Effect of Tuned Parameters on a LSA MCQ Answering Model

Alain Lifchitz, Sandra Jhean-Larose, Guy Denhière

Effect of Tuned Parameters on a LSA MCQ Answering Model

Alain Lifchitz[#1], Sandra Jhean-Larose [*2], Guy Denhière[*2]

[#]*LIP6, DAPA, Université Pierre et Marie Curie – Paris 6, CNRS UMR 7606*
104 avenue du président Kennedy, F-75016, Paris, France
[1]alain.lifchitz@lip6.fr

[*]*Équipe CHArt: Cognition Humaine et Artificielle, EPHE-CNRS*
41, rue Gay Lussac, F-75005, Paris, France
[2]sandra.jhean-larose@ephe.sorbonne.fr
[2]guy.denhiere@ephe.sorbonne.fr

***Abstract* — This paper presents the current state of a work in progress, whose objective is to better understand the effects of factors that significantly influence the performance of the Latent Semantic Analysis (LSA). A difficult task, which is answering to biology MCQ, was used to test the semantic properties of truncated singular space and to study the relative influence of several parameters. An original and dedicated software eLSA1 has been used to fine tune the LSA semantic space for MCQ purpose. With the parameters of best configuration, the performances of our model were equal or superior to 7th and 8th grades students. Besides, global entropy weighting of answers was an important factor in the model's success.**

I. Introduction

In this paper, we have following goals: (i) to search for a method that enables us to obtain best input features, of type TFIDF, for the LSA as a non-supervised learning method (ii) to define a concrete task (answering to Multiple Choice Questions (MCQ)) that permits, on one hand, to evaluate the semantic nature of the obtained vector spaces and, on the other hand, to measure the relative influence of the parameters used to build these spaces (iii) to describe some original aspects of the dedicated tool developed to realize these processes (iv) to compare the model to the results obtained from 7th and 8th grades' students.

A. Looking for better TFIDF features as input of LSA

Latent Semantic Analysis (LSA) [2] has proved to provide reliable information on long-distance semantic dependencies between words in a context, using the "bag of words" model. LSA combines this classical vector-space model with singular value decomposition (SVD). Thus, "bag of words" representations of texts can be mapped into a modified vector space that reflects, to some degree, their semantic structure. It is the consequence of the reduction of dimensionality induced allowed by the truncation of the singular space restricted to the orthogonal components associated with the higher singular values.

This paper presents the state of our on going work, which is similar to the work of Wild & al. [24]. We tried to measure the effects of tuning the parameters of the input textual features TFIDF of LSA [21] [20], and more precisely, the effects of lemmatisation, stop-words list, weighting of terms in the Terms*Documents matrix, pseudo-documents, and normalization of document vectors.

B. Semantic spaces: at which degree semantic?

Unlike free answer questions that are frequently used in LSA research ([9], [6]), this paper addresses how to automatically find the right answer to multiple-choice questions (MCQ) using LSA. An answer to this question could be interesting both from a cognitive point of view and in practical applications.

We have built a model to answer MCQ, which is a non trivial problem and not frequently studied, even if LSA is frequently used for the e-learning, and for the processing of questionnaires. The limited number of available terms to choose the correct answer to the MCQ determines the difficulty of the task and the small size of our corpora further increases this difficulty [19].

The model we propose is based on the following two assumptions: (i) each question and its associated answers are represented by a "bag of words" model, and (ii) the correct answer is the one out of three, which has the higher similarity with the question. The results presented below indicate how much these two rough assumptions are effective and what their limitations are. We also take into account the specific properties of MCQs by adopting original entropy global weighting of answers.

C. eLSA1: motivation for a dedicated tool

Quesada [19], in his chapter entitled "Creating Your Own LSA Spaces", does not recommend to build one's own LSA toolkit because of its complexity and presents the most frequently used LSA software's (see also [23], [1]). Nevertheless, facing the links between the successive steps of processing, the desire to understand in detail the stages of processing, we find it necessary to develop our own software in order to implement some specific algorithms. This MCQ dedicated *eLSA1* software could be extended to other tasks in the future as needed.

D. Comparison eLSA1 model with student performance from 7th and 8th grades

LSA can be considered as a theory of meaning [13] and as a model of semantic memory [3]. According to this, LSA permits to compute the relative importance of textual statements necessary to summarize a text [4] or to predict the eye movements of readers as a function of relative importance of statements [22]. If the cognitive relevance of LSA for learning and summarizing is generally accepted, it is still to be proved in the case of MCQ. So, we will compare the results obtained from *eLSA1* to the performances of students on the same MCQ by varying some properties of the corpora that are known to influence the performances of learners such as titles of documents, quantity and nature of information.

E. The structure of the paper

The rest of this paper is structured as follows. The original aspects of the eLSA1 software and the sequence of LSA processing specific to MCQ are presented in section II. The data used in the experiments: corpora, optimized semantic spaces and MCQ, will be presented in section III. A typology of questions and answers and various forms of "non differentiation" between answers will be presented in section IV. The relative influence of the parameters on the quality of results will be described in section V. Finally, comparisons between the *eLSA1* model and the student performances will be presented in section VI.

II. *eLSA1*: THE TOOL AND ITS IMPLEMENTATION

eLSA1 has been developed in freeware Python interpreted language [18] with numerous ready to use libraries, in particular numerical matrix calculation library NumPy [15].

A. eLSA1 features

The main original *eLSA1* features are:

- triggered lemmatisation for a couple of words, with the same prefix, based on predefined couples of suffices;
- joint lemmatisation for both the corpus and the MCQ;
- building of a stop list specific to the content of the learning corpus;
- global entropy weighting of the MCQ answers;
- automatic detection of questions that lead to "indiscernible" answers for the "bag of words" model.

B. Triggered lemmatisation

The effects of stemming and lemmatisation are controversial ([3], [12]) and probably depend on the size of the used corpora. Stemming and lemmatisation have the same effect: similar vector components are merged to create an equivalence class (the stem or the lemma) with less statistical noise; as a consequence, the vector space dimension is reduced. For limiting the risks of spurious equivalence classes, we developed our own solution. Our lemmatiser used rules like Porter's stemmer [16][17], but triggered by a co-occurrence of predefined suffices present in each couple of words in the corpus that share the same prefix.

C. Joint lemmatisation

In LSA, similarity can only be computed between terms that belong to the learning corpus. So, the similarity computed between the MCQ pseudo-documents can only take into account the terms from the corpus. To increase the number of common terms between corpus and MCQ, a joint lemmatisation was conducted.

D. Computer aided stop list design

For building our stopword lists, we make an original use of a specific property of the entropic global weighting (1 - entropy) [7] of the terms vector of the Terms * Documents matrix which, by definition, varies between 0 and 1 : 0 when the term is present in all documents with the same frequency, 1 when the term is present in only one document. A good candidate of a stopword list must have low global weighting values, but the reverse is not necessarily true for specialized corpora as we used here. So the following procedure was adopted:
(i) *eLSA1* lists the first 150-200 terms ranked by increasing (1 – entropy) values,
(ii) discard too general terms,
These corpus specific stopword lists proved to be very effective with just the need to inspect very few words.

E. Global entropic weighting of MCQ answers

In our model of MCQ the question and each of the three answers are pseudo-documents [14]. Each pseudo-document "answer" will be compared with the pseudo-document "question" in semantic space of the training corpus. To produce these pseudo-documents it is recommended to use weightings which were used for the corpus [14]. Here being given the very low number of terms, their frequencies are non significant. We thus made profit of MCQ specificity: there are 3 concurrent answers for the same question. That makes it possible to again apply entropic global weighting (1 - entropy) to the 3 answers as a whole instead of considering them individually: the contrast of the terms differentiating more the answers is increased with the very beneficial effect expected on the results.

III. CORPORA, SEMANTIC SPACES AND MCQS

A. Corpora

4 corpora dealing with the 7th grade Biology program were built from two different sources: public scholar book (C) and private remedial course (M), either in a "basic" (Cb and Mb) format restricted to the content of the course, either in an extended (Ce and Me) version containing definitions and explanations of the concepts and some additional relevant information. Two chapters were extracted from the part "Functioning of the body and the need for energy": "muscular activity and need for energy"

and “Need of organs for dioxygen in the air”. The main characteristics of these 4 corpora are presented in Table I.

TABLE I.
CORPORA DATA

		Without Titles			With Titles		
	Docs	**Tokens**	**Words**	**Terms**	**Tokens**	**Words**	**Terms**
Cb	149	*11799	1944	1418	14298	1972	1433
Ce	425	*34331	4664	3174	40295	4729	3216
Mb	191	15169	1362	966	*19138	1377	976
Me	294	23549	1560	1072	*29663	1576	1083

Legend: Docs = documents (paragraphs in our case), Words = unique tokens (vocabulary), Terms = class of words after lemmatisation. * See section V. A.

The essential characteristics of the vector spaces filtered by the specific stop lists (cf. II.D), used in our experiments are presented in Table II.

TABLE II.
VECTOR SPACE MODELS PROPERTIES USING STOP LISTS

	Stopwords words => terms	**Words =>Terms**	**TxD Matrix Sparsity**
Cb	67 => 35	1877 => 1383	2,14%
Ce	83 => 39	4581 => 3135	1,00%
Mb	66 => 37	1311 => 939	3,42%
Me	64 => 34	1512 => 1049	3,02%

Appendix A exhibits, as an example, the stop list used with the Cb corpus.

B. Semantic spaces

The essential characteristics of resulting semantic spaces, used in the experiments of the section V, are presented in Table III below.

TABLE III.
BEST SCORE ACCORDING TO THE SEMANTIC SPACE DIMENSIONS

	Best Reduction		**No Reduction**		**Worst Reduction**	
	Dim	**Cor. Ans.**	**Dim**	**Cor. Ans.**	**Dim**	**Cor. Ans.**
Cb	14	27 / 31	149	18 / 31	148	16 / 31
Ce	13	25 / 31	425	17 / 31	3	15 / 31
Mb	5	22 / 31	191	14 / 31	191	14 / 31
Me	5	22 / 31	294	13 / 31	294	13 / 31

Legend: Dim = dimensionality, Cor. Ans. = number of correct answers.

C. MCQ31

Table IV presents statistics for the MCQ31 considered as a whole corpus. As there are 31 questions, the number of documents (here rather LSA pseudo-documents) is 124=31*(1 question+3*answers). The two last columns are the number of words and terms of MCQ31 present in the different corpora. These terms, and only them, are involved in building pseudo-documents to find the 31 correct answers to queries.

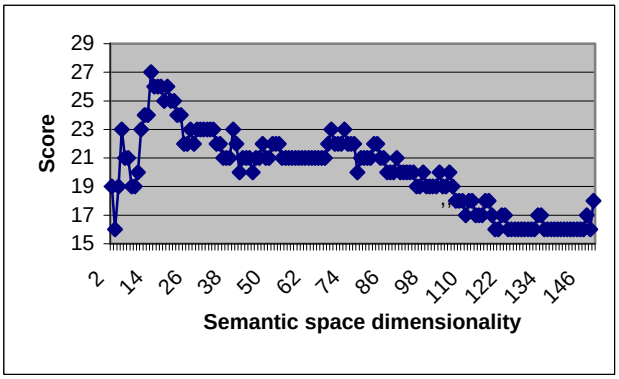

Figure 1: Number of correct answers as a function of the number of dimensions of the Cb semantic space.

TABLE IV.
MCQ31 VECTOR SPACE MODEL

	Queries Docs	**Tokens**	**Words**	**Terms**	**Words in corpus**	**Terms in corpus**
Cb	31 / 124	1311	307	255	224	188
Ce	=	=	=	=	241	203
Cb	=	=	=	=	225	187
Ce	=	=	=	=	230	191

IV. TYPOLOGY OF MCQ QUERY / ANSWERS

To conduct a useful experiment, we have to take into account the consistency between the basic assumptions of our model and MCQ data, namely:

- Each question and each answer of the MCQ is represented by a “bag of words” model.
- The correct answer is the one, from the three candidates, which has the higher similarity with the question.

This leads us to introduce a typology of questions / answers and reject the questions that are inconsistent with the model.

A. Out of subject questions

Two questions (n° 29 and 36) are rejected because they are related to topics which are not treated any more in our corpora, like the use of the cigarette and the associated harmful effects: words concerned are not even present in each corpus vocabulary.

B. Question / answers lack of correlation

The question n° 7 is characterized by an absence of correlation (meaning of the textual contents) between the question and the answers. This contradicts the basic assumptions of our model: « *Parmi les trois affirmations suivantes, une seule est juste. Laquelle ?* » (“Among the three following assertions only one is right. Which?”)… sounds as an universal wording for every query of the MCQ.

C. “Bag of words” indiscernibilities of answers

1) Hard indiscernibility

The lost of order of words due to the “bag of words” model can lead easily to indiscernible answers. We define

indiscernible answers as follows: when a correct answer and at least an incorrect answer have identical "bag of words" representation, hard indiscernibility occurs.

We call this indiscernibility "hard" to distinguish it from the "soft" one described later. For example, the question n° 24 leads systematically (whatever the corpus, with or without lemmatisation) to the following situation:

```
RMCQ24  best: 1 ref: 3
= 2, 3 indiscernible for a bag of words.
Question: [Quel] est le [sens] des [échanges] de
[gaz] [respiratoires] se [produisant] au [niveau]
des [alvéoles] [pulmonaires] ?
 1) Le [dioxyde] de [carbone] [quitte] l'[air]
    [alvéolaire] pour [rejoindre] le [sang].
 2) Le [dioxygène] [quitte] le [sang] pour
    [rejoindre] l'[air] [alvéolaire].
*3) Le [dioxygène] [quitte] l'[air] [alvéolaire]
    pour [rejoindre] le [sang].
```

eLSA1 has automatically pointed out that 4 questions (8, 24, 30, 35), are hard indiscernible for the "bag of words". It is illusory to seek to distinguish the correct answer among identical representations whatever the subsequent algorithms.

2) Soft indiscernibility

The previous indiscernibility was qualified as "hard" because it leads to indiscernibility between correct and incorrect answers. There is another kind of indiscernibility with less serious consequences. We define this kind of indiscernible answers as follows: when two incorrect answers have identical "bag of words" representations, soft indiscernibility occurs.

For example, the answers to the question n° 38 undergo this soft indiscernibility. With such soft indiscernible queries, *eLSA1* is able to choose the correct answer, so these questions are not discarded.

3) Stopwords and lemmatisation effect

Stopwords and lemmatization necessarily reduce the diversity of words in corpora. This reduction of the vocabulary, in spite of its very beneficial effects (as can be seen in the next section), can create indiscernibility: therefore indiscernibility detection of *eLSA1* remains activated during all our experiments as a protection.

Finally we have to reject 7 queries (n° 7, 8, 24, 29, 30, 35 and 36). Therefore, we use only 31 queries MCQ31 subset from the original 38 queries MCQ.

V. Relative influence of the parameters

A. Experimental conditions

Here we give the results of optimization by varying main parameters. Due to the interdependency between the parameters [24], we examined the discrepancy from the best score, one parameter at a time.

Since most authors confirmed that the best result is obtained with the logarithm as local weighting and the entropy (or more exactly 1 – entropy) for the global weighting [10], [7], the so-called log-entropy weighting was used.

Best scores (maximum number of correct answers) were obtained without paragraphs titles for the corpora Cb / Ce and with titles for Mb / Me (Table I): So "Title" in Table VI means flip-flop from best score tuning.

"Document Normalisation" means normalisation of columns document vectors, in Term * Document matrix, before log-entropy weighting (cf. V A).

"Frequency Normalisation" means that the sum of frequencies, components of document vectors, is normalized to 1 (empirical probabilities) before log-entropy weighting.

"Query 3-set entropy" in Table V and Table VI means that the weighting scheme describe in II E was used (or not) for the three answers associated to each query.

TABLE V.
BEST SCORE PARAMETERS SELECTION

	Cb	Ce	Mb	Me
Titles	-	-	+	+
Document Normalisation	-	-	-	-
Cooperative Lemmatisation	+	+	+	+
Frequency Normalisation	-	-	-	-
Query 3-set entropy	+	+	+	+
Stopwords	+	+	+	+
LSA truncation	+	+	+	+

In the case of corpora Mb and Me, if no cooperative lemmatisation is done, *eLSA1* detects occurrence of hard indiscernibility for the two first answers of question n° 6 even if the correct one is found by chance, because it is just the first of two answers with the same cosine:

```
RMCQ06  best: 1 ref: 1 :-)
=> 1, 2 indiscernible for a bag of words.
Question: [Quels] sont les [mouvements] des
[côtes] et du [diaphragme] lors d'une
[expiration] ?
*1) Les [côtes] s'[abaissent] et le [diaphragme]
monte.
 2) Les [côtes] et le [diaphragme] s'[abaissent].
 3) Les [côtes] se [soulèvent] et le [diaphragme]
s'[abaisse].
```

As the word "monte" ("raise") in the first answer in not present in the Mb and Me corpora, the "bag of words" representations are identical, it leads to hard indiscernibility described above (see IV.C.).

On the other hand if the joint lemmatisation occurs between the MCQ and the corpus, the words "monté" and "montée" of the corpus and the word "monte", of the answer, fall in the same class "monte". The "bags of words" representations of the answers 1 and 2 become discernible:

```
*1) Les [côtes] s'[abaissent] et le [diaphragme]
[monte].
 2) Les [côtes] et le [diaphragme] s'[abaissent].
```

This example shows the relevance of the joint lemmatisation, for not only for adding semantics when one works with relatively few words, but also in our case to limit the risk of parasitic phenomena as hard indiscernibility. Nevertheless, this does not mean that the correct answer will be found in this particular case.

TABLE VI.
RELATIVE PARAMETERS DECREPANCY FROM THE BEST SCORE

- Relative Influence +		Cb	Ce	Mb	Me
	Best score	27	25	22	22
	Titles	26	25	21	19
	Document Normalisation	24	23	20	18
	Cooperative Lemmatisation	24	22	-	-
	Frequency Normalisation	22	21	20	19
	Query 3-set entropy	22	22	18	17
	Stopwords	18	20	16	16
	LSA truncation	18	17	14	13

B. Discussion

Normalisations of documents and term frequencies have a negative effect on the results. The positive role of the recommended [24] pre-processing TFIDF features of the vector space model (before SVD) is confirmed. The optimal truncation (number of dimensions) of the semantic space and the stopwords list play a major role. Entropy weighting, specific to our problem, has a crucial influence (see discussion in section II.E).

C. About the best low dimensionality

The best score is obtained for relatively low values of the semantic spaces dimension (Table III, Figure 1), which may appear unusual. Wild et al. [24], who also obtained low dimensionalities, treat this question of the best dimensionality remained open since 20 years: for a long time "magic" values such as 100-300 [7] or even 50-1500 [19] were proposed in the literature. It now directs towards better founded statistical methods [8]. Wild et al. give for them better founded, simple methods in their base and their implementation, but apparently little known. The simplest of them is to consider a fraction (1/50) of the number of terms: in our case respectively 28, 63, 19, and 21, which appears to be a correct order of magnitude and is very satisfactory given the simplicity of implementation.

Let us now make some comments and assumptions concerning this point of our results:

- The fact that we could carry out an exhaustive scanning of the interval of dimensionality eliminated the phenomenon of the false optimum in results.
- The optimal dimension must not be completely independent of the task evaluating it, i.e. it does not rely solely on the corpus: in our case, there would be a filtering of the dimensionality by the low number of concepts denoted by the 31 MCQ questions.
- The high redundancy of corpora Mb and Me induces a relative poverty, from a numerical point of view, of concepts (conceptual focusing), and consequently of the number of important singular vectors (dimensionality), in comparison with corpora Cb and Ce.

VI. EXPERIMENT WITH STUDENTS

A. Participants and tasks

2 classes of 7th and 8th grades from Jean-Baptiste Say[1] high school (Paris, 75016) participate in the three phases of the experimentation : paper and pencil questionnaire, « classic » and « evidential » MCQ [5] and free answer questions [11] on the chapters : « Respiration » from the 7th grade biology program. Two equal 7th and 8th grades groups were formed according to the results of the paper and pencil questionnaire, one assigned to the « evidential » MCQ (n=26) and the other assigned to the « classic » MCQ (n=29). This « classic » MCQ has been supplied by « Maxicours », a private course enterprise, with whom we collaborate on the Infom@gic[2] project. This MCQ was composed of 38 questions, each of which has 3 candidate answers.

B. 7th and 8th grades results

The mean percentages of correct answers of 7th and 8th grades were very similar (79.5% and 81.2 %) and the distributions of their performances were close as shown by the significant correlation between their results ($r = 0.89$, $p < .01$). For example, the 9 questions that lead to the worst results (one standard deviation below the mean) are common to the 2 groups (n° 4, 6, 7, 8, 9, 14, 23, 24, 34).

C. eLSA1 indiscernibility of answers and student results

We should notice that the 7 questions eliminated by *eLSA1* (see section IV), are among the questions that lead to the lowest 7th and 8th grades' performances: 69 and 70% respectively. The mean percentage of correct answers of *eLSA1* with the Cb-149-14 semantic space (27/31 = 87%) is higher than the students' performances, while the results with the Ce 425-13 semantic space (25/31 = 81%) is equal to the students' performances. Performances of *eLSA1* with the Mb et Me semantic spaces (22/31 = 71%) are lower than the 7th and 8th grades' performances.

D. Correlation between eLSA1 and the students' performances

The correlations between the angle values corresponding to the cosines[3] affected by eLSA1 to the 3 answers of the remaining 31 questions and the frequency of choice of these answers by the 7th and 8th grades' are presented in

[1] Nous remercions Monsieur Patenotte, Proviseur de la cité scolaire, Madame Linhart, Principale adjointe ainsi que Mesdames Lopez et Lechner, professeurs de SVT de nous avoir permis d'utiliser les installations informatiques du Collège nécessaires au travail avec les élèves.

[2] Infom@gic est financé par le pôle de compétitivité à vocation mondiale de l'Île de France et coordonné par Pierre Hoogstoel (Thalès Communication) et Bernadette Bouchon-Meunier (LIP6, CNRS, UPMC).

[3] We substitute the angle of the vectors to their cosine, in a trend to be more linear and thus probably nearer to the spreading out of the answers of the pupils.

Table VII. These correlations indicate a significantly strong link between *eLSA1* and student performances.

TABLE VII.
CORRELATION BETWEEN *eLSA1* AND THE STUDENTS' PERFORMANCES

	Cb	**Ce**	**Mb**	**Me**
7th grade	.66	.56	.58	.47
8th grade	.59	.51	.54	.51
7th+8th grades	.63	.55	.57	.48

VII. CONCLUSIONS

We have demonstrated that LSA can be used to analyse MCQ and that its performances are similar to the students' results. Entropy weighting is proved to be a very influential parameter. The dedicated tool *eLSA1* enables us to build a typology of MCQ answers and to take into account their specificity. Thanks to *eLSA1*, it is now possible to continue our study on the effects of parameters and to extend our model to other tasks.

The strong correlations between *eLSA1* and student performances are encouraging despite of the simplicity of our model. This model could be improved easily to deal with more complex tasks.

Acknowledgments: We would like to thank Murat Ahat, computer science doctoral student from LAISC laboratory, EPHE-Paris, for his help in translating this paper and Nicolas Usunier of LIP6 for his kind proof reading.

APPENDIX A. STOPWORDS

67 stopwords / 35 stop lemmatized terms (bold words) list used for Cb corpus: « **ai, au,** auraient, aurait, aux, avait, **avec,** avoir, avons, **ce,** ces, **cet,** cette, **chez, comme, dans, de,** des, du, **en, est, et, grâce, il,** ils, **la,** le, les, **leur,** leurs, **ne, on,** ont, **ou, par, pas, permet,** permettant, permettent, permis, **peut,** peut-on, peuvent, **plus, pour, qu, quand,** que, **qui, sa, se,** ses, soient, soit, sont, **sous,** suis, **sur, très, un,** une, unes, **vers,** étaient, était, été, être ».

REFERENCES

[1] H. Baier, W. Lenhard, J. Hoffmann, W. Schneider, "SUMMA – A LSA Integrated Development System", *Special edition of Mobile Ad-hoc NETworkS (MANETS)*, to be published.

[2] S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. Harshman, "Indexing by Latent Semantic Analysis", *Journal. of the American Society of Information Science*, vol. 41, no 6, pp. 391-407, Sep. 1990.

[3] G. Denhière and B. Lemaire, "Representing children's semantic knowledge from a multisource corpus", *Proceedings of 14th Annual Meeting of The Society for Text and Discourse*, Chicago (USA), 1-4 August 2004, Lawrence Erlbaum Associates (publisher, Chicago), p. 10.

[4] G. Denhière, V. Hoareau, S. Jhean-Larose, W. Lehnard, H. Baïer, & C. Bellissens, "Human Hierarchization of Semantic Information in Narratives and Latent Semantic Analysis", *Proceedings of the 1st International Conference on Latent Semantic Analysis in Technology Enhanced Learning (LSA-TEL'07)*, Heerlen (The Netherlands), 29-30 March 2007, pp. 15-16.

[5] J. Diaz, "Diagnostic et modélisation de l'utilisateur : prise en compte de l'incertain", *thèse de doctorat*, Univ. P. et M. Curie, Paris, France, 23 septembre 2008.

[6] J. Diaz, M. Rifqi, B. Bouchon-Meunier, S. Jhean-Larose, and G. Denhière, "Imperfect Answers in Multiple Choice Questionnaires", *Proceedings of 3rd European Conference on Technology-Enhanced Learning (EC-TEL 2008)*, Maastricht (The Netherlands), 16-19 Sep. 2008, LNCS, P. Dillenbourg and M. Specht, editors, Springer-Verlag, vol. 5192, pp. 144–154.

[7] S. T. Dumais, "Improving the retrieval of information from external sources", *Behavior Research Methods, Instruments, and Computers*, vol 23, no 2, pp. 229–236, 1991.

[8] S. T. Dumais, "LSA and Information Retrieval: Getting Back to Basics", In *"Handbook of Latent Semantic Analysis"*, T. K. Landauer, D. S. McNamara, S. Dennis, W. Kintsch (editors), ©L. Erlbaum (publisher), pp. 293-321, 15 Feb. 2007.

[9] A. C. Graesser, K. Wiemer-Hastings, P. Wiemer-Hastings, R. Kreuz and Tutoring Research Group, "AutoTutor: A simulation of a human tutor", *Cognitive Systems Research*, vol 1, no 1, pp. 35-51, December 1999.

[10] D. Harman, "An experimental study of the factors important in document ranking", *Proceedings of the 9th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, Pisa (Italy), 1986, pp. 186-193.

[11] S. Jhean-Larose, V. Leclercq, J. Diaz, G. Denhière, and B. Bouchon-Meunier, "Knowledge evaluation based on LSA: MCQs and free answer questions", *Special edition of Mobile Ad-hoc NETworkS (MANETS)*, to be published.

[12] M. Kantrowitz, B. Mohit, and V. O. Mittal, "Stemming and its effects on TFIDF ranking", *23rd Annual International ACM SIGIR'2000 Conference on Research and Development in Information Retrieval*, Athens (Greece), 24-28 July 2000, pp. 357-359.

[13] W. Kintsch, "Meaning in Context", In *"Handbook of Latent Semantic Analysis"*, T. K. Landauer, D. S. McNamara, S. Dennis, W. Kintsch (editors), ©L. Erlbaum (publisher), pp. 89-105, 15 Feb. 2007.

[14] D. I. Martin and M. W. Berry, "Mathematical Foundation Behind Latent Semantic Analysis", In *"Handbook of Latent Semantic Analysis"*, T. K. Landauer, D. S. McNamara, S. Dennis, W. Kintsch (editors), ©L. Erlbaum (publisher), pp. 35-55, 15 Feb. 2007.

[15] NumPy. [Online]. Available: http://numpy.scipy.org/

[16] M. F. Porter, "An algorithm for suffix stripping", *Program*, vol. 14, no. 3, pp. 130-137, July 1980.

[17] M. F. Porter, "Snowball: French stemming algorithm", Oct. 2001. [Online]. Available: http://snowball.tartarus.org/algorithms/french/stemmer.html

[18] Python Software Foundation Home Page. [Online]. Available: http://www.python.org/about/

[19] J. Quesada, "Creating Your Own LSA Spaces In *"Handbook of Latent Semantic Analysis"*, T. K. Landauer, D. S. McNamara, S. Dennis, W. Kintsch (editors), ©L. Erlbaum (publisher), pp. 71-85, 15 Feb. 2007.

[20] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval", *Information Processing and Management*, vol. 24, no 5, pp. 513-523, Sep. 1988.

[21] G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing", *Communications of the ACM*, vol. 18, no 11, pp 613–620, Nov. 1975. (The article in which the vector space model was first presented).

[22] D. Tisserand, S. Jhean-Larose, and G. Denhière, "Eye movement analysis and Latent Semantic Analysis on a Comprehension and Recall Activity", *Proceedings of the 1st International Conference on Latent Semantic Analysis in Technology Enhanced Learning (LSA-TEL'07)*, Heerlen (The Netherlands), 29-30 March 2007, pp. 17-19.

[23] F. Wild, "An LSA Package for R", *Proceedings of the 1st International Conference on Latent Semantic Analysis in Technology Enhanced Learning (LSA-TEL'07)*, Heerlen (The Netherlands), 29-30 March 2007, pp. 11-12.

[24] F. Wild, C. Stahl, G. Stermsek, and G. Neumann, "Parameters Driving Effectiveness of Automated Essay Scoring with LSA", *Proceedings of the 9th CAA Conference*, Loughborough (UK), July 2005.