ENSURING HIGH SURRENDER VALUE FOR CORPORATE CLIENTS AND INCREASING THE AUTHORITY OF THE LANGUAGE INSTRUCTOR: THE DIVIDENDS OF A DATA-DRIVEN LEXICAL APPROACH TO ESP

Jonathan Clifton
Diana Phillips
University of Antwerp

One source of a language instructor’s authority is his/her knowledge of the target language. However, this paper draws attention to the potential gap between a language instructor’s knowledge of the language in use required by learners from a specialized professional community, and the expectations of stakeholders that the language instructor does indeed have such knowledge. The potential existence of such a gap thus calls into question the instructor’s warrant to teach as “an authority” on the specialized language required by the learners. This paper suggests that a data-driven lexical approach to syllabus and materials design, based on a corpus of naturally occurring data from within the target discourse community itself (TDC), is one way of increasing the authority of the instructor and providing him/her with the knowledge base necessary for the design of English for Special Purposes (ESP) courses with “high surrender value” for corporate clients.

Introduction

A language instructor has two sources of authority. The first is “institutional authority,” which is derived from his/her social position as an instructor. This can be summed up by the expression “being in authority.” However, this kind of authority is often derived from a teacher-pupil relationship typical of schools and is less relevant in the world of in-company language training where the learners are often very skilled, competent, and highly regarded members of their own professional communities. In ESP courses, the instructor has to rely on a second kind of authority which “implies a relationship in which both parties recognize that one of the parties is competent to direct, guide, or instruct the other” (Stevick, 1998: 177). This can be seen as being
"an authority" rather than being "in authority." In order to be "an authority," the instructor must therefore have the linguistic knowledge that the professional community implicitly considers to be essential if the instructor is to be able to effectively instruct the learner. However, being "an authority" is becoming increasingly difficult, as stakeholders in language training are demanding that the instructor has in-depth linguistic knowledge of increasingly specialized domains. Moreover, they are looking for language courses that have "high surrender value." In other words, language training must increase the L2 performance of a company's employees in increasingly specialized fields of professional practice in as short a time span as possible. Consequently, a "long-haul approach" to global language proficiency is no longer economically viable when the primary objective of language training is not learning a new language as such but acquiring the ability to carry out a professional task using a second language as a medium of communication. In short, stakeholders in language courses have a desire for relevance and economy which more readily equates with specialized training needs than traditional approaches to language learning. However, each learner belongs to a specific discourse community which, following Swales (1990), can be glossed as the "particular styles and genres of language" in use by members of a profession. Such discourse communities determine what is shared or given in a particular profession, and if the language instructor does not share such frames of reference, he/she is effectively excluded from the target discourse community (TDC) and therefore lacks the knowledge required to instruct effectively.

This paper suggests one possible way in which language instructors can regain their status as "an authority" in ESP situations and therefore obtain a credible warrant to teach. The solution we propose is a data-driven lexical approach to ESP whereby the instructor's linguistic authority is derived from a language description which is based on a corpus of naturally occurring data drawn from the TDC. The corpus and a subsequent analysis of the most frequently occurring lexical items and their co-text (i.e., set of words which come immediately before or after a word) are thus employed by the instructor as the key resource which can bridge the gap between learners' specialized needs and the instructor's limited knowledge of the language of a particular discourse community. Consequently, working on the premise that the most frequently occurring lexis is the most useful to the practitioner, an instructor can gain in legitimacy as an "authority" on the language required by the TDC.

In this paper, we first briefly discuss the theoretical underpinnings of the lexical approach and outline why it can be an effective tool in fulfilling clients' specialized training needs. Second, we examine the nature of the potential "gap" between an instructor's knowledge and the TDC's language needs and suggest how a data-driven lexical approach based on the use of a corpus of language specific to the TDC can be used as a way of increasing the linguistic authority of the instructor. Finally, we illustrate how a data-driven approach to syllabus and materials design can be implemented by drawing on our own
experience of preparing a syllabus and materials for an international firm of lawyers which uses English as its working language.

The Lexical Approach: Basic Concepts

Anecdotally, a non-English-speaking immigrant to England sat down with a copy of the Bible and a dictionary and started to "learn English." Yet, statistics generated by computational analysis of corpora show that around half of the words in the Bible occur only once. Consequently, the immigrant's attempts to learn English in such a way resulted in a considerable waste of time and effort. The task would have been performed more efficiently had somebody been able to provide her with a list of the most frequently occurring words in that text and the co-text in which they occur. Such an approach to learning would involve a prototypical lexical syllabus which basically sets out to make an inventory of the most useful words and their co-texts for the learner to know. The underlying justification for a lexical syllabus is thus based on the usefulness and power of the most frequently occurring words within a particular discourse community. For example, Willis (1990) points out that the 700 most frequently occurring words in English account for approximately 70 percent of all English text, the most frequent 1,500 words for 76 percent, and the most frequent 2,500 words for 80 percent. Consequently, any general English syllabus that seeks to achieve "efficiency" in terms of exposure to frequently occurring lexis should obviously determine the most frequently occurring lexis first. This applies even more in the case of specialized discourse communities where key "jargon" is characteristic of practitioners' language use in the workplace.

However, a lexical syllabus has much more to offer than just picking out the most frequently occurring lexical items (i.e., head words that one might find in a dictionary entry) and then "teaching" them. Thirty years ago, Becker (1975: 1) wrote:

"I suspect that we speak mostly by stitching together swatches of text that we have heard before; productive processes have the secondary role of adapting the old phrases to the new situation."

Such an approach to language was a departure from the post-Chomskian focus on syntax as the basis of a speaker's internalized language knowledge and the consequent influence of such ideas on language teaching which dominated linguistics at that time. Following Becker, the lexical approach to language teaching argues that language is essentially made up of building blocks of frequently occurring prefabricated lexical chunks which are socially recognized pointers to areas of shared experience. As Lewis (1996: 10) puts it, "the fundamental idea [of the lexical approach] is exceptionally simple—much of our supposedly original language use is, in fact, made up of prefabricated chunks, often, perhaps usually, much larger than single words." The exact
nature of such chunks remains fuzzy and open to debate but following Lewis' (1993: 91 ff.) generally accepted taxonomy, such chunks can be divided into three main categories:

- Polywords, or groups of words that function as single lexical items. Examples of polywords are: of course, by the way, put off, DVD player, etc.
- Collocations, or words that are statistically more likely to appear together than random choice may suggest. So, for example, collocates for "wedding" might be white, cake, or shotgun. Collocations may range from the weak (e.g., a tall man or a loud noise) to strong collocations that are institutionalized and relatively fossilized (e.g., "rancid butter" or "take up a challenge"). One type of relatively strong collocates is, of course, polywords.
- Institutional expressions. Lewis subdivides these into three categories:
  - short grammaticalized utterances such as: not yet, just a moment;
  - sentence heads such as: that's all very well but, I see what you mean;
  - complete sentences that have a readily identifiable pragmatic meaning.

Moreover, such prefabricated chunks seem to provide the raw material for language acquisition. For example, Nattinger and DeCarrico (1992) argue that learning a language (both L1 and L2) involves memorizing chunks of words that frequently occur together in certain predictable situations. If, as Lewis (1997: 41) argues, "lexis helps us handle highly probable events fluently and effortlessly by providing us with prefabricated ways of dealing with them," then the teaching of such lexical chunks has to be a cornerstone of language teaching.

The lexical approach can be an effective pedagogic tool for three main reasons. First, the lexical approach argues that an overemphasis on teaching single (decontextualized) words may hinder the development of an L2 lexicon and deny the learner the possibility of rapid and fluent use of the L2. Simply put, it is more effective for the learner to learn the whole and break it into its parts than to build up to the meaning of a lexical chunk by focusing on the separate parts. Research carried out by Wood (2004) confirms that by helping learners to recognize the lexical chunks that make up language and, more specifically, the lexical chunks that are statistically more likely to appear within the genre used by a TDC, the learner will be able to perform tasks in an L2 more fluently. Second, not only does a lexical approach improve the fluency of language output, but it also improves the learner's receptive skills. Partington (1998), for example, suggests that the lexical approach can help students to recognize lexical chunks in use and thus enables the learner to decode rapid colloquial speech more quickly and effectively. Third, a lexical approach to language instruction can improve accuracy as well as fluency since errors in collocation are particularly frequent amongst language learners. An approach
that specifically targets collocations rather than "words" in isolation can therefore help to address this problematic area of language learning (McAlpine & Myles, 2003: 75).

Moreover, not only is there a substantial amount of research to suggest that the lexical approach is an effective way of instruction, but it has also been convincingly argued that a corpus-driven approach to learning provides samples of "real" English (e.g. Gavioli & Aston, 2001; McNerney & Wilson, 2001). Many commercially available textbooks, on the other hand, teach "a kind of school English which does not seem to exist outside the foreign language classroom" (Mindt, 1996: 232).

The Lexical Approach to Syllabus and Materials Design

The essence of the "problem" in ESP is that the language instructor is rarely a member of the learners' discourse community. Consequently, the instructor is unacquainted with the specialized language games, in the Wittgensteinian sense, that are played by members of the community that he/she is asked to teach and with the specialized lexis that these games entail.

The instructor, as an outsider, cannot perceive the L2 needs of his/her learners because he/she is not "an authority" on the language used by members of the TDC. If the syllabus is thus to have high surrender value and not waste the learners' time with language that is not pertinent to their discourse community, it is widely assumed that it is essential that the instructor first carry out a language audit or needs analysis (e.g., Bosher & Smalkoski, 2002; Graves, 1996; Hutchinson & Waters, 1987; Nunan, 1988). Such a language audit should take the form of "ethnographic" research by collecting as much data concerning the working practices of the learners as possible. This could be achieved by interviewing members of the TDC, observing them in situations where they are already using the target language or examining written documents that the learners have to use or have written themselves in the target language. In this way, the instructor will discover what specialized lexis is in use, what tasks the practitioners have to perform in the L2, what registers need to be mastered, and so on. Prodromou (1997), however, gives a word of warning concerning this issue: if a corpus is based on a high proportion of written, planned language (which is most easily available), it will not necessarily reflect the needs of the learners if they are required to understand or produce spontaneous spoken language. Consequently, the instructor as ethnographer/needs analyzer has the responsibility of ensuring that the corpus correctly reflects the nature of the interaction in which the learners will be or are involved. Thus, by building up a corpus of texts (both spoken and written) that are pertinent to the target discourse community's field of practice, the instructor can take a more objective look at what language is useful to the learner. This ensures that the syllabus and the teaching materials are based on the genres that are in use within a particular discourse community and that they are not based on an intuitive (and often false) grasp of what the syllabus/
materials designer thinks the learner needs. When the syllabus is based upon a corpus of naturally occurring language specific to the TDC, the course will be student-centered and will reflect the needs of practitioners. As Hutchinson and Waters (1987: 19) note, "ESP is not a particular kind of language or methodology, nor does it consist of a particular type of teaching material. Understood properly, it is an approach to language learning which is based on learner need." After the audit, the syllabus designer is able to carry out a discrepancy analysis which "measures" the gap between the linguistic level required to perform specific tasks in an L2 within the TDC and the actual level of performance of the language learners. The gap between the two thus forms the basis of a needs analysis (i.e., what the learners need to know to perform competently in an L2 as members of a particular discourse community).

Once the needs analysis has been carried out, the syllabus can be defined in terms of objectives which fall into three mutually reinforcing and reflexive categories: performance skills, language knowledge, and learning skills. Performance skills are related to the ability to perform certain tasks within a TDC. Language knowledge refers to passive rather than active knowledge of both structural elements (such as lexis, phonology, and grammar) and sociopragmatic elements (such as register and politeness). The learning skills strand of a syllabus should encourage the development of strategies for autonomous learning after the course has ended. For the purpose of this paper, we will concentrate on knowledge objectives and, more specifically, the objective of designing a syllabus and materials that allow the learners to acquire the lexis that is used in the specific genres within their discourse community. However, it is probable that such a lexical syllabus would not be a "stand-alone" syllabus but would be one strand in a multi-syllabus that consists of several organizing principles (such as "grammar," "functions," or "phonology") woven into a coherent unit. Staying with the concept of a lexical syllabus, it is therefore necessary to research the TDC and build up a corpus of the lexis used by community members. Learning objectives will be defined in terms of acquisition of the most frequently occurring lexis.

The next step is to enter the data into a computer and carry out an analysis of the most frequently occurring lexical items within that corpus. To this end, a search is carried out to exclude grammar words (i.e., words such as "and," "but," "if," and so on) from the analysis and leave only content words or lexical items. This ensures that the frequency count will only be made up of lexical items pertinent to the TDC.

Finally, the lexis thus selected by means of a frequency count must be converted into suitable learning materials. The most frequent words will be used as a basis for the course material but it is the material designer's role to ensure that the learners do not focus on the "words" but that their attention is drawn to the co-text that makes up the lexical chunks.

As Fowle (2002) points out, the lexical approach to language learning is often accompanied by the use of vocabulary notebooks which can be used to
raise the learners' awareness of the existence of lexical chunks and encourage autonomous learning skills.

The Lexical Approach to ESP Syllabus and Materials Design Exemplified

In this final section, based on our own experience with an international firm of lawyers, we exemplify how a lexical approach to syllabus and materials design was implemented. The Brussels office of the law firm asked us if we could prepare a team of lawyers before they went to the U.K. to follow an intensive specialist course on English contract law which would use English as a medium of instruction. Since neither of the two instructors/materials writers had any legal training, we had no first-hand knowledge of the linguistic needs of the learners or of the specialized lexis that would be of use to them. Our warrant for being competent instructors in a genre of which both of us had a limited understanding was thus in doubt. We therefore carried out a needs analysis together with the training manager of the law firm. As a result of this analysis and subsequent discrepancy analysis, we devised a multi-syllabus. Some of the needs that we identified in our language audit were addressed by means of traditional teaching methods such as simulations of business meetings and negotiations, coaching of presentation skills, and writing "clinics," but for the purpose of this paper we wish to concentrate on the lexical strand within this multi-syllabus. The training manager stated that the lawyers would be given a textbook as an introduction to English contract law, and this textbook would form the basis of their legal training in England. We decided to use this textbook as the corpus on the basis of which the lexical part of the syllabus would be designed. To address the issues raised in this context by Proctor (1997), we recognized that a complete ESP corpus for our clients would have to include spoken and written legal English as well as a corpus drawn from a wider database than one textbook. However, because of the confidential nature of much of a lawyer's work, we had only limited access to data from sources such as meetings, telephone calls, and company correspondence. A textbook was therefore "safe" material for us to use as a corpus. Secondly, our aim for this part of the syllabus was fairly limited. The material was primarily designed to cover some of the short-term needs of the lawyers who were due to follow the course in contract law in England. Following Flowerdew (1993: 90), who asserts that "the greater specificity of the ESP corpus means that a smaller corpus is more likely to be adequate for ESP applications," we felt that a relatively limited but well-targeted corpus was justified for the lexical strand of the syllabus.

To design the materials, we used commercially available concordancing software (Wordsmith) to find the most frequent lexical items. The ten most frequently occurring items were:
1. contract  2. law  3. breach  4. unlawful  5. defendant
6. liability  7. negligence  8. against  9. liable  10. action

The teaching materials which we designed on the basis of this analysis aimed to cover the most frequently occurring lexical chunks that would facilitate the reading of the textbook and the production of key concepts and ideas from the book during the intensive course in legal English in the U.K. The materials were essentially of two types: consciousness-raising, and practice exercises. Consciousness-raising activities provide an inductive way of increasing the learners’ awareness of prefabricated lexical chunks (Carter, 2003; Johns, 1991). For example, students were asked to underline lexical chunks in sections of the text or, working from the list of the most frequent lexical items in the corpus, to study the co-text in order to find the collocates for these key words. The practice activities were mainly variants of classic “gap-fill” type exercises, as well as matching exercises and cloze texts. Even though “gap fills” are currently out of fashion, they are an ideal way of activating learning because the most frequent words are often used in relatively fixed collocations, and by the simple act of deleting key words, the learners’ attention can be drawn to the surrounding co-text. In short, by focusing on and deleting the most frequent words from short extracts of text drawn from the corpus, the instructor can use them as “bait” for retrieving the co-text that makes up the lexical chunks relevant to the TDC. Similarly, matching exercises that encourage the learners to match the constituent parts of key collocations are also a relatively simple way of promoting a focus on lexical chunks.

In addition, the learners were provided with a commercially available vocabulary notebook (Wordsflo) which encourages students to note the new lexis in a way that corresponds with the principles of a lexical approach.

The computational analysis of the text thus gave the instructors knowledge of the key lexis used by the lawyers and the co-text surrounding these key lexical items. Consequently, they were able to bridge the initial gap caused by their lack of knowledge of legal lexis, and more specifically the legal lexis required by the lawyers of a particular firm of lawyers for a particular course in England. Linguistic authority was thus conferred on the instructors, and the course achieved “high surrender value” in terms of directing the learners to specific lexis in use by their discourse community. In-house company evaluation of the training stressed that the learners had found that the syllabus included “words” that were useful to them and that it had also provided opportunities for them to develop their analytical ability to “notice” concordances. This was recognized as a skill that would continue to be useful to them after the end of the course.

Conclusion

With the increasing specialization within professional communities and the consequent demand for more specialized language training, language in-
structors often do not have sufficient knowledge of the specific language required by their learners. Thus, the instructors often lack the authority that will guarantee them acceptance as credible sources of language knowledge by learners who have a high level of competence within their own community of practice and in a sense know more than their instructors. A lexical approach to ESP constitutes one way of resolving this problem and of increasing the authority of the language instructor. The collection and analysis of a corpus of naturally occurring language that the TDC uses will provide the most frequent lexis in use. By using such lexis as the basis of at least one strand in a syllabus, the language instructor gains status as an authority on the specialized language in use by the TDC and so ensures the "high surrender value" that corporate clients expect.

The Authors

Jonathan Clifton teaches business English and business communication in the Faculty of Applied Economics at the University of Antwerp. His research interests include workplace interaction and the design of authentic teaching materials for business students. He is at present studying for a Ph.D. in Applied Linguistics.

Professor Diana Phillips currently heads the Department of International Business Communication in the Faculty of Applied Economics at the University of Antwerp. Her research interests are lexicography and business communication. She has published two financial dictionaries and books and articles in the fields of business communication and error analysis.

Notes

1 Following Lewis, we will be using the term "lexical chunk" throughout this paper. However, the same phenomenon has been referred to by many different labels including gambits, speech formulae, and lexical phrases.

References


