced.

The students have to know that it is not enough to know the conventions from the outside; they have to be brought inside and incorporated into their cognitive structure. The conventions cannot just be memorized. In Bakhtin's terms, they have to be assimilated and appropriated, so that they can be submitted to the writer's own intentions. Students as writers cannot just quote the members of the discourse community they want to belong to; they have to be able to rephrase what the other members say in their own terms. They should not only identify the information they read; they have to transform it, incorporate what they receive into what they already know, metamorphosing it — while not extrapolating their role as novices.

These are in my view some of the problems I find in helping graduate students with the writing of their theses. In terms of the text produced by the student, the problems affect the linguistic, textual and conceptual levels. In terms of the process underlying this textual production, there is a social aspect to be mentioned, starting with the interaction between advisor and advisee.

The traditional solution to help the student acquire competence in the rules and conventions of the discourse community is through the mediation of the thesis advisor — a process that usually goes through long and repeated sessions and that is often characterized by inability to express what is meant. In a conference with one of my students, for example, I had marked the following sentence as contradictory:

> Explicit information in a text is not really present in the text, which would make it too long.

I could not understand how something that was explicit in a text was not present in the text. When I told that to my student she was completely surprised

and told me that that was not what she meant. Of course, my ready-made answer for these occasions is "I am not interested in what you meant, I am interested in what you said". In this case, however, she made her point, showing that the sentence could have a second reading, which went more or less like this: "Explicit information is so obvious that it is not necessary to write it down, so it is not present, it is not put in the text. Writing all the explicit information would also make the text too long."

The student produced a text that was coherent according to her own thinking, but not according to the institutionalized and conventionalized expectations of our discourse community, which require the writer to have consideration for the reader, avoiding idiosyncrasies and ambiguous passages.

Misunderstandings go in the other direction too, that is, not only from the student to the advisor but also from the advisor to the student. According to Craswell (xerox), the metalanguage used by the advisor cannot be assumed to be understood by the student. Telling a student that he or she has to improve paragraph structure, maintain unity of focus or signal the line of discussion may be of very little help if the student does not know what the advisor means by that.

In my own experience, when I start with a new student I have to devote some time to adjust the terminology so that we can talk the same language — on the illusion, of course, that we both can say what we mean. I believe we all have already gone through of experience of telling the student to do one thing and get something totally different. If we complain, the answer is invariably the same: "but this is what you told me to do!"

I think the best way to help students in the writing of their theses is to provide them with a community, so that they can socialize what they write, reading each other's papers and exchanging ideas. The great

author as solitary genius is regarded as a myth, even in literature. Postmodernists such as Barthes have proclaimed the death of the author, since authors exist only at the expense of the reader. They do not exist if what they write is not read.

Universities are gregarious places where learning does not occur only through individual reflection but mainly through discussion with others. At least in the places where I work I very rarely see a student working alone. They get together to read papers, prepare seminars, analyze data in an project, and even to write assignments (although my personal feeling is that this gregarious behavior is more frequent with girls). The assumption is that meaning is created between the participants, teachers and students, through continual negotiation; truth is not an individual possession but a commodity that is shared by everybody in the group. As things stand at the moment, the safest thing to do is to create a solidarity community of learners where students support each other. These communities include special training programs, research groups and other collective projects. Their main advantage is that they help the student writer to get the inside status in the writing culture of their own discipline, learning the secrets.

The greatest advantage in such communities is that students learn not only to cope with the constraints of academic writing but also learn to revert the process and use the constraints for their own benefit.

REFERENCES

Bakhtin, Mikhail. *Questões de literatura e estética.* São Paulo: Hucitec, 1993.

Bartholomae, D. Writing assignments: Where writing begins. In: Stock, P. (Ed.), *Forum.* Montclair, NJ: Boynton/Cook.

Brazil, David. *A grammar of speech.* Oxford

University Press, 1995.

Chomsky, Noam. *The minimalist program.* Cambridge, Massachussetts: The MIT Press, 1995.

Craswell, Gail. An interdisciplinary model of textual analysis: discourse conventions in thesis writing. *Handout*. Canberra: Study Skills Centre, Australian National University.

Craswell, Gail. Maintaining textuality: a case study of the problematic use of academic discourse conventions in the thesis text of an international graduate student. *Handout*. Canberra: Study Skills Centre, Australian National University.

Katoh, Naoto and Aizawa, Teruaki. Machine translation of sentences with fixed expressions. *Proceedings of the 4th Conference on Applied Natural Language Processing*, Stuttgart, p. 28-33, 1994.

Swales, John M. *Genre analysis; English in academic and research settings.* Cambridge: University Press, 1990.

UNL. *Universal Networking Language.* Tokyo: Institute of Advanced Studies, The United Nations University, 1996.