APPLYING THE FRAMEWORK FOR SECOND LANGUAGE VOCABULARY ASSESSMENT TO WORDSMITH TOOLS¹

Soraya Moreno Espinosa Almudena Fernández Fontecha María Pilar Agustín Llach Universidad de La Rioja

RESUMEN. Nuestro objetivo es proporcionar tanto al profesorado de inglés como lengua extranjera como a los investigadores, una exhaustiva descripción de *WordSmith Tools* (SCOTT 1997), un instrumento tecnológico que proporciona una nueva visión del comportamiento de las palabras en el texto, con el fin de que puedan considerar si dicha herramienta encaja dentro de sus objetivos tanto pedagógicos como de evaluación .

PALABRAS CLAVE. Evaluación, vocabulario, aprendizaje y enseñanza lenguas extranjeras, tecnología.

ABSTRACT. The aim of this paper is providing ESL/EFL teachers and researchers with a thorough description through description of *WordSmith Tools* (Scott 1997), a technological instrument which provides insight on how words behave in texts. Despite being a legitimate tool for teaching, learning and researcvh, employed in different studies as an assessment tool, we find a lack of description of it on the basis of the relationship between its purpose and design; data which reports relevant information for ESL/EFL teachers and researchers, so that they will be able to consider whether it really fits their pedagogical and/or assessments goals.

KEY WORDS. Assessment, L2 embedded vocabulary, teaching and learning EFL, technological tool.

1. INTRODUCCIÓN

WordSmith Tools (M. SCOTT 1997) is a technological instrument which provides insight on how words behave in texts. It seems to be a legitimate tool for teaching, learning and research, which has either been used as an assessing instrument or as a corpus analyser.

Despite being employed in different studies as an assessment tool, we find a lack of description² of it on the basis of the relationship between its purpose and design; data which

¹ This paper was supported by the Spanish Ministry of Science and Technology (Grant no. BFF2003-04009-CO2-02), University of La Rioja, and Comunidad Autónoma de La Rioja (Grant F.P.I. 2005).

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

reports relevant information for ESL/EFL teachers, so that they will be able to consider whether it really fits their pedagogical and/or assessment goals. Thus, we consider relevant to present a thorough description of this technological tool on the basis of J. READ and C. CHAPELLE (2001)'s framework for second language vocabulary assessment.

Our paper will be structured as follows: firstly, we will provide a brief review of studies that have employed *WordSmith Tools*, so as to state the gap we aim to fill in; secondly, we will present our framework of reference; and subsequently, we will describe this technological instrument, by taking into account which inferences can be drawn from its outcome, in order to get a rounded picture of testees' lexical profile.

2. REVIEW OF STUDIES

As we have already pointed out, in this section we aim to highlight different studies which have made use of this electronic analyser in order to provide different insights into lexis: (a) as a teaching instrument; (b) as a corpus comparison tool; and (c) as an assessing instrument.

The use of concordancers as teaching instruments has been highlighted by different scholars (S. MURISON-BOWIE 1993, S. GRANGER and C. TRIBBLE 1998, J. WEBER 2001, C. TRIBBLE 2004, B. DAVIS and L. RUSELL-PINSON 2004, J. MUKHERJEE 2004, J. OSBORNE 2004), since it enables teachers to come up with real examples of language use, to compare native speaker language use as opposed to non-native speakers' and to analyse learner corpora for error analysis, among other things. If the most common errors that learners produce are detected, teachers can devise different learning activities to repair such deficiencies; moreover it can be used by learners to raise language awareness.

There are a wide range of studies which have employed this instrument as a corpus comparison tool from a wide variety of perspectives such as those that have investigated language use in speech (S. DE COOK *et al* 1998, A. O'KEEFFE 2004, W. CHENG 2004), language use in writing (M. BONDI 2004, C. SAMSON 2004), language change (B. DAVIS and L. RUSELL-PINSON 2004), and L2 learners' writing such as word collocations (B. ALTENBERG and S. GRANGER 2001), among other.

Regarding its use as a lexical measure, in conjunction with other instruments, we would like to comment on the following investigations:

• Y. LI (2000): This scholar examines the relationship between: (a) objective computerised text analysis by making use of *WordSmith Tools*; and (b) subjective evaluation performed by human raters. She analyses 132 writing samples produced by ESL undergraduate students at an American university, in order to assess the written texts in terms of syntactic complexity, lexical complexity and grammatical accuracy; however, she does not specify neither the level, the age, nor the mother tongue of her sample of informants. Y. Li finds that there is no significant correlation at the level of measuring lexis by computer or by hand, whereas there seems to be a well-founded one with regard to grammar.

² For a description of *WordSmith Tools*' software capabilities compared against other retrieval tools used for text analysis, see P. RAYSON (2002); and for a comparison against MonoConc Pro, see R. REPPEN (2001).

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

- In our view, Y. Li's study presents a number of pitfalls such as: (a) we do not know whether informants represent a homogeneous entity, and whether the results obtained would have been the same with a different sample of testees; (b) with regard to lexis, the different tools employed one manual and the other technological (*WordSmith Tools*) were measuring the same elements, but from different viewpoints; on the one hand, the computer analysis was quantitative in nature, whereas on the other, human raters carried out a qualitative assessment. Therefore, we believe that her analysis should be reviewed by undertaking more research within this field on the basis of a much more uniform assessment.
- T. NAVES NOGUÉS (2001): This author examines: (a) whether two computerised tagged text-analyses of linguistic features of L2 writing (being one of them retrieved by handling *WordSmith Tools*) correlate; and (b) whether there is any correlation between those text analysers, with manually calculated writing measures. She explores 35 writing samples produced by Spanish learners of English as a Foreign Language, who were enrolled in secondary education, and she concludes that: (i) there is a significant correlation between the two computerised analysis; and (ii) no significant correlation seems to be found between analytical measures automatically scored by computer programs and manual band scale scoring.
- P. SAGASTA ERRASTI (2000): This scholar makes use of *WordSmith Tools* in order to assess the lexical complexity of a sample of written texts in English, Spanish and Euskara.
- P. DE HAAN and K. VAN ESCH (2004): They use *WordSmith Tools* for a general analysis of essays written by Dutch learners of English of three different levels of proficiency in order to report a quantitative analysis of word length, type/token ratio and other lexical features to highlight the relationship between certain linguistic features and proficiency levels.

By reviewing the existing literature, we can conclude that even though this technological instrument has already been used within teaching and testing, there is a lack of description of it as a lexical measure. Besides, we would like to stress our agreement with P. NATION (1990), when he remarks that before testing, it should be clear why learners are being tested and what it will be the use of the information retrieved from the testing procedure. Thus, by bearing all this in mind, we consider necessary to give a thorough account of *WordSmith Tools* by taking into account Read and Chapelle's framework.

3. A FRAMEWORK FOR L2 VOCABULARY ASSESSMENT

In this section, we aim to present briefly J. READ and C. CHAPELLE (2001)'s framework for L2 vocabulary assessment. A framework that we will take as a point of reference in our description of the practical value and construct definition of *WordSmith Tools*. This framework refers to J. READ (2000)'s three dimensions of vocabulary assessment. Thus, before focusing on the framework, we would like to start by making reference to J. READ

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

S. MORENO ESPINOSA, A. FERNÁNDEZ FONTECHA Y Mª. P. AGUSTÍN LLACH

(2000)'s dimensions, which form part of a continuum: (a) vocabulary can be assessed either as a discrete or as an embedded element within a larger construct; (b) vocabulary assessment can be selective or comprehensive; and (c) vocabulary can be measured as a context-dependent or as a context-independent element (see figure 1).

Discrete	Embedded
A measure of vocabulary	A measure of vocabulary which
knowledge or use as	forms part of the assessment of
an independent construct	some other, larger construct
Selective A measure in which specific vocabulary items are the focus of the assessment	Comprehensive A measure which takes account of the whole vocabulary content of the material (reading/listening tasks) or the test- taker's response (writing/speaking tasks)
Context-independent	Context-dependent
A vocabulary measure in which	A vocabulary measure which assesses the
the test-taker can produce	test-taker's ability to take account of
the expected response	contextual information in order to
without referring to any context	produce the expected response

Figure 1. Dimension of vocabulary assessment (J. READ 2000: 9)

Specifically, J. READ and C. CHAPELLE (2001)'s framework for L2 vocabulary assessment, aims to specify the relationship between test purpose and test design. Thus, they claim that there are a number of issues that should be regarded when developing any L2 vocabulary test such as: (a) Test purpose and validity considerations, by taking into account: (a1) Inferences to be drawn from test performance at various levels, for instance: item level, sub-test level and/or whole test level; (a2) uses of test results such as instruction, research and/or evaluation of programs amongst others and (a3) impacts that the test is intended to have and actual consequences; (b) Mediating factors which can be: (b1) construct definition; (b2) the way in which performance should be reported and interpreted and (b3) test presentation; and (c) implications for test design and validation (see figure 2).

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

806

TEST PURPOSE	Inferences	Uses	Intended impacts
VALIDITY CONSIDERATIONS	Construct validity	Relevance and utility	Actual consequences
MEDIATING FACTORS	Construct definition	Performance summary and Reporting	Test presentation
TEST DESIGN	Decisions about the dimensions Discrete -Embedded Selective - Comprehensive Context-independent - context dependent		
VALIDATION	Arguments based on theory, evidence and consequences		

Figure 2. A framework for L2 vocabulary testing (J. READ and C. CHAPELLE 2001:10)

We do agree with them, in the sense that there are a wide range of vocabulary tests, used for instructional and research purposes which lack a comprehensive basis for carrying out their evaluation; being this, one of their main reasons for putting forward their framework. Moreover, they also deal briefly with the specification of the purposes of eight exemplary tests³ and its implications for test design. However, none of their exemplary assessing instruments are concordancers, that is why, we consider that our description is largely justified, since on the basis of their framework, there is a lack of description of *WordSmith Tools*, which would enable ESL/EFL teachers to decide whether it really fits their goals.

4. DESCRIPTION OF *WORDSMITH TOOLS* ON THE BASIS OF OUR FRAMEWORK OF REFERENCE

In this section, we will specify the relationship between test purpose and test design. We would like to highlight the fact that we have adapted J. Read and C. Chapelle's framework for second language vocabulary assessment, in order to fit into our tool's characteristics. Thus, we will deal with the main following points: Assessing instrument purpose and validity considerations; and its Mediating factors.

³ It should be noted that the exemplary tests are: *Vocabulary Levels Test, Lexical Frequency Profile, ESL Composition Profile*, TOEFL vocabulary items, multiple-choice cloze, c-test, *Vocabulary Knowledge Scale* and lexical density index.

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

4.1. TEST PURPOSE AND VALIDITY CONSIDERATIONS

According to J. Read and C. Chapelle's framework for second language vocabulary assessment, test purpose consists of three components: Inferences and construct validity; uses of the test results and impacts that the test is intended to have. Likewise, we will look at each one in turn:

4.1.1. *Inferences and construct validity.* On the basis of the analysis provided by *WordSmith Tools*, we can draw conclusions on testees' vocabulary. This software package enables us to create word lists (in both alphabetical and frequency order), retrieve concordance output, and get collocation information; data which will put forward a thorough description of testees' embedded lexis. Moreover, not only will we be able to analyse each single text individually, but we may also study a sample of written texts.

As it has already been confirmed (C. A. ENGBER 1995, S. MORENO ESPINOSA *et al* 2005), there seems to be some relationship between lexical richness and writing ability. Therefore, we believe that the outcome of *WordSmith Tools* may also benefit teachers and researchers in order to ponder testees' writing competence. Nevertheless, we would like to stress the fact that, this software program does not draw conclusions in itself, but it helps teachers and researchers to spot lexical patterns, so that they will able to provide their value judgements on the basis of objective data. Therefore, as M. SCOTT (1997) noted, the quantitative data retrieved through *WordSmith Tools* and the linguist's intuition will be complementary, rather than antagonistic.

4.1.2. Uses of the gathered results. As J. Read and C. Chapelle observe, the practical outcomes of assessing instruments may be divided into three categories: instructional, research and evaluation uses. From our viewpoint, *WordSmith Tools* can be mainly used for instructional and research uses.

If we focus on *instructional uses*, we consider that this mechanical instrument can be employed effectively for diagnostic and placement purposes. Moreover, we would like to draw attention to the fact that:

- ESL/EFL teachers, linguists and/or students amongst others can examine English lexical patterns such as collocations, by analysing English corpora through *WordSmith Tools*. Therefore, *Concord* (i.e. the concordancer included within *WordSmith* package) can be used as a teaching, researching and/or learning instrument. Besides, it offers the possibility of blanking out the search word (i.e. the word we have analysed the company it keeps), in order to create vocabulary learning activities.
- When analysing learner corpora, it can be handled as an assessing instrument.

This test can also be used for *research uses*, being a clear example of this, the different studies found in the existing literature (Y. LI 2000, P. SAGASTA ERRASTI 2000, T. NAVES NOGUÉS 2001, P. DE HAAN and K. VAN ESCH 2004, B. ALTENHERG and S. GRANGER 2001, J. WEBER 2001, S. DE COOK *et al* 1998, A. O'KEEFFE 2004, W. CHENG 2004, M. BONDI 2004, C. SAMSON 2004, B. DAVIS AND L. RUSSELL-PINSON 2004).

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

808

4.1.3. *Intended impacts and actual consequences.* The intended effects of this tool have not been explicitly articulated, however it should be noted that, it was developed in order to fit the author's need as an applied linguist. This tool analyses corpora, therefore, we believe that: (a) it can be used to study L1 written texts, in order to draw inferences on lexical patterns amongst others; and (b) we can examine learner corpora to assess texts on the basis of lexis.

4.2. MEDIATING FACTORS

There are three factors that mediate between instrument purpose and test design which are: construct definition; performance summary and reporting; and instrument presentation.

4.2.1. *Construct definition*. By bearing in mind, J. READ (2000)'s three dimensions, we can state that *WordSmith Tools*' trait definition entails a tool which assesses embedded, comprehensive and context dependent vocabulary.

4.2.2. *Performance summary and reporting. WordSmith* package is composed of three computer programs: *Concord, KeyWord* and *Wordlist*, which will report the data according to their different characteristics. Thus, the former makes concordances, the second identifies key words in texts and the latter generates word lists (in frequency and alphabetical order), at the same time that it displays statistics such as the type/token ratio and word length among others. Thus, we can see that *WordSmith Tools* enables to examine machine-readable corpus from different perspectives, by providing a multiple component profile of testees' productive lexis. Issue which may be useful for teachers in order to know approximately what stage of vocabulary development students are at.

4.2.3. Instrument presentation. This test seems to be a practical instrument for teachers, linguists, learners and researchers in order analyse embedded vocabulary. It is available on purchase, and a demo version can also be downloaded from the author's web $page^4$.

Finally, we would like to provide our own summary of *WordSmith Tools* by taking into account J. READ and C. CHAPELLE (2001)'s framework of reference (see figure 3).

⁴ http://www.lexically.net/wordsmith/

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

TEST PURPOSE AND VALIDITY CONSIDERATIO NS	INFERENCES Description of testees' embedded vocabulary on the basis of : * Wordlists * Collocates information * Keywords 	USES Instructional Diagnostic Placement Research	 IMPACTS It should help teachers in their vocabulary assessment on the basis of the quantitative data retrieved. It can be used to analyse different corpora.
MEDIATING FACTORS	CONSTRUCT DEFINITION • Embedded • Comprehensive • Context-dependent	 PERFORMANCE SUMMARY Outcome of the test can be reported as a profile containing multiple insights into informants' lexis. 	 TEST PRESENTATION The test was presented in such a way so as to appeal a particular audience of L2 teachers, researchers and students among others.

Figure 3. WordSmith Tools according to the framework of reference

BIBLIOGRAPHICAL REFERENCES

- ALTENBERG, BENGT and GRANGER, SYLVIANE (2001): «The Grammatical and Lexical Patterning of Make in Native and Non-native Student Writing», *Applied Linguistics*, 22 (2), pp. 173-194.
- BONDI, MARINA (2004): «The discourse function of contrastive connectors in academic abstracts», in AIJMER, K. and STENSTRÖM, A. (eds.), *Discourse Patterns in Spoken and Written Corpora*, Amsterdam, John Benjamins Publishing Company, pp. 139-156.
- CHENG, WINNIE (2004): «FRIENDS, Ladies and GENtlemen: Some Preliminary Findings from a Corpus of Spoken Public Discourses in Hong Kong», in CONNOR, U. and UPTON, T. A. (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam, Rodopi, pp. 35-52.
- DAVIS, BOYD and RUSSELL-PINSON, LISA (2004): «Concordancing and Corpora for K-12 Teachers: Project MORE», in CONNOR, U. and UPTON, T. A. (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam, Rodopi, pp.147-169.
- DE COOK, SYLVIE; GRANGER, SYLVIANE; LEECH, GEOFFREY and MCENERY, TONY (1998): «An Automated Approach to the Phrasicon of EFL Learners», in GRANGER, S. (ed.), *Learner English on Computer*, Harlow, Longman, pp. 67-79.
- DE HAAN, PIETER and VAN ESCH, KEES (2004): «Towards an Instrument for the Assessment of the Development of Writing Skills», in CONNOR, U. and UPTON, T. A. (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam, Rodopi, pp. 267-279.
- ENGBER, CHERYL A. (1995): «The Relationship of lexical proficiency to the quality of ESL compositions», *Journal of Second Language Writing*, 4 (2), pp. 139-155.

INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.

810

- GRANGER, SYLVIANE and TRIBBLE, CHRISTOPHER (1988): «Exploiting Learner Corpus Data in the Classroom: form-focused instruction and data-driven learning», in GRANGER, S. (ed.), Learner Language on Computer, Harlow, Longman, pp. 199-209.
- LI, YILY (2000): «Assessing Second Language Writing: The Relationship between Computerized Analysis and Rater Evaluation», *Assessing Second Language Writing*, 127-128, pp.37-51.
- MORENO ESPINOSA, SORAYA; AGUSTÍN LLACH, MARÍA PILAR and FERNÁNDEZ FONTECHA, ALMUDENA, (2005): «Can P_Lex Accurately Measure Lexical Richness in the Written Production of Young Learners of EFL?», *Porta Linguarum* 4, (forthcoming).
- MUKHERJEE, JOYBRATO (2004): «Bridging the Gap between Applied Corpus Linguistics and the Reality of English Language Teaching in Germany», in CONNOR, U. and UPTON, T. A. (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam, Rodopi, pp.239-250.
- MURISON-BOWIE, SIMON (1993): MicroConcord Manual: An Introduction to the Practices and Principle of Concordancing in Language Teaching, Oxford, Oxford University Press.
- NATION, PAUL (1990): *Teaching and Learning Vocabulary*, New York, Newbury House Publishers.
- NAVES NOGUÉS, MARÍA TERESA (2001): «To what extent do computerised POS tagging textanalysis programs correlate with manually calculated analytical writing measures?», Paper presented at *AEDEAN*, Granada, 13-15 December 2001.
- O'KEEFFE, ANNE (2004): «'Like the Wise Virgins and All that Jazz': Using a Corpus to Examine Vague Categorisation and Shared Knowledge», in CONNOR, U. and UPTON, T. A. (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam, Rodopi, pp.1-20.
- OSBORNE, JOHN (2004): «Top-down and Bottom-up Approaches to Corpora in Language Teaching», in CONNOR, U. and UPTON, T. A. (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam, Rodopi, pp. 251-265.
- RAYSON, PAUL (2003): Matrix: A statistical method and software tool for linguistic analysis through corpus comparison, Ph.D. thesis, Lancaster University.
- REPPEN, RANDI (2001): «Review of MonoConc and WordSmith Tools», *Language Learning and Technology* 5 (3), pp. 32-36.
- Retrieved from http://llt.msu.edu/vol5num3/pdf/review4.pdf
- READ, JOHN and CHAPELLE, CAROL A. (2001): «A framework for second language vocabulary assessment», *Language Testing*, 18 (1), pp. 1-32.
- READ, JOHN (2000): Assessing Vocabulary, Cambridge, Cambridge University Press.
- SAGASTA ERRASTI, MARÍA PILAR (2000): La producción escrita en euskara, castellano e inglés en el modelo D y en el modelo de inmersión, PhD Thesis, University of the Basque Country.
- SAMSON, CHRISTINA (2004): «Interaction in written economics lectures», in AIJMER, K. and STENSTRÖM, A. (eds.), *Discourse Patterns in Spoken and Written Corpora*, Amsterdam, John Benjamins Publishing Company, pp. 199-216.
- SCOTT, MIKE (1997): Wordsmith tools, Oxford, Oxford University Press.
- TRIBBLE, CHRISTOPHER (2004): «The Text, The Whole Text ... or Why Large Published Corpora Aren't Much Use to Language Teachers and Learners», in LEWANDOWSKA-TOMASZCZYK, B. (ed.), *Practical Applications in Language and Computers*, Frankfurt am Main, Peter Lang, pp. 303-320.
- WEBER, JEAN-JACQUES (2001): «A concordance- and genre-informed approach to ESP essay writing», *ELT Journal*, 55 (1), pp. 14-20.
- INTERLINGÜÍSTICA. ISSN 1134-8941. 16 (2), 2005, pp. 803-811.