

In Poce, A., (ed.), *Individualizzazione del messaggio di apprendimento in ambiente adattivo / Individualisation of the learning message in adaptive environments*, Milano: FrancoAngeli, IT 49 – 60 / EN 47 – 56, ISBN 978-88-917-0994-3

Subject lexicons, corpora analyses and vocabulary assessment

by Andrea Zini¹

Introduction

The FIRB project “Adaptive Message-Learning” (2009-2014) has had the primary scientific goal of developing a didactic model of individualization which is the adaptation of the «learning message» to the «verbal skills» of the recipient (Vertecchi, 2010).

According to Chris Whetton (2011, 32) the general construct underlying the project can be formulated in a very simple way. Each student has a personal vocabulary, which can be measured. This information can be used «to construct individualized learning materials, such that the learning for that person is supported and enhanced». For the group as a whole, in terms of measurement, «this should lead to a reduction in variance of scores and an overall increase in mean of their scores». Within the am-learning model, assessment activities have different times, contents and objectives, which Whetton (2011, 31) leads back to two basic elements: «the initial measurement of learners’ ability to understand, and the final measurement of the outcomes of the learning, measured at the conclusion of the course or module». For the first part, the project has developed a measure of the personal lexicon that uses the cloze test to assess the availability of relevant vocabulary.

The initial and continuous assessment of the lexical competence of students, entrusted to the software LexMeter (Agrusti, 2010), is a core element in the am-learning model. It requires on the one hand the lexical analysis of the input to which students are exposed in a given context of formal learning, on the other hand the collection of measures on the reader in relation to the vocabulary of the texts and to the specific language of a subject area. In more detail, software-generated multiple-choice rational-deletion cloze-tests are used to measure the extent to which learners can presently handle relevant vocabulary in discipline specific texts, in order to estimate how effectively they could cope with standard course materials and then use this information for adapting the wording of the study text to the individual student.

The validation of the diagnostic assessment of am-learning required specific attention. In am-learning, quantitative analysis of words in a corpus of LSP texts is used for identifying the relevant vocabulary, that is to define by statistical criteria the vocabulary list to take into exam (which words may be deleted from a given text). The assessment of the individual vocabulary, then the automated construction and evaluation of the tests, are based on the frequency of occurrences of the content words in a specific language. In a nutshell, the software:

- produces a specific frequency word list through the lexical-statistical analysis of a representative sample of the language of the subject area;
- given a text (analogue to the materials of the course), constructs, administers and corrects rational-deletion cloze with multiple choice answers. The function words are excluded from consideration. Once you determine which grammatical classes of words are eligible for deletion (in the case of experiments conducted: noun and adjective), the hidden words must fall within a range of frequencies determined by the author of the test. It provides examinee a text that presents a certain number of gaps, accompanied by a list of alternatives from which he must choose one to

¹ Andrea Zini, Unità di Ricerca Università degli Studi di Modena e Reggio Emilia, andreazini3@gmail.com, English translation edited by the author.

reintegrate. The list includes all the hidden words of the original text plus other that might be added by the author of the test.

Surveys conducted by national research group has allowed the testing and development of this model within the OrbisDictus environment.

In addition, it is possible to establish a system of automatic assessment on different aspects of vocabulary, provided that they can be described with reference to explicit and quantifiable data of the language. In one of the exploratory studies conducted in Reggio Emilia, assessment is related to the stratification of the vocabulary of a specific corpus not in terms of frequency, but peculiarity of use.

Overview of the field tests

1. The preliminary phase of the work was devoted to the validation of the automatic evaluation tool, in particular:

- testing of cloze tests generated automatically by the software LexMeter;
- examining the difficulty of different types of items;
- the verification of the validity of the frequency of use of the hidden words as a criterion to differentiating the difficulty of the tests.

The tests administered at the University of Modena and Reggio Emilia have involved

- 28 students enrolled in the degree course in Physiotherapy and 120 students of the Degree in Nursing (first year);
- a group of 100 students enrolled in the second year of the degree course in Nursing.

The frequency of use of words has been shown since the earliest studies in the 1800s to be an important and useful quantitative criteria for predicting the readability of written texts for a determined category of readers (Amizzoni, Mastidoro, 1993). The aspect of the frequency of vocabulary use is considered influential both on the receptive performance (decoding and comprehension) and productive performance (Colombo, 1993), in as much that it has an effect on the strength and availability of the representation of a word in the mental vocabulary. Confirming these assumptions, the frequency of use of the hidden words has shown to be an effective criteria for predetermining the difficulty of the test.

2. After the definition of the environment OrbisDictus, each research unit has conducted a tryout in order to test the functionality of the system and start the development of technical solutions. The tryout in Reggio Emilia was hosted by Degree Course in Physiotherapy (Faculty of Medicine) in the integrated course of Biomechanics and Applied Kinesiology and has involved a cohort of 30 students (first year). Of these, 26 have completed all the scheduled assessments: two rational cloze tests (both administered at entry and exit) and two achievement tests relative to the content of the two learning units which make up the course. The cloze tests, built on two excerpts from a Kinesiology manual of around 750 words, contained 15 gaps and proposed 15 alternative answers. The two passages looked equivalent in terms of lexical profile and in terms of readability (comparison with the Italian basic vocabulary De Mauro, 1980; reading ease index "Gulpease"). A comparison of the hidden words in the two tests with De Mauro Italian dictionary (2000) showed that for almost all of them a specific definition can be found in the medical science field (mostly anatomy and physiology).

In terms of measurement, the cloze test where the hidden words were chosen within the medium and low frequency bands (excluding hapax) in a frequency list derived from a corpus specific to Rehabilitation medicine, it was revealed at the same time, more homogeneous, difficult and discriminative of the other, in which the low-frequency band was excluded from the examination (more details are in Zini, 2011).

3. A different field experience has involved a group of 74 students in the second year of the degree

course in Educational Sciences (Faculty of Education).

The type of entrance test that we have studied in this occasion was conceived as a reading essay of a specific monograph (considered as mini-corpus), indicated in the course bibliography (Cecconi, 2002). The two text samples (approximately 600 words each) selected to produce cloze tests were extracted from the first chapter of the same text.

With respect to previous experience, the most important difference lies in the different degree of formalization of the language used to write the texts: those on Educational Research are written in an academic register of standard Italian, while in those on Kinesiology the special language of Rehabilitation Medicine was prevalent.

The target words were chosen in the list of two hundred content words that occur most frequently in the book. The list comprises 3% of the types attested (200 of about 7300), which correspond to about 21% of the total occurrences in the mini-corpus. Almost none of the 200 types is original of a specialized technical vocabulary (one of the rare examples is “postulate”). By consulting the De Mauro dictionary (2000), we find for the half of the hidden words a technical-scientific meaning applicable to the use of that word in the text. Rarely this acceptance falls in the specific use of social sciences and humanities or philosophy.

The critical point of the cloze has proved to be the lack of internal consistency. In the discussion of the results this fact has been linked to the degree of ambiguity of the language of the text, and in particular of the target words.

The administration of the cloze alongside other tests that measure known constructs (terminology mastery; certain aspects of text comprehension) allowed us to observe a significant rank correlation (Spearman's rho) between the results obtained by the same subjects in the cloze test and in the specialized vocabulary test.

In addition, an “erratic” item was added (and separately considered) to the 15 items which make up the cloze test, obtained by deleting the connective adverb “Consequently” which occurs in both texts and links in a coherent manner two syntactically independent periods. It is a request for completion which is intended to directly arouse the textual competence of the reader. For this gap four specific alternatives of completion are offered. In addition to the correct one, they appear the distractors “However,” “Similarly”, “Instead”. The results of the tests suggest that this type of item, which has a different nature from the others and evidently does not contribute to the estimation of the available lexicon, it can nevertheless show an high index of discrimination within the cloze test and a significant association with the answers to text comprehension questions that require to produce bridging inferences. This first experience has opened the way for further developments. As we experienced even later on, it exists, although it appears limited, the possibility of constructing this type of item starting from explicit data of the language, and then respecting the constraints of automated processing.

4. The activities of the Research Unit of the University of Modena and Reggio Emilia has focused further on the content and tools of the initial assessment. A new review of the literature has accompanied the analysis of partial results obtained, enabling to design a new experiment aimed at the assessment of the peculiar vocabulary of a disciplinary language in the context of the study materials. The subsequent field test has involved a new cohort of 30 first-year students of the degree course in Physiotherapy. In the following sections we will discuss some aspects of this work (for more details, see Zini, 2012, 2014).

Highlights from the literature review

The literature attests the crucial role of vocabulary knowledge in reading comprehension, although this role is not to be understood ingenuously.

Indeed, the vocabulary knowledge and the mastery of the morphosyntactic structures of the language are immediately required for those who want to understand even a single utterance, but not necessarily assure the understanding of the (expressed or underlying) contents of a text, intended as a cohesive and coherent discourse, which also implies other processes. The same proper

attribution of meaning to words, both known and novel, encountered in the reading of a text, is normally influenced or permitted by the context and the co-text in which they are placed. Consequently, “in order to understand a text it is not only necessary to have a large vocabulary, but, more importantly, to understand the meaning that words take on when they are used, how their meaning changes in a co-text” (Salerni, 1989).

The language proficiency of an individual and the breadth of its vocabulary are closely associated with the processes of learning. Care must be taken that the “knowledge of the meaning” can have various senses, and can be assessed in different ways. Indeed, to assess the vocabulary different procedures can be used (Passolunghi, De Beni, 2001).

The most obvious difference separates an explicit test of vocabulary knowledge, which is obtained “using a test consisting of discrete points items, each of which tests a single lexical component”, from an implicit assessment, “as in an integrated test which requires the knowledge of a particular set of lexical items in order to complete the assigned task” (Davies, 2008).

Asking the meaning of an isolated word, we only get information on the extension of the vocabulary of the subject, while we remain unaware of the degree of organization of its lexical system, ie, the subject's ability to master the rules of combination of words (Boschi, Aprile, Scibetta, 1992).

The literature review has confirmed the rational-deletion cloze procedure to be a suitable test format to complement measures of vocabulary size with more contextualised measures of vocabulary use, which take into account the depth of vocabulary knowledge in specific contexts of use, such as academic disciplines (Read, 2000, 2004; Salerni, 1989).

As Salerni stated (1989):

Certainly the use of a test is a function the objective that you want to check. Indeed, if the aim is to measure the lexical skills of students in relation to the total process of reading, it seems essential that the test measures the knowledge of vocabulary in context. The test should measure to which point a student is able to use the syntactic, semantic and morphological clues in order to derive the specific meaning of a word as it is used in a particular context. For this purpose rational cloze tests or gap-filling tasks may give a useful indication.

According to our task-analysis, the answers to rational cloze tests with multiple choice answers should give an account of three operational skills, the second of which applies upon reading the incomplete text, the other two on reading the list of words offered to fill in the blanks:

- recall the meaning of familiar words (of which is already known the applicable meaning);
- infer the meaning that partially known or unknown words assume based on the context;
- correctly anticipate the meaning of the individual missing words using clues in the text and our own prior knowledge (general and specific).

The three operational skills could be traced back to as many components of the specific cognitive process of reading comprehension: vocabulary, lexical inference and semantic inference (as defined in De Beni, R., Cornoldi, C., gruppo MT, 2004).

In Reed (2007) references are found to the current and potential contribution of corpus analysis to extend our understanding of the lexical features of academic registers (Read, 2007), and to the application of word lists in the assessment of learners' lexical abilities: “For vocabulary assessment corpora can provide the basis for more accurate word lists from which target words can be sampled, taking account of frequency, range of occurrence and other criteria” (Read, 2007: 106).

We need to complement discrete vocabulary tests with embedded measures of the learners' ability to handle lexical items in context. Traditionally, context has been conceived in linguistic terms as a sentence or larger co-text in which a vocabulary item occurs. Corpus analysis now gives us powerful tools to enrich our understanding of context through providing detailed descriptions of vocabulary use in specific registers or disciplinary genres. The applications of these richer descriptions to vocabulary assessment are not entirely clear as yet but they represent fertile ground for further investigation.

Among the experiments conducted by Read, one called “Words in Context Test” (2004) has particular relevance to the subject of this study, also because the format adopted is the multiple-choice, selective-deletion cloze. In the terms of the Reed model (2000), it regards the exploration of a kind of ‘discrete test’ (which measures the knowledge and use of vocabulary as an independent construct), which is ‘selective’ (aimed at a specific vocabulary: the words which belong to the Academic Word List) and at the same time “context-dependent”.

The main issue in writing the test is how to make the items context-dependent. The initial intention was to have, as the options for each item, phrases expressing possible meanings of the target word so that the test-takers would need to process contextual information to be able to choose the correct option. [...] However, in the test as a whole, it proved to be difficult to maintain this approach consistently. It seems that AWL words are not as polysemous as expected, at least not to the degree that three phrases can be composed representing actual meanings of the word which are plausible in the particular context of the target word. [...] For the majority of the items, then, I reversed the logic in a sense. [...] the options tend to represent meanings which are plausible in the context and the test-takers need to have some specific knowledge of the target word to be able to choose the correct answer. (Read, 2004: 25, 27)

Very similar issues have arisen in our own experience, where the availability of plausible alternative answers depends directly on the choice of which words to hide.

Words in specific and general use

A recent study (Truffelli, 2010) on reading comprehension of academic texts conducted on students in Educational Sciences at Bologna University, describes a complex picture of the different components of the process, which shows that the knowledge of the meaning of specific terms or syntagms and the ability to search and identify specific information contained in the text are the most critical for the whole sample group. “In particular, the widest gap between the full sample and 'poor' readers is observed both with respect to the knowledge of the terms, in line with the findings reported in the literature [...] and with respect to the overall understanding of the text”.

Assuming the distinction between terms and words, “the terms which have determined meanings, words which have meanings which are indeterminate and expansive in their use and dilatable in the use by extensions and metaphors”, it can be stated that “in a field of discourses, the greater the use of terms, the greater the share of technicality and scientificity” (De Mauro, 1998). This does not imply that the terms are “words of different physiognomy than the ordinary” nor that the repertoire is vast, but “it implies only that much of the discourses must be constructed with well-defined terms, also drawn from a very limited repertoire”. The incidence of the terms within the texts varies between different fields of discourses.

In our study, take on particular importance the specification and the automatic extraction of a category of terms to be taken into consideration. According to Adriano Colombo (2002: 33), “When the technical terms have been specially created, usually they do not constitute difficulties: they will mostly be novel words to the students, but their (unique) meaning will be learned along with disciplinary concepts”. However, many terms, “far from being newly coined, are nothing more than words of the general language that undergo a redetermination and clarification of meaning when they are taken within a given specialized vocabulary” (Lavinio 2004: 101). This type of terms can offer students a particular difficulty in understanding, because “the words which become terms will inevitably lead behind the halo of meanings they continue to possess in general language” (Ibid) and <switch between common acceptations to those of the specific lexicons may not be easy> (Colombo, 2002: 33).

A cloze test on peculiar language of Rehabilitation Medicine

In the last of the experiments conducted, the words to hide in the cloze test were selected within a

particular repertoire, namely the peculiar vocabulary of rehabilitation medicine. A corpus of Rehabilitation sector language was collected (Corpus of Rehabilitation Medicine, CMR), which mainly represents the language of Rehabilitation Medicine intended as a scientific academic field (the majority of the texts collected consists of manuals, lectures and dissertations). The corpus, stratified according to three variables (author, recipient and text type) that can assume different modalities, has about 1,100,000 occurrences (tokens) of 64,000 words (types). The extraction of peculiar language consists in “selecting the words to be analyzed not in function of their absolute frequency, but of their over/under use, compared to the medium use in a language of reference” (Bolasco, 2008). For this purpose, not only endogenous resources but also exogenous resources to the specific corpus have been employed. The statistical-linguistic resource used as a reference model to extract useful information from the specific corpus is a frequency list called the Italian standard (Bolasco, 2008).

Comparing the vocabulary of the corpus CMR with the frequency list of standard Italian POLIF2002 contained in the TaLTaC2 database, has been defined the set of types that are represented both in the standard Italian use and in the use of rehabilitation medicine (intersection) and that within the latter are over-represented (with a variance on the occurrences of positive and significant value).

The final list of peculiar types includes 1230 elements, accounting for about 2% of different types in the CMR, which correspond to about 16% of the total occurrences (tokens) in the same corpus.

Which words belong to this list and which ones are excluded?

Firstly, the “banal” words are excluded, which have a frequency of occurrence equivalent in CMR and in Polif. In addition, are excluded all “original” words of the Corpus of Rehabilitation Medicine, namely, all types not attested in the lexicon of standard Italian Polif.

The included words are in many cases technical terms whose meaning is related to the discipline, in other cases academic or common words (if not fundamental), which have a looser semantic relationship with the specific field of discourses. Therefore, the words that fall within the considered repertoire seem differentiated enough to allow an assessment of lexical competence that is both extensive and specific.

Results

In entry tests, which in this case coincide with the object of the study and were conducted in presence, the two cloze tests are flanked by other kind of tests regarding text comprehension, technical vocabulary and prior knowledge.

In order to understand how the cloze-test worked on the group of students who completed all of the tests (28 subjects out of 30), we use some classical item analysis tools. The items which make up the cloze test cover well differentiated difficulty levels. In the first test, three items out of fifteen turn out to be critical for excessive ease and low discrimination, while in the second test we note just one critical item out of twenty. The reliability of cloze tests as measured by the correlation between parallel components, defined as even and odd items, corrected by the Spearman-Brown formula, assumes in the case of the first test the value 0.72 and in the case of the second test the value 0.82.

An examination of the relationship between the cloze test scores and those of the other tests conducted using the Spearman's index of rank correlation (ρ) allows to observe a significant correlation ($r_s = 0.52$, $p 0.01$) between the scores obtained by the same subject to the first cloze and the test of specialized vocabulary (definition of 22 terms out of context). In the case of the second cloze test, there is a positive relationship, although weaker ($r_s = 0.38$, $p 0.05$), with the test results of prior disciplinary knowledge.

Conclusions

The relationship found with concurrent measurements of the extension of terminological heritage of

the readers and tests that concern the learning of disciplinary concepts explains in part the nature of these cloze tests as tests of specialized vocabulary and partly as tests of knowledge. In the context of the project “am-learning” and for the purpose that the considered assessment has within it, the instrument proves to be appropriate.

The experience also suggests the possibility of enriching with additional tools a computerized assessment system that uses both lexical and linguistic (in a wider sense) indicators to intervene in different comprehension difficulties. In particular, we explored the possibility of instructing an automatic procedure to generate gap-filling items removing some connectives from a text, in order to arouse the sensitivity to text cohesion and coherence.

References

- Agrusti, F. (2010). From LexMeter to Adapter. Towards a match up between the Virtual and the Real Reader, *Cadmo. Giornale Italiano di Pedagogia sperimentale. An International Journal of Educational Research*, 1, 97-108.
- Amizzoni, M, Mastidoro, N. (1993). Linguistica applicata alla leggibilità: considerazioni teoriche e applicazioni, *Bollettino della Società Filosofica Italiana*, 149, 49 – 63.
- Bolasco, S. (2008). *Corpora e liste di frequenza d'uso: criteri e tecniche per l'analisi automatica dei testi*, in in Barni, M. , Troncarelli, D. , Bagna C. (eds.), *Lessico e apprendimenti*, Milano: Franco Angeli: 113 – 142.
- Boschi, F., Aprile, L., Scibetta, I. (1992). *Le parole e la mente. Studi e ricerche sullo sviluppo della competenza lessicale nel ragazzo*, Firenze: Giunti, 69.
- Cecconi, L. (ed.) (2002). *La ricerca qualitativa in educazione*, Milano: Franco Angeli.
- Colombo, A. (2002). *Leggere. Capire e non capire*, Zanichelli: Bologna.
- Colombo, L. (1993), *Locus o loci dell'effetto frequenza?*, in Laudanna, A. e Burani, C. (eds.), *Il Lessico: processi e rappresentazioni*, Roma: NIS.
- Davies, A. (2008). *Che cosa ci dicono i test lessicali riguardo alla padronanza linguistica?*, in Barni, M., Troncarelli, D., Bagna C. (eds.), *Lessico e apprendimenti*, Milano: Franco Angeli: 104.
- De Beni, R., Cornoldi, C., gruppo MT (2004). *Nuova guida alla comprensione del testo*. Trento: Erickson.
- De Mauro, T. (1980). *Guida all'uso delle parole*, Roma: Editori Riuniti.
- De Mauro, T. (1988). *Linguaggi scientifici e lingue storiche*, in Guerriero, A.R. (ed.), *L'educazione linguistica e i linguaggi delle scienze*, Firenze: La Nuova Italia, 9 - 19.
- De Mauro, T. (2000). *Dizionario italiano*, Milano: Paravia.
- Lavinio, C. (2004). *Comunicazione e linguaggi disciplinari. Per un'educazione linguistica trasversale*, Carocci: Roma.
- Passolunghi, M.C., De Beni, R. (2001). *I test per la scuola*, Bologna: Il Mulino: 95.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2004). Second Language Vocabulary Testing: Taking a Broader Perspective, *2004 International Conference on English Instruction and Assessment*.
- Read, J. (2007). Second Language Vocabulary Assessment: Current Practices and New Directions, *International Journal of English Studies IJES*, 2, 105 - 125.
- Salerni, A. (1989). *La verifica della competenza lessicale*, in Lucisano, P., Salerni, A., Benvenuto, G., Siniscalco, M.T. (a cura di), *Lettura e comprensione*, Torino: Loescher, 117 - 151.
- Truffelli, E. (2010). Comprendere per migliorare lo studio: analisi e riflessioni a partire da un'esperienza biennale di sostegno alla matricole universitarie, *Giornale italiano della ricerca educativa*, giugno 2010, pp. 91 - 104.
- Vertecchi, B. (2010). Nuove ipotesi per lo sviluppo della didattica in linea, *Journal of e-Learning and Knowledge Society*, 6(1), 31 – 42.
- Whetton in Poce, A., Angelini, C. (2011). *Towards a new educational culture. Reflections on the am-learning project*, FrancoAngeli: Milano.
- Zini, A. (2011). *Analisi del lessico e individualizzazione del messaggio di apprendimento. La preparazione di un corso nell'ambito del progetto “adaptive message learning”*, in Minerva T.,

- Colazzo L. (eds.), *Connessi! Scenari di Innovazione nella Formazione e nella Comunicazione. VIII Congresso Nazionale della Società Italiana di e-Learning*, Milano: LediPublishing, 961-970.
- Zini, A. (2012). Misurare la competenza lessicale in contesto specifico attraverso prove di cloze, *Giornale Italiano della Ricerca Educativa - Italian Journal of Educational Research*, 9, 108 - 119.
- Zini, A. (2014). *Una prova di cloze sul lessico peculiare di un linguaggio scientifico*, in Colombo, A., Pallotti, G. (eds.), *L'italiano per capire, "I Quaderni del GISCEL"*, Roma: Aracne, 223 - 239.