Specifying the Construct of Academic Vocabulary: Functional and Discursive Perspectives

Moisés Damián Perales Escudero

moisesd@umich.edu

Suggested Citation:

Abstract

Problem Statement: Models of academic vocabulary use need to address discursive and functional linguistics perspectives in order to enhance construct validity and authenticity in academic vocabulary assessment, particularly with regard to the testing of word families.

Purpose of Study: To provide a clearer picture of academic vocabulary use that can inform the design of academic vocabulary testing by enhancing construct validity and authenticity.

Method: A survey of the literature on academic vocabulary has been conducted to present various perspectives and their shortcoming. Then, findings from the functional linguistics and discourse analysis traditions with regard to vocabulary use are discussed. An example of a testing item incorporating functional perspectives on the use of derivatives within a lemma is presented.

Conclusion: Incorporating functional and discursive findings related to the discourse-structuring functions of multiple derivatives of a single lemma may lead to more valid and

authentic academic vocabulary tests and teaching practices.

Keywords: construct validity, academic vocabulary, peripheral lexis, functional linguistics, discourse

Introduction

The testing of foreign languages for academic purposes has become a very important activity around the world. Every year, large numbers of candidates to both undergraduate and graduate programs have their knowledge of a second/foreign language -mostly English- tested in order to ensure that their level of language ability enables them to pursue university-level studies in Anglophone institutions. However, the practical experience of many ESL instructors and professors of other disciplines indicates that high scores in tests such as the TOEFL do not necessarily correlate with students' ability to use language effectively for the particular tasks that their academic activities involve. These perceptions are supported by Spolky's assertion that the need for practicality in the mass administration of commercially successful tests has led to construct validity problems (Spolsky, 1995, p. 133).

As stated by Read (2000), test takers "need to have a thorough knowledge of words that occur frequently in different academic texts" (p. 92). However, for tests of academic vocabulary to be useful in Bachman and Palmer's terms (Bachman and Palmer, 1998), they need to be based on a definition of academic vocabulary (and of the vocabulary ability that it implies) that is theoretically sound. Such soundness can bring about enhanced construct validity and authenticity by ensuring that what is tested corresponds to real language use. This paper aims at showing how systemic-functional linguistics (SFL from now on), schema theory and discourse theory can inform our understanding of the constructs of academic vocabulary. The final section of the paper presents suggestion for further research regarding the nature of academic vocabulary and vocabulary ability, and the manner in which test

takers' cognitive traits interact with test tasks.

Academic Peripheral Lexis

In general, vocabulary researchers have construed three different levels of vocabulary according to their contribution to the realization of different registers: general service vocabulary, academic vocabulary (non-specialized), and technical vocabulary (specialized). The middle level -academic vocabulary- consists of words that occur more frequently in academic texts than in non-academic texts, but do so consistently across different disciplines and discourse-genres (see Carter, 1988 for a definition of discourse-genre) without being field-specific. This middle-level academic vocabulary has received different names in the literature, such as 'sub-technical vocabulary' (Cowan, 1974; in Nation, 2001, p. 187) or 'specialized non-technical lexis.' (Cohen, Glassman et al., 1988; in Nation, 2001, p. 187). A more adequate term for academic vocabulary might be that of peripheral lexis (Jones et al., 1998, p. 262). The use of this term provides a way of avoiding the confusion between non-technical and technical vocabulary.

What constitutes academic peripheral lexis is an elusive concept, yet an important one since this is the kind of lexis that is often most problematic for students of English for Academic Purposes (EAP) (Jones et al.,1998; Read 2000). Three questions regarding the definition of peripheral lexis are important for second language teaching and learning purposes: a) which lexical items are to be classified as peripheral lexis?, b) when can we say that a learner 'knows' any given lexical item that follows under the category of peripheral lexis?, and c) how many such items does a learner need to 'know' in order to participate successfully in the discourse community of his/her field of study? Before proceeding to the discussion of academic vocabulary, the concept of discourse community and its implications for learning/teaching vocabulary are addressed below.

The term "discourse community" has been used to refer to groups of academics working in a specific discipline. Swales (1990) famously defined the features of discourse communities. According to Swales, a discourse community:

1. has a broadly agreed set of common public goals.

2. has mechanisms of intercommunication among its members.

3. uses its participatory mechanisms primarily to provide information and feedback.

4. utilizes and hence possesses one or more genres in the communicative furtherance of its aims.

5. in addition to owning genres, it has acquired some specific lexis http://en.wikipedia.org/wiki/Lexis_%28linguistics%29.

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

It is point number 5, the use of academic lexis that is of concern here. Specifically, this paper concerns itself with specifying what counts as academic, but not discipline-specific, lexis, or peripheral lexis, and how its use by academic writers contributes to their successful participation in an academic discourse community. The paper discusses the relevance of this kind of lexis in academic writing and considers its implications for enhancing construct validity and authenticity in assessment.

Academic peripheral lexis does not fall into any one specific grammatical category (Jones et al., 1989, 263). Relational verbs, nouns, and adjectives can be classified as peripheral lexis. Interestingly, the nouns presented as examples by Jones et al. belong to the category of discourse nouns identified by Carter (Carter, 1988), which are nouns that perform a discourse function by referring to other parts of the text, usually in an anaphoric manner. Different attempts have been made to arrive at a comprehensive list of peripheral lexis words,

such as Praninska's American University Word List (Praninskas, 1972 in Nation, 2001) , Farrell's semitechnical vocabulary list (Farrell, 1990 in Nation, 2001), Salager's Fundamental Medical English list (Salager, 1984), Xue and Nation's University Word List (UWL) (Xue & Nation, 1984 in Nation, 2001), Coxhead's Academic Word List (AWL) (Coxhead, 1998 in Nation, 2001), and Coxhead's New Academic Word List (Coxhead, 2000). These lists are drawn from corpora of academic texts, either from a single discipline (Farrell's and Salager's) or from a variety of disciplines (Praninskas, 1972; Xue & Nation, 1984; Coxhead, 2000). Despite the fact that different word selection criteria and corpus-design principles were used for each of these lists (Coxhead, 2000, 217), there is "substantial overlap" between them (Nation 2001, 193), which points at the existence of an academic peripheral lexicon.

A major characteristic of those lists is that they group words into families or lemmas in order to solve the problem of what to count as a word. Thus, the UWL contains over 800 word families and the AWL contains 570 word families. Underlying this approach to establishing the size of the peripheral vocabulary is the assumption that knowledge of a base word or lexeme can greatly facilitate the comprehension of its derivatives (Coxhead, 2000, p. 218), as well as the notion that word families are an important organizational unit of the mental lexicon (Coxhead, 2000, p. 218).

Considering word families as the constitutive units of peripheral lexis poses several problems for the testing of both depth and breadth of word knowledge for this type of lexical items. Considering only word families in lists of peripheral lexis seems to focus on knowledge as representation (Bialystok & Sharwood Smith 1985 in Gass & Selinker 2001). According to Tyler (1989, 444 in Gass & Selinker 2001, p. 376) "the representation of a word cannot contain all the various and subtle interpretations that the word could have in different real-world contexts." It follows that representational knowledge is not sufficient if learners are to comprehend and understand words in a way that approximates that of competent

speakers (Gass & Selinker, 2001). Schmitt and Zimmerman's finding that knowledge of any given stem does not amount to knowledge of its derivatives, especially at the productive level (Schmitt & Zimmerman, 2002) supports this claim. Furthermore, there is evidence suggesting that competent speakers do not associate words based on morphological/phonological factors of the type involved in word family construction, but on semantic (paradigmatic and syntagmatic) factors (Meara, 1978; Schmitt & Meara, 1997; in Gass & Selinker, 2001, 378). In addition, the type of vocabulary lists mentioned above fail to take into account both single-word general service lexical items and multi-word lexical items that occur frequently in academic registers and perform particular functions within them. Further exploration of the nature of peripheral academic lexis, word knowledge and vocabulary ability is therefore crucial to arrive at a better specification of the construct of peripheral lexis.

Defining what it means to know a lexical item is a problem that has been tackled by many researchers (Richards, 1976; Nation, 1990; Bogaards, 2000; Chapelle, 1994). All of them include considerations of morphology and syntax as part of word knowledge, but only Bogaards (2000), and Chapelle (Chapelle, 1994) add a dimension of discourse/context to word knowledge. Chapelle's work is particularly relevant for vocabulary testing, as it distinguishes between the construct of vocabulary knowledge and that of vocabulary ability. Vocabulary use in real communication tasks is therefore a combination of both vocabulary knowledge and ability, with the latter being the mediating force between the mental lexicon and the target language use situation. Test designers need to take into consideration that both constructs are to be measured if increased validity and authenticity (i.e. closer mimicking of target language use situations) are sought.

It follows that theoretical positions defining lexical knowledge as independent and separate from vocabulary ability, such as the one advocated by Nation (Nation 2001), are not fit for approaching context-dependent vocabulary testing. In contrast, Chapelle's

interactionalist model of vocabulary ability (Chapelle, 1994) provides a useful framework for approaching this kind of testing. She considers three components of language ability, namely context of vocabulary use, vocabulary knowledge and fundamental processes, and metacognitive strategies. These components are not independent. Rather, they interact with one another and overlap in actual vocabulary use.

Chapelle (1994) operationalizes context of vocabulary using the SFL constructs of field (what is being talked about), tenor (the interpersonal relationships between discursive participants), and mode (the role language plays in the construction of discourse). Along the same SFL guidelines, several researchers in the field of science education have explored the role played by vocabulary in the construction of academic registers (Jones et al., 1989; Halliday & Martin, 1993; Martin, 1993; Celce-Murcia, 2002; Schlepegrell, 2002; Gee, 2002). Underlying their work is a common social semiotics view of language as a meaning-making device that enacts different social functions depending on the lexicogrammatical choices made by speakers/writers. In this view, different types of texts are realized by different registers, a register being "the constellation of lexical and grammatical features that realizes a particular situational context." (Halliday & Hasan, 1989; in Colombi & Schleppegrell, 2002, p. 9) In the paragraphs below, examples of some of those researchers' findings will be presented and discussed regarding their relevance to the definition and testing of academic peripheral lexis.

One of the shortcomings of present operationalizations of academic peripheral lexis (i.e. academic word lists) is their lack of attention to the situated meanings that general service words can acquire in academic registers. Gee (2002, p. 165) gives an example that illustrates how the words *name* and *describe* (none of which appear in Coxhead's AWL) help a student in the mapping of meaning from a colloquial to an academic register. Gee presents the following excerpts from two drafts of the same paper written by a high school student:

First Draft:

Then to let people know there are different types of Albinism, I will tell and explain all this.

Second Draft:

*Finally, to let people know there are different types of Albinism, I will **name** and **describe** several.*

In these two examples, Gee shows how general service words such as "name" and "describe" can acquire specific meanings and functions that assist the realization of academic registers. That is, they could be considered part of peripheral lexis. Gee emphasizes that, in order to use these words in a way that's consistent with an academic register, the student "needs to have experienced certain sorts of acts of classification within certain sorts of Discourses" (106) which is what allows the student to use words with their situated meanings.

Lexical choices are also known to shape writers' syntactic choices (Winter, 1977 in Carter & McCarthy, 1988, p. 207), which has an effect on their communicative effectiveness (i.e. how effectively they convey meaning within according to the register requirements/conventions of the specific genre and discourse community they participate in). In another study, Schleppegrell (2002) analyzed the presentation of assumptions in the theory sections of four lab reports of a Chemical Engineering course, one of which was written by register-competent native speaker of English (the model report). The other three were written by ESL students (Writers 1, 2, and 3). After analyzing and comparing the four reports, Schleppegrell found that the writer of the model report was able to use all the derivatives of the lexeme *"assume"* in a way that enabled her to create a coherent report (the SFL variable of mode), to project an authoritative stance that engaged readers interpersonally (tenor) and present a well-organized statement. Examples of phrases from the model report with this

lexeme are below.

> *The accumulation term can be neglected if **one assumes** a quasy-steady state condition.*
>
> *The error introduced by this simplification **is assumed** to be negligible.*
>
> *An order of magnitude analysis will show that this is **a valid assumption***
>
> ***Assuming** equation 4 holds, the solution of equation 3…*
>
> *B represents **the assumed** stagnant air.*
>
> *Here, **it was also assumed** that Na>Nb. This assumption depends on…*

By contrast, the ESL writers were not able to control the derivative of "assume." Each of them only used two derivatives. For example, Writer 1 listed her assumptions using the imperative form of "assume":

> ***Assume** the diffusion occurs at quasy-steady state; and the concentration is zero at the top diffusion tube…*

Writer 2 introduced the experiments' assumptions using the noun "assumption":

> *There were a lot of **assumptions** associated with this experiment which could cause some discrepancy in the final result.*

Writer 3 used the participal form "assumed" in a passive voice construction, but did not control the passive construction well.

> *An important analysis of Stefan diffusion tube **is assumed that** the diffusion process occurred at a quasi-steady state.*

By manipulating derivates of the lexeme "assume," the native speaker was able to create coherent, register-congruent texts. In contrast, the other reports displayed "no lexical variation in the way they present assumptions, using only the word *'assume'* in one of its forms to state assumptions," (Schleppegrell, 2002, p. 127) which resulted in impoverished communicative effectiveness as defined at the beginning of the paragraph. Thus, Schleppegrell's findings provide a way of operationalizing Chapelle's vocabulary ability

model in that they show how the context of vocabulary use intersects with vocabulary knowledge in real texts. These findings along with those of Gee have implications for the estimation of academic vocabulary size as it will be discussed in the following paragraph.

Schleppegrell's analysis highlights the importance of controlling different derivatives within a lemma. By contrast, current indicators of academic peripheral vocabulary size take into account word families only. However, in the light of the fact that different tokens of the same lemma might perform different discourse functions and even convey different meanings, the size of the lexicon as measured by word families might not truly reflect a learner's ability to use the lexical items being measured for real communicative purposes. Furthermore, Gee's suggestion that general service words also play a role in the realization of academic registers also bears on the estimation of academic peripheral vocabulary size.

The social semiotics approach to construing vocabulary knowledge/ability reviewed above can be complemented by interactive models of schema theory in order to arrive at a more comprehensive interactionalist model of discrete, context-dependent vocabulary testing that could result in enhanced validity and authenticity. Interactive models see text comprehension as "the outcome of generating hypotheses [about lexical choices] and confirming or disconfirming these hypotheses by resorting to what exists in the texts." (Khodadady & Herriman, 2000, p. 205). Khodadady and Herriman applied this notion to the construction of schema-based multiple choice cloze tests in which the distracters all shared some semantic features with the correct response. This allowed for the testing of depth of word knowledge and fundamental processes (i.e. drawing the required meaning from the context). A testing item of this kind could be constructed by, for example, requiring learners to choose from different derivatives of "assume" to complete a statement. For example:

1) _____ that equation four is true, the solution of equation 3 is also true.

   a) Assume

b) Assuming

c) The assumption

d) Assumed

This kind of item would require the test-taker to comprehend not only the semantics but also the syntactic restrictions and discourse-structuring functions governing the choice of a derivative of "assume." This item would then test knowledge of the derivatives' functions, which also involves testing depth of world knowledge and fundamental processes, or the ability to draw the required meaning from context). By reflecting real-world vocabulary used, the item would enhance construct validity and authenticity.

Conclusions

As seen above, the constructs of academic vocabulary and vocabulary ability need to be further understood. The integration of social semiotics approaches (SFL and discourse theory) with schema theory can provide an interesting research avenue to explaining how test takers and test tasks interact in the production of academic peripheral lexis. For instance, to what extent does the linguistic input (register) presented in tests contribute to the engagement of a certain discourse domain and the activation of schemata? Is the process of answering a context-dependent vocabulary test a matter of choosing different words within the same schema, with the register being the governing factor for the hypotheses formulate by the test taker? Exploring these issues is interesting in its own right, but such exploration is also necessary if we wish to obtain more accurate and informed inferences from vocabulary tests.

Furthermore, ways of incorporating the theory to actual testing practices in order to increase the usefulness of vocabulary tests need to be explored. Several attempts have been made to address discourse competence via the use of adapted cloze tests (Alderson, 1979 in Read, 2000; Jongsma, 1980 in Carter 1988; Deyes, 1984; Bensoussan and Ramraz, 1984 in Read, 2000; Carter, 1988; Singleton and Little, 1991). I suggest that the cloze test has the

potential to be adapted to integrate both trait (i.e. schema theory) and textual considerations (i.e. SFL, discourse theory) into useful vocabulary tests that are nonetheless practical.

Finally, the construct of peripheral academic lexis needs to be expanded to include multi-word lexical items. Recent corpus-based studies (Gledhill, 2000; Luzón Marco, 2000; Cortés, 2002) show that certain items such as collocational frameworks and lexical bundles occur idiosyncratically in certain registers and discourse-genres within those registers, i.e. some collocational frameworks tend to occur in the methods section of research articles, whereas others occur more frequently in the discussion or conclusions sections. The ability and knowledge constructs underlying the use of these lexical items need to be addressed, as do potential ways of testing them.

Although it would be naïve to assume that principled modifications of vocabulary teaching/learning methods and tests would automatically translate in positive backwash (Spolsky, 1995), the potential of such modifications to generate that kind of backwash cannot be overlooked. Therefore, efforts should be made to bring theory and practice together for the benefit of L2 language learners and users of L2 tests.

References

Alderson, J.C. (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly 13,* 219-227.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: OUP.

Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second language acquisition. *Applied Linguistics, 6,* 101-117.

Bensoussan, M. & Ramraz, R. (1984). Testing EFL reading comprehension using a multiple-choice rational cloze. *Modern Language Journal 68,* 230-239.

Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: the case of the Euralex French Tests. *Applied Linguistics 21* (4), 490-516. Oxford: OUP.

Carter, R. (1988). Vocabulary, cloze and discourse: an applied linguistics view. In R. Carter & M. McCarthy (Eds.) *Vocabulary and language teaching*, (pp. 161-180). New York: Longman.

Carter, R. & McCarthy, M. (1988). Lexis and discourse: Vocabulary in use. In R. Carter & M. McCarthy (Eds.) *Vocabulary and language teaching*, (pp. 201-220). New York: Longman.

Chapelle, C A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research 10, 2*, 157-187.

Chapelle, C. A. (1998). Constructing definition and validity inquiry in SLA. In L.F. Bachman & A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*.

Celce-Murcia, M. (2002). On the use of selected grammatical features in academic writing. In M.J. Schleppegrell and M.C. Colombi (Eds.), *Developing advanced literacy in first and second languages*, (pp. 143-158). Mahwah, NJ: Lawrence Erlbaum

Cohen, A.D., Glassman, H., et al. (1988). Reading English for specialized purposes: Discourse analysis and the use of student informants. In P. Carrell, J. Devine & D.E. Eskey (Eds.), *Interactive Approaches to Second Language Reading,* (pp. 152-167). Cambridge: Cambridge University Press.

Cortés, V. S. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23* (in press).

Cowan, J.R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly, 8,* 389-400.

Coxhead, A. (1998). An academic word list. *Occasional Publication Number 18, LALS,* Victoria University of Wellington, New Zealand.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213-235.

Christie, F. (1999). Genre theory and ESL teaching. *TESOL Quarterly, 33*, 759-763.

Deyes, T. (1984). Towards an authentic discourse cloze. *Applied Linguistics 5* (2)*,* 128-137.

Farrell, P. (1990). Vocabulary in ESP: a lexical analysis of the English of electronics and a study of semi-technical vocabulary. *CLCS Occasional Paper no. 25.* Trinity College.

Gass, S. & Selinker, L. (2001). *Second language acquisition: An introductory course.* Mahwah, NJ: Lawrence Erlbaum.

Gee, J.P. (2002). Literacies, identities and discourse. In M.J. Schleppegrell & M.C. Colombi, (Eds.), *Developing advanced literacy in first and second languages*, (pp. 159-176). Mahwah, NJ: Lawrence Erlbaum

Gledhill, C. (2000). The discourse function of collocation in research articles. *English for Specific Purposes, 19*, 115-135.

Halliday, M.A.K., & Martin, J.R. (1993). *Writing science: Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Jones, J., Gollin, S., Drury, H., & Economou, D. (1989). Systemic-functional linguistics and

its application to the TESOL curriculum. In R. Hasan and J.R. Martin (Eds.), *Language development: Learning language, learning culture*, (pp. 257-328). Norwood, NJ: Ablex.

Kobayashi, M. (2002). Cloze tests revisited: exploring item characteristics with special attention to scoring methods. *The Modern Language Journal*, 86, 571-586.

Khodadady, E. & Herriman, M. (2000). Schema theory and selected response item tests: From theory to practice. In A.J. Kunnan (Ed.), *Fairness and validation in language assessment: selected papers from the 19th Language Testing Research Colloquium*, (pp. 201-224). Cambridge: Cambridge University Press.

Lemke, J.L. (1990). *Talking science: language, learning and values*. Norwood, NJ: Ablex.

Luzon Marco, M.J. (2000). Collocational frameworks in medical research papers: a genre based study. *Engish for Specific Purposes, 19*, 63-86.

Martin, J.R. (1993). Life as a noun. In M.A.K. Halliday & J.R. Martin (Eds.), *Writing science: Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Meara, P. (1978). Learner's word associations in French. *Interlanguage Studies Bulletin, 43*, 192-211.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Praninskas, J. (1972). *American University word list.* London: Longman.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing 18, 1*, 1-32.

Richards, J.C. (1976). The role of vocabulary teaching. *TESOL Quarterly 10,* 77-89.

Salager, F. (1984). The lexis of fundamental medical English: classificatory framework and rhetorical function. *Reading in a Foreign Language, 1* (1), 54-64.

Schleppegrell, M.J. (2002). Challenges of the science register for ESL students: Errors and meaning-making. In M.J. Schleppegrell & M.C. Colombi, (Eds.), *Developing advanced literacy in first and second languages*, (pp. 159-176). Mahwah, NJ: Lawrence Erlbaum

Schmitt, N. & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition, 19,* 17-36.

Schmitt, N. & Zimmermann, C.B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36,* 145-171.

Singleton, D. & Little, D. (1991). The second language lexicon: Some evidence from university-level learners of French and German. *Second Language Research 7* (1), 61-81.

Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.

Tyler, L. (1989). The role of lexical representations in language comprehension. In W. Marslen-Wilson (Ed.), Lexical representation and process, (pp. 439-462). Cambridge, MA: MIT Press.

Winter, E.O. (1977). A clause-relational approach to English texts: a study of some predictive lexical items in written discourse. *Instructional Science 6* (1), 1-92.

Xue, G. & Nation, I.S.P. (1984). A university word list. *Language Learning and Communication 3* (2). 215-229.

Moisés Damián Perales Escudero is an Associate Professor of Language and Education at the Universidad de Quintana Roo, Mexico. He holds a PhD in English and Education by the University of Michigan. His research interests are in the areas of academic literacies, discourse analysis, and corpus linguistics.