



# Applying lexical statistics to rational cloze test to assess the relevant vocabulary in course readings

Andrea Zini, University of Modena and Reggio Emilia, Dept. of Education. andrezini3@gmail.com

## Background

This study is part of the "adaptive-message learning" project (FIRB, 2009-2013), aimed at the individualization of learning messages in e-learning systems (Vertecchi, 1998, 2010), which consists of adapting the wording of a study text to the ability of a reader cooperating with the text. Within the project, the objective of the assessment is therefore to measure indirectly this individual ability in relation to a precise context (linguistic and extra linguistic), that is to know how effectively the students cope with course readings. In more detail, computer generated, multiple-choice, rational-deletion cloze-tests (Agrusti, 2010) are used to measure the extent to which learners can handle relevant vocabulary in discipline specific texts. Quantitative analysis of words in a corpus of LSP texts is used for identifying relevant vocabulary.

## Study purpose

The validation of this tool required specific attention. This is due to the conditions of use of the test within the project, which requires the definition of a procedure which must be largely automatic so as to satisfy the need for continuous assessment (the tests are administered at the beginning of each learning unit and periodically throughout). For this reason it was necessary to compare the effects of different statistical criteria on classifying and selecting content words to delete in a passage. Moreover, it is necessary to consider the risks of validity connected to the use of the cloze as a lexical test. Precisely because an evaluation linked to the context is obtained through the cloze test, it is difficult to discern what the specific contribution of the vocabulary is in the execution of these tests (Read, 2000).

## Research questions

- Is the multiple-choice rational cloze-test a suitable kind of test for measuring the lexical abilities of university students with relation to authentic academic texts?
- Is it possible to define a deletion procedure guided by the analysis of the textual data?

## Certain premises. Vocabulary and comprehension

1. 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.

2. Assuming the distinction between *terms* and *words*, it can be affirmed that in a field of speech, the greater the use of terms the higher the proportion of technicality and scientificity. This does not imply the terms to have an unusual look nor that the repertoire is vast, but it implies only that much of the speech must be constructed with well defined terms (De Mauro, 1998).

3. The literature attests to the important relationship between the knowledge of vocabulary and the ability to understand the text, although this relationship is not clear. Understanding the significance of individual words does not necessarily ensure the comprehension of the content expressed or implicit in a text (intended as cohesive and coherent discourse), which implies other processes, and conversely, a limited vocabulary knowledge does not always compromise comprehension (Oakhill, Cain, Bryant 2003).

## Cloze testing

Cloze-tests consist of a passage which has had words deleted from it. The procedure was originally codified by W. Taylor as a measurement tool for the readability of texts. J.R. Bormuth (1967, 1969) developed the application of the technique in the evaluation of instructional programs that utilize written materials. Experience in Italian (Lucisano, 1989, 1993, 2010) confirms the validity of cloze for the same surveys, as well as explore its potential as a teaching tool (Marello, 1989; Salerni, Siniscalco, 1991). The cloze-test has been widely used as integrative language test (Oller, 1979) suitable for evaluation of the overall level of competence (proficiency) and there are also important applications in this field and studies in Italian (Grassi, Nuzzo, 2011).

There are two main types of cloze-test: the "natural" cloze-test with fixed-ratio deletions (blanks occur at regular intervals) and the "rational" cloze test, in which the categories of items are predetermined and not the distances. In the literature are evidence in support of different hypotheses as to which skills are measured by the two types of cloze (review in Oller, Jonz, 1994). Assuming that within a text a larger number of words have a "syntactic function" rather than a "cohesive" one, Bachman (1982) considered that the fixed-ratio deletion procedure tends to effect a higher proportion of words whose completion depends only on the immediate co-text and therefore measures mostly skills of a lower order, while by using a rational deletion procedure it is possible to measure also skills of a higher order.

In J. Read references are found both to the application of corpus analysis to vocabulary testing (2007), and to the use of the multiple-choice rational cloze-test format, as happens in the experience known as "Words in Context test" (2004).

Even in A. Salerni (1989, 138) the rational cloze-test (as well as the gap-filling task, as also described by Salerni, Siniscalco, 1991) is indicated for the objective of measuring to which point students are able to use the syntactic, semantic and morphologic indications so as to arrive at the specific significance of a word as it is used in a particular context.

## Exploring various criteria for rational deletion

We consider the results of a set of tests conducted at the University of Modena and Reggio Emilia. Since tests were developed as part of a course, students who were actually taking the course were used as subjects in grupus of 30-40. Self-contained excerpts from Kinesiology and Research Methodology in Education course readings were sampled, which are characterized by different levels of formalization of the language used for writing the texts: the latter written in an academic register in standard Italian while in the former a language specific to Rehabilitation Medicine is prevalent.

## References

Bachman, L.F. (1985). Performance on Cloze Tests with Fixed-Ratio and Rational Deletions, *TESOL Quarterly*, 3, 535-556.  
 Barni, M., Troncarelli, D., Bagna C. (eds.), *Lessico e apprendimenti*, Milano: Franco Angeli.  
 Lucisano, P., Salerni, A., Benvenuto, G., Siniscalco, M.T. (eds.) (1989), *Letture e comprensione*, Torino: Loescher.

TAB.1. Measures on selected passages.

	Education - Passage 1	Education - Passage 2	Kinesiology - Passage 1	Kinesiology - Passage 2
	La ricerca qualitativa	Lo studio di caso	La postura eretta	La prensione
Fundamental*	55%	53%	48%	46%
High usage	16%	17%	14%	13%
High availability	14%	18%	17%	20%
Not in Basic Vocabulary	13%	10%	19%	20%
Gulpease Index**	34	33	30	34

\*Fundamental Lexicon; Lexicon of High Usage; Lexicon of High Availability. The combination of words that fall into one of these repertoires is the Basic Vocabulary of the Italian language, which has about 7000 entries (De Mauro, 1980).

\*\*Respect to the level of education "higher diploma", values 30-40 correspond to the difficult level (Piemontese, 1996, 102).

For the activity on the research methodology texts the monograph from which the two chosen passages were taken (Cecconi, 2002) was considered as corpus (RME, Research Metodology in Education). So as to analyze the kinesiology texts a specific corpus (RMC, Rehabilitation Medicine Corpus) was constructed and stratified on the basis of three variables (author, recipient and text type).

TAB.2. The reference lexicon: statistics on corpora vocabulary.

Statistics on corpora vocabulary	Research in Education	Rehabilitation Medicine
token	N	56.140
type	V	7.319
type/token ratio	(V / N) * 100	13,04
hapax ratio	(V <sub>1</sub> / V) * 100	47,82
average frequency	N / V	7,67
coefficient G (Guiraud)	V / sqrt(N)	30,89
coefficient a	logN / logV	1,23

In order to avoid that the choice of the correct answer is necessitated by grammatical values required by the co-text, the am-learning system limits the choice of the items to two classes of words (names and adjectives) and expects that the list of alternatives can be integrated with the addition of distracters. Moreover, by applying to the text analysis a stopwords list, it sees to the exclusion of function words. These two barriers have been respected in each case, even when the system for the construction of the tests was not used (Test a) or variants were used (Test c).

The criteria for selecting which words to hide are shown in BOX 1.

## Test Analysis

For each test are observed in particular the difficulty (average scores) and internal consistency reliability, measured by indices of homogeneity of the items (Alfa) and correlation between two components (the split-half subdivision of the items among peers and odd it is considered to be particularly suited to the nature of the cloze-test).

TAB.3. Test Analysis.

test	Passage	Examinees	Classical item and test analysis										IRT item parameter calibration		
			Items	Alpha	S-B Odd-Even	SEM	Mean	SD	Min Score	Max Score	Item Mean	Item R	Calibrated Items	Alpha	Mean
a	La ricerca qualitativa	40	15	0,20	0,32	1,65	9,53	1,84	6	12	0,64	0,06	9	0,56	6,25
b	Lo studio di caso	34	15	0,18	0,42	1,58	10,06	1,74	6	14	0,67	0,06	11	0,43	6,82
c	La postura eretta	26	15	0,50	0,70	1,39	12,15	1,95	8	15	0,81	0,17	11	0,59	10,50
d	La prensione	26	15	0,69	0,87	1,38	11,00	2,47	3	14	0,73	0,35	14	0,72	10,12
e	La postura eretta (1 <sup>st</sup> half)	28	15	0,56	0,71	1,49	10,96	2,25	6	15	0,73	0,20	13	0,63	10,11

1. Comparing the cloze-tests of Education (Test a) with those of Kinesiology, it may cautiously be assumed that the internal consistency reliability of this type of cloze-test is in relation:

- a) positive to the degree of formalization of the language of the text;
- b) negative with the choice of high-frequency words in the reference lexicon.

2. In Test b, the cloze passage named "La prensione" (Grasp), in which the words to hide were chosen from a list that includes all medium and low-frequency words (excluding hapax), was both more consistent and difficult and discriminative than the one named "La postura eretta" (The upright posture), where, instead, most of the words belonging to the low-frequency band had been excluded by the choice.

3. Observations on the criterion of words' peculiarity (Test c).

a) The "banal" words, which have a frequency of occurrence equivalent in RMC and POLIF (a reference lexicon of standard Italian), they normally have a less determined meaning and are less significant in context.

b) The kinesiology texts presuppose the knowledge of notions and terms from anatomy and physiology which are extraneous to standard Italian vocabulary but are accessible to whoever has already received, as is the case here, the relative teaching.

As a consequence:

- a) the exclusion of banal words (in addition to fuction words) is beneficial;
- b) the exclusion of the original vocabulary of rehabilitative medicine reduces informative power of the test.

4. Examining the relationship (Spearman's rho) between the scores obtained by the same subjects to the cloze-tests and to tests of specialized vocabulary, is observed on all occasions a moderate significant rank correlation.

## Conclusions

The relationship found with concurrent measurements of the extent of readers' knowledge of terminology, which mainly implies the prior learning of concepts, explains in part the nature of these cloze-tests, built on texts with a high information content, as knowledge tests. Further studies may deepen the dependence of the task from the context.

The test offers the possibility of being used both as an evaluation tool and as an exercise. The administration in e-learning does not require that the reading is interrupted, and may be concurrent with the study of the text.

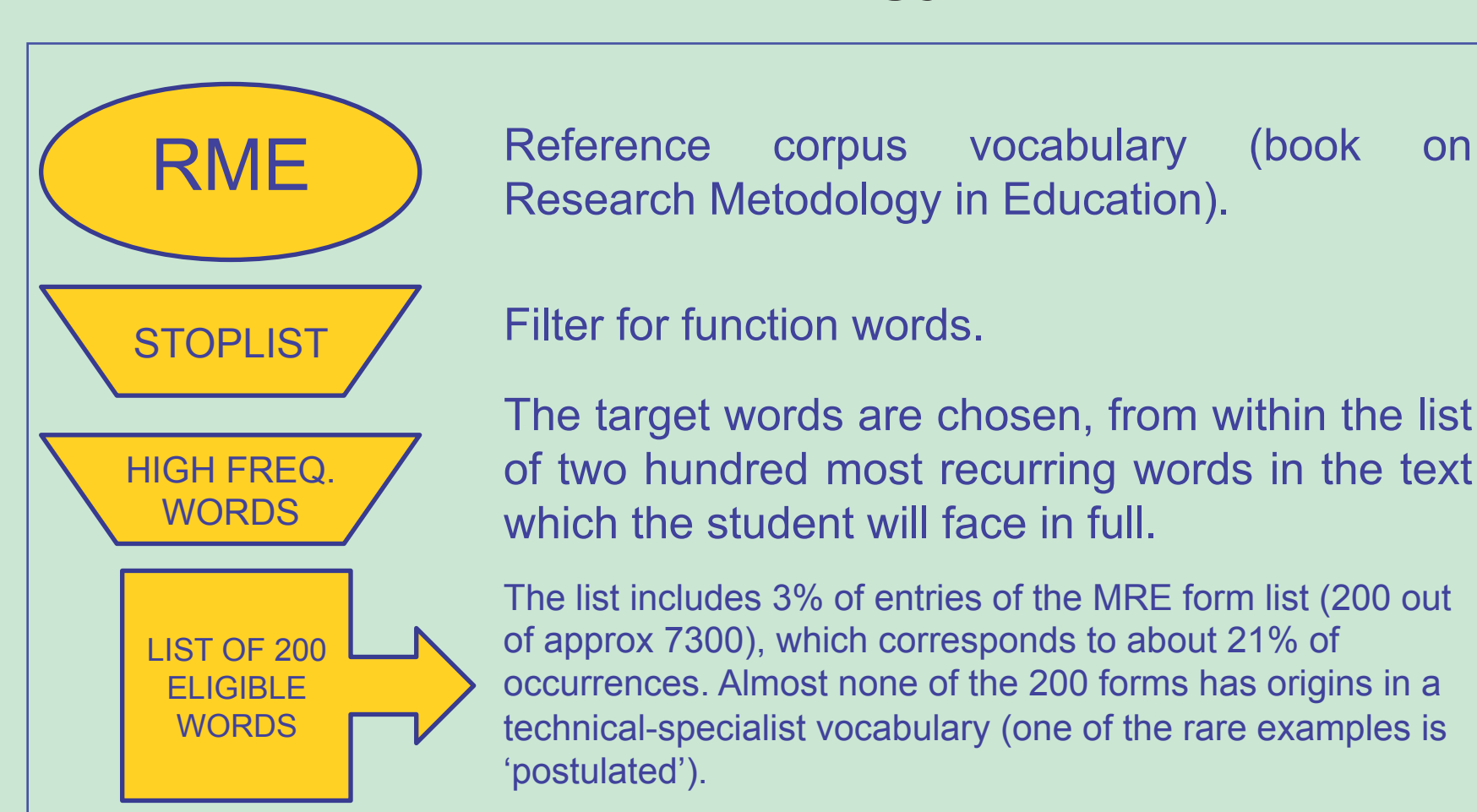
Lucisano, P. (1993). *Misurare le parole*, Roma: Kepos.

Read, J. (2004). Second Language Vocabulary Testing: Taking a Broader Perspective, *2004 International Conference on English Instruction and Assessment*.

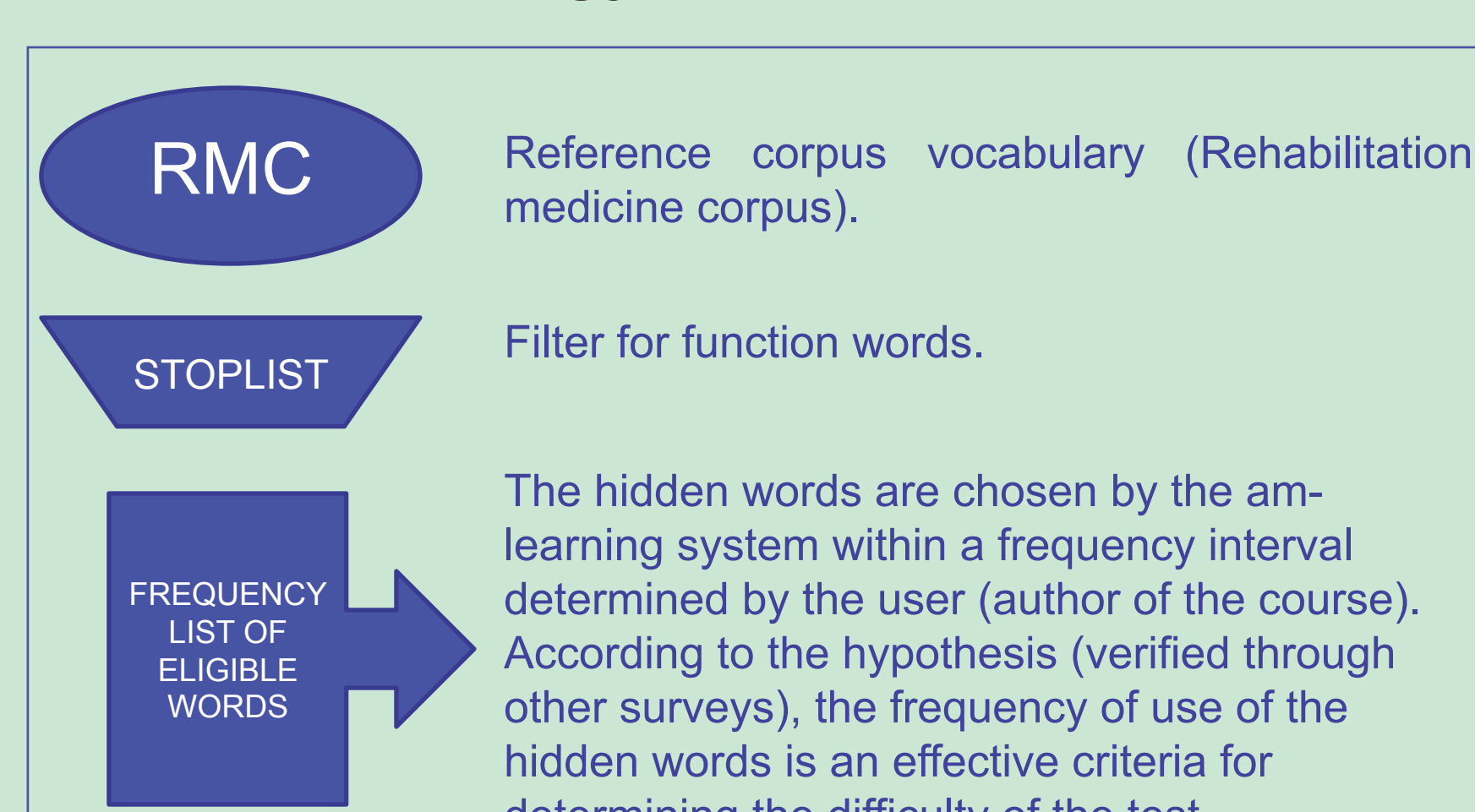
Vertecchi, B. (2010). Nuove ipotesi per lo sviluppo della didattica in linea, *Journal of e-Learning and Knowledge Society*, 1, 31-42.

## BOX 1. Criteria for rational deletion

### Test a - Research Metodology (Education)



### Test b - Kinesiology - 2010-2011 Cohort



### Test b - Kinesiology - 2011-2012 Cohort

