Knowledge evaluation based on LSA: MCQs and free answer questions

Sandra Jhean-Larose, Vincent Leclercq, Javier Diaz, Guy Denhière, Bernadette Bouchon-Meunier

Équipe CHArt : Cognition Humaine et Artificielle
École Pratique des Hautes Études
CNRS UMS 2809
Laboratoire des Usages en Technologies d’Information Numériques (LUTIN)

{sandra.jhean, guy.denhiere}@ephe.sorbonne.fr
Université Pierre & marie Curie-Paris 6, CNRS UMR 7606 DAPA, LIP6
{vincent.leclercq, javier.diaz, bernadette.bouchon-meunier}@lip6.fr
Knowledge evaluation based on LSA: MCQs and free answer questions

Sandra Jhean-Larose* — Vincent Leclercq** — Javier Diaz** — Guy Denhière* — Bernadette Bouchon-Meunier**

* Équipe CHArt: Cognition Humaine et Artificielle
École Pratique des Hautes Études
CNRS UMS2809
Laboratoire des Usages en Technologies d’Information Numériques (LUTIN)
41, rue Gay-Lussac
Paris, F-75005, France
{sandra.jhean, guy.denhiere}@ephe.sorbonne.fr

** Université Pierre et Marie Curie - Paris6, CNRS UMR 7606
DAPA, LIP6
104, Av. du Pdt. Kennedy
Paris, F-75016, France
{vincent.leclercq, javier.diaz, bernadette.bouchon-meunier}@lip6.fr

ABSTRACT. In this paper, we present an application using the SUMMA-LSA platform developed by Baier, Lehnard, Hoffmann & Schneider (this volume). SUMMA-LSA was used to evaluate biology knowledge of 7th and 8th grade students dealing with «The human body energy requirements». Student knowledge has been measured by means of classical and “evidential” multiple choices questions (MCQs) as well as open questions. The highly significant correlations between student answers and LSA cosine values opened encouraging perspectives in the development of e-learning systems that allows self-evaluations in real time and adaptation to learners’ knowledge.

RÉSUMÉ. Dans cet article, nous présentons une application utilisant la plateforme SUMMA de développement et de test de l’Analyse de la Sémantique Latente développée par Baier, Lehnard, Hoffmann & Schneider (ce numéro). SUMMA-LSA a été utilisé pour évaluer les connaissances d’élèves de 5e et de 4e dans le domaine de la biologie, «le fonctionnement du corps humain et ses besoins en énergie», à l’aide de QCM classiques et «évidentiels» et de questions
ouvertes. Les corrélations hautement significatives entre les performances des élèves et les valeurs de cosinus fournies par LSA ouvrent des perspectives prometteuses dans le développement de systèmes d’apprentissage en ligne, adaptés au niveau de connaissances des apprenants, et permettant des (auto)-évaluations en temps réel.

KEYWORDS: Latent Semantic Analysis, Biology, e-learning, Classical and «Evidential” MCQs, Free answer questions.


1. Introduction

Though new technologies developed learning by means of electronic support (e-learning), little progress has been made in the development of self-acquisition of knowledge, assumedly because of three issues: (i) it is difficult to automatically process online generation of natural language, (evaluation of answers to open questions for example) and then (ii) it is hard to approximate users’ knowledge online and then (iii) it is quite impossible to adapt the content of the electronic support to users’ knowledge in order to improve learning online. In this paper, we propose a solution to issues (i) and (ii). We used Latent Semantic Analysis (LSA: [1],[2]) to automatically process natural language answers and approximate learner’s knowledge. We built three tasks, classical multiple choice questions, evidential multiple choice questions, and open questions. Many studies in cognitive psychology have shown that remembering and learning from text depend on both textual characteristics and the cognitive properties of readers ([3],[11], [5],[6]). We assume that learners differ not only in the quantity of prior knowledge they possess, but also in the organization of that knowledge. According to previous results we obtained ([7], [8]), the knowledge of advanced subjects is organized in a hierarchical goal-subgoal structure, whereas that of intermediates and beginners is organized in a temporal-causal chain.

Baudet and Denhière ([9]) studied the structure of the representation in long-term memory of different systems. They first described the units
composing the system and the relations between these units in terms of a causal path and then prepared a teleological description of the system organized as a tree of goals-subgoals. The first description of a system considered the relations among actions, events and states according to the causal attribute ([10]). The sequences of actions, events and states expressed the chronological order of the system’s functioning. The second description considered these sequences of actions, events and states as a hierarchical structure of goals-subgoals. The nodes subordinate to the original node represented the subgoals of the system; attainment of these subgoals was a condition for the realization of the main goal of the system.

Several experiments ([11],[12]) studied the structure of the mental representation of a complex system in car mechanics in three groups of students with different levels of knowledge in the domain. To do so, the authors constructed four types of tasks that differed in the kind of activities required to retrieve information stored in memory: free interview, causal questioning, supplying the second event in a three-events sequence and a recognition task. The results were consistent with the hypothesis of a mental representation organized in a functional system and showed that the acquisition of knowledge may be characterized by (i) more information units stored in memory and (ii) greater structuring of the information in the system. The construction of a hierarchical, teleological organization presupposes the construction of an organization based on a temporal and causal ordering of actions. Although the beginner group could not differentiate between the system and the causal subsystems during these tasks, the group with high knowledge constructed a representation that was organized into only one functional system structured into subsystems. Based on this result, we assume that advanced learners have a mental representation of the knowledge domain organized in a teleological structure, whereas beginner and intermediate learners have a mental representation organized in a causal path. Having constructed a temporal-causal structure seems to be a necessary condition for constructing a hierarchical goal-subgoal structure. A beginner who is learning the functioning of a system memorizes the information in a linear fashion and organizes it in a causal path. With further knowledge, the beginner will become intermediate, and will eventually progress to an advanced level.
The work described in this paper is a multidisciplinary collaboration between psychologists, teachers and computer scientists. It deals with knowledge evaluation through a series of multiple choice and free answer questions. We chose a particular 7th grade biology lesson dealing with “The functioning of the human body and its energy requirements”, that can be generalized to treat any type of knowledge evaluation.

Participants were 7th and 8th grades students from a public school in Paris (collège Jean Baptiste Say). They first answer a knowledge questionnaire. We built subgroups of students as a function of prior knowledge and education levels. Then, students were asked to answer classical MCQs ([13]) or evidential MCQs ([14], [15], [16]) followed by a series of open questions (why and how questions).

2. Semantic spaces and prior knowledge

Many cognitive psychology studies have shown an effect of learners’ prior knowledge on learning performances ([17], [18],[19]). We built LSA corpora supposed to reflect different levels of education. Our goal was to vary levels of education and study the effects of that variation on the way LSA performed evaluation of students answers to MCQs and open questions[20],[21],[22]. We built two “basic” corpora (1 and 2 below) and two “elaborated” corpora (3 and 4), with reference to education standards:

– Corpus 1. Biology textbooks (7th grade).
– Corpus 2. Biology textbooks (8th, 9th and 10th grades)
– Corpus 3. Corpus 2, elaborated with definitions of concepts such as “flora”, “fauna” or “geologic”.
– Corpus 4. Corpus 2, elaborated with definitions of concepts and with encyclopaedic articles referenced in the original textbooks.

1. Contrat Infom@tic, pôle de compétitivité, Cap Digital, Commission Ingénierie des connaissances, Paris-Région, 2006-2009. We particularly thank Pierre Hoogstoel, Principal Investigator (Thalès L & J).
2. We want to thank Madame Linhart, co-principal, and Monsieur Patenotte, principal, as well as all those biology teachers in 7th and 8th grades for allowing us to complete our experimental program.
3. Experimentations

3.1. Platforms

To test the LSA semantic spaces we built, we used two interfaces: one to collect data from students answering tests, and one to analyse the answers using LSA.

The first interface we developed is a web platform allowing for the evaluation of students through traditional multiple choice questions (see figure 1), evidential multiple choice questions ([15],[16]) and free answer questions (see figure 2).

Students had to log into the platform and go through a small tutorial explaining the simple interaction, so that they could then begin answering the questions assigned to their particular profile. Questions were presented one at a time and students could browse them (using navigational arrows) without being forced to answer them. When the students were certain of their answer, they were asked to validate them.

The second platform we developed is a LSA analysis interface, also designed as a web interface. It is built on top of the same database as the interrogation platform in order to analyse the collected data.

---

3. We used an Apache web server with Tomcat, supporting J2EE technologies on the server side and AJAX interactivity on the client side.

4. In this article, we will not deal with the particularities of the answers given via evidential multiple choice answers; we will report on this analysis in future works.

5. Developed on J2EE technologies as well.

6. The database is handled using the Spring framework and Hibernate technology.
3.2. Protocol

Two 7th grade and two 8th grade classes of the secondary School J.B. Say (Paris, 75016) participated on the three phases of this experiment. First they were presented with a paper-pencil questionnaire. Secondly, the students had to answer a web-based evaluation, containing either a series of “classic” multiple-choice questions (MCQs) questionnaire or of “evidential” multiple-choice questions (Ev-MCQs). Eventually, they had to answer a set of free answer questions through the web-based evaluation platform. All questions dealt with the contents of the chapters “Muscular activity and energy needs” and “Organ dioxygen supply” from the “Functioning of the human body and its energy requirements” lesson of the 7th grade official biology program.

The paper-pencil questionnaire was built in 3 parallel formats. Each included 8 tasks more and more demanding.

8. We would like to thank Marie Mérouze and her colleagues for their help in the implementation of this questionnaire in the 7th grade biology curriculum proposed by the e-learning Maxicours Society.
Task 1: Participants had to judge whether 10 statements were “true” or “false”. Examples of statements: “oxygen consumed by organs comes from the air”; “muscles consume glucose only during physical exertion”. («Le dioxygène utilisé par les organes provient de l’air»; «les muscles consomment du glucose uniquement pendant un effort physique»).

Task 2: Participants had to find 10 words from their respective definition. For example: “name that is given to the thinnest vessels inside an organism”. Answer: “capillaries”. («Nom donné aux vaisseaux les plus fins à l'intérieur d'un organisme», réponse : «capillaires»).

Task 3: Participants were asked to make a sentence by using specific words. For example: “energy” “chemical reaction” “heat” “muscle”; possible answer: “In a contracting muscle, chemical reactions release energy that appears either as mechanical work or as heat.”
« Énergie | Réaction chimique | Chaleur | Muscle »; phrase possible : « Lors de la contraction d’un muscle, des réactions chimiques produisent de l’énergie et de la chaleur ».

- Task 4: By using “because”, participants were asked to connect pairs of sentences (number of sentences = 8).

- Task 5: Participants needed to categorize given properties of a concept as “true” or “false”. Example: “during respiratory movements, oxygen . . . (a) is exhaled in room’s ambient air; (b) is the only gas to enter respiratory system.” (« Au cours des mouvements respiratoires, le dioxygène . . . (i) est rejeté dans l’air ambiant », (ii) est le seul gaz à pénétrer dans l’appareil respiratoire », . . ).

- Task 6: Participants had to explain WHY. For example: “Explain why the blood getting out from the lungs is richer in oxygen than the blood that enters the lung?” (« Expliquer Pourquoi le sang sortant des poumons est-il plus riche en dioxygène que le sang entrant »).

- Task 7: Participants had to explain HOW. For example: “Explain how the oxygen is brought to the muscles” (« Expliquer comment le dioxygène est apporté aux muscles »).

- Task 8: Participants were asked to write down answers to 5 open questions. Example: “The smoke of cigarettes causes harmful effects on human health. The effects are observed on the respiratory system and on the other organs of the human body. Describe these harmful effects of the smoke of cigarettes and explain the mechanisms of its action on the respiratory system and on the other suffered organs.” (« La fumée de cigarettes a des effets néfastes sur la santé humaine. Ces effets s’exercent sur l’appareil respiratoire et sur d’autres organes du corps humain. Décrivez ces effets nocifs de la fumée de cigarette et expliquez les mécanismes de son action sur l’appareil respiratoire et sur les autres organes atteints »).

3.3. Methodology

3.3.1. Multiple Choice Questions

MCQ items can be analysed as to the proximity of each choice:

- to the other choices,
Knowledge evaluation based on LSA

– to the question lemma,
– to a portion of the text containing the correct choice’s explanation.

The analysis platform also enables us to select a test containing MCQ items and a semantic space, so that a similarity matrix containing these comparisons can be calculated (see figure 4).

![Knowledge evaluation based on LSA](image)

**Figure 4:** Similarity table for a MCQ test for a given semantic space.

Another functionality of the analysis platform permits us to select a question and to calculate the previous similarities on a set of semantic spaces.

Every line on figure 5 shows the similarity measures for a different semantic space that considers the same set of corpus documents, but that variates the number of dimensions used to calculate the semantic
Figure 5: Similarity table for a MCQ with all semantic spaces.

space. Columns 2, 3 and 4 show the similarity between the question lemma and each choice of the MCQ question, column 4 presents the similarity between the question and the portion of the text from the corpus containing the correct choice explanation, and columns 5, 6 and 7 show the similarity between this corpus extract and the 3 choices.

On figures 4 and 5, a color code (red and green) enables us to distinguish the similarities of the correct choice from the similarities of the other choices and indicates whether the semantic space would have correctly (green) or not (red) “answered” this question. When the similarities to the question for the good choice and another choice are not significantly different, the result is undetermined and we can distinguish two cases: either the greatest similarity corresponds to the third choice and it is treated like a wrong answer (in brown on figure 4); or there is a real undecidability (in yellow on figure 4) and this is treated as
additional information (the graphic on figure 7 shows the correct and undecidable answers to a set of MCQ for several semantic spaces).

By looking at figure 5, we can observe the importance of considering the right number of dimensions for a semantic space: the first 10 semantic spaces are not able to detect the correct choice (by obtaining a similarity greater than the one of incorrect choices), but the remaining spaces are.

Another analysis enables us to create a similarity matrix for the choices regarding a particular question, considering a particular semantic space (as shown in figure 6).

**Figure 6: Similarity table for the choices of a MCQ on a given semantic space.**

3.3.2. *Free Answer Questions*

Free answers can be analysed by using LSA to compare them to each other or against a golden standard response. As explained previously, students were asked to answer nine open questions in detail. The teachers of the Lycée JB Say kindly provided us with textual golden standard answers for every free answer question.

The LSA analysis platform presented in section 3.1 enabled us to select a particular test (set of questions) and a semantic space to calculate different LSA similarity measures. As shown on figure 8, the analysis
platform creates a similarity matrix presenting the comparison of the student answers against:

- the question lemma (first column “Question”),
- the golden standard (second column “Model”),
- the answers given by the other students (remaining columns).

4. Results

4.1. Multiple Choice Questions

The MCQs results were analyzed in great detail and published elsewhere ([13]). We tested the behaviour of several semantic spaces with regard to the MCQ test defined previously (see 3.2).

9. The students remain anonymous and are displayed with a code containing their grade and group.
Figure 8: Similarity matrix of the open answers given by the students.

The test (see 3.3.1) revealed some question profiles; two of these profiles showed up as irrelevant to an LSA analysis: the “undecidable” profile and the “out-of-context” one. Two of the 38 questions were classified as “out of context”. Indeed, there was no correlation between the question and the answers. For example, the content of such a question: “Among the three following sentences, only one is correct. Which one?” (“Parmi les 3 affirmations suivantes, une seule d’entre elles est juste, laquelle?”) would unlikely overlap any possible answers. The same problem can occur when possible answers are “yes” or “no”. We showed that rephrasing these questions could make them more relevant.

The following question is an example of “undecidable” profile:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 35 Similarity matrix of the open answers given by the students.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
— “Quelle est la composition du sang qui quitte les alvéoles pulmonaires ?” (What is the composition of the blood leaving the pulmonary alveoli ?)

(a) “Le sang est riche en dioxygène et en dioxyde de carbone” (The blood is rich in both dioxygen and carbon dioxide)

(b) “Le sang est riche en dioxygène et pauvre en dioxyde de carbone” (The blood is rich in dioxygen and poor in carbon dioxide)

(c) “Le sang est riche en dioxyde de carbone et pauvre en dioxygène” (The blood is rich in carbon dioxide and poor in dioxygen)

We see in this example that the correct choice (b) and a bad one (c) have a similarity of 1 in all the semantic spaces and cannot be distinguished. The use of LSA (in the current implementation) is irrelevant because the correct choice cannot be picked. By comparing to students answers, this profile of question seems to trouble them.

3 questions belong to « undecidable » profile. The following question is an example of “out-of-context” profile : “Parmi les trois affirmations suivantes, une seule est juste. Laquelle ? […]” (Among these three sentences only one is true. Which one ?) Here whatever the choices are, comparing them to the question is irrelevant as it doesn’t contains information about the subject of the question. The same problem occurs when the choices are just between “yes” or “no” to answer the question. We proved that rephrasing these questions can make them become relevant.

If this MCQ test is used mostly to evaluate the performance of semantic spaces (see figure 7), it can also give information about the question profiles, profiles that are not uncorrelated to student results. This information can then be useful for the test editor (in our case the Maxi-cours society) to improve the quality of the test or when editing a new test.

Below are two examples of questions and cosine values obtained between questions and answers

**Question N°37**

Question : “How is used the energy produced by muscles?” (Comment est utilisée l’énergie fabriquée par les muscles ?)
Knowledge evaluation based on LSA

A1: “It is used for the digestive system to work” (Elle est utilisée pour le fonctionnement de l’appareil digestif.)
A2: “It is used for the respiratory system to work” (Elle est utilisée pour le fonctionnement de l’appareil respiratoire.)
A3: “It is used for the muscles to work and is liberated as heat” (correct answer) (Elle est utilisée pour le fonctionnement des muscles et libérée sous forme de chaleur.)

Cosines between the question and the answers (LSA comparisons have been computed in French)

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>0.057</td>
<td>0.136</td>
<td>0.795</td>
</tr>
</tbody>
</table>

Question N°38

Question: “Which apparatus is used to measure the quantity of oxygen present in the environment?” (Quel appareil permet de mesurer la quantité de dioxygène dans un milieu?)
R1: “Thermometer” (Le thermomètre.)
R2: “Oxymeter” (correct answer) (L’oxymètre.)
R3: “Oscilloscope” (L’oscilloscope.)

Cosines between the question and the answers (LSA comparisons have been computed in French)

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>0.034</td>
<td>0.365</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Correlation between LSA and students’ performances.

The correlations between the cosines affected by SUMMA to the 3 answers of the 33 remaining questions and the frequency of choice of these answers by the 7th and the 8th grades were all high and significant with all spaces: they varied from .56 to .68, the correlations being higher with «basic» corpora than «elaborate» ones. These strong correlations are promising because the model could be easily improved.
4.2. Free Answer Questions

We also evaluated the performance of a semantic space for automatic correction of free answer questions by comparing student answers to golden standard (see 3.3.2).

Here is an example illustrating the results:

– **Question 4**: “Explain how the oxygen is brought to the muscles” (Expliquez comment le dioxygène est apporté jusqu’aux muscles.)

– **Golden standard**: “The blood brings to the muscles the oxygen and nutriments they need. Muscles, as all organs, take dioxygen O2 from the blood and reject carbon dioxide CO2.” (Réponse type : “Le sang apporte aux muscles le dioxygène et les nutriments dont ils ont besoin. Les muscles, comme les autres organes, prélèvent en permanence du dioxygène O2 dans le sang et rejettent du dioxyde de carbone CO2”)

– 7th Grade. A.7 : 0.3943

“The dioxygen is breathed in through the mouth or the nose, go through the trachea then the bronchus, the bronchioles, then enter the alveoli” (“Le dioxygène est aspiré par la bouche ou le nez, passe par la trachée puis par les bronches, les bronchioles et les alvéoles”).

– 7th Grade. A.13 : 0.5573

“The oxygen is brought to the muscles by trachea, bronchus, bronchioles and alveoli.” (Le dioxygène est apporté jusqu’aux muscles par des conduits : la trachée, les bronches, les bronchioles et les alvéoles.)

– 7th Grade. A.4 : 0.5454

“The oxygen is breathed in through the mouth or the nose, then goes through bronchus, bronchioles, then enters the muscles.” (Le dioxygène est aspiré par la bouche ou le nez, passe par la trachée puis par les bronches, les bronchioles puis entre dans les muscles.)

– 7th Grade. A.2 : 0.9201

“The oxygen is brought to the muscles by the blood.” (Le dioxygène est apporté jusqu’aux muscles par le sang.)

– 8th Grade. A.14 : 0.9330

“The oxygen is brought to the muscles by the blood, which goes through the pulmonary alveoli to take the oxygen and to throw out the carbon
The answers are ranked by increasing cosine value between the golden standard and students’ answers (from 0.3943 to 0.9339), this ranking corresponding also to teacher’s judgments.

– **Question 9**: “What are the characteristics of the alveolar wall that allow the passage of oxygen in the blood?” (Quelles sont les caractéristiques de la paroi alvéolaire qui favorisent le passage du dioxygène dans le sang ?)

– **Golden Standard**: “The oxygen enters the blood as the lung alveolar wall is very thin, highly vascularized, and the exchange area between air and blood is very large. The alveolar wall, very thin, richly irrigated by numerous blood capillaries, forms a wide exchange surface which allows the passage of oxygen from inspired air to the blood.” (Réponse type : Le dioxygène O2 passe dans le sang car la paroi alvéolaire des poumons est très fine, très vascularisée, et que la surface d’échange entre l’air et le sang est très grande. Cette paroi alvéolaire, très mince, richement irriguée par de nombreux capillaires sanguins, forme une vaste surface d’échange qui favorise le passage du dioxygène O2 de l’air inspiré vers le sang.)

– 7th Grade. A.5 : 0.2764

“The oxygen goes through the blood by a membrane.” (Le dioxygène passe dans le sang grâce à la membrane.)

– 7th Grade. A.13 : 0.3031

“The oxygen goes through the blood by capillaries.” (Le dioxygène passe dans le sang grâce aux capillaires.)

– 7th Grade. A.9 : 0.4208

“The oxygen goes through the blood by capillaries, very thin and very
numerous blood vessels.” (Le dioxygène passe dans le sang grâce aux capillaires, des vaisseaux très fins et très nombreux.)

- 8th Grade. A.4 : 0.5314

“The oxygen goes through the blood because the alveolar wall is large and permits to the oxygen and the blood to circulate more easily.” (Le dioxygène passe dans le sang car la paroi alvéolaire est grande pour permettre au dioxygène et au sang de circuler plus librement.)

- 8th Grade. A.2 : 0.6064

“The oxygen goes through the blood because the alveolar wall is a large and very thin surface. This large surface and this very thin alveolus lining permits to the oxygen to go from the blood to the organs.” (Le dioxygène passe dans le sang car la paroi alvéolaire est une grande surface très mince. Cette grande surface et cette paroi très fine permettent au dioxygène de passer du sang aux organes.)

- 8th Grade. A.15 : 0.9141

“The surface of all alveolar walls is very large and richly being supplied with many very thin blood capillaries. These characteristics make easier the crossing of the oxygen in the blood.” (Déjà, la surface de toutes les parois alvéolaires d’un poumon est très étendue mais elle est également très fortement irriguée par des capillaires très fins. Ces caractéristiques favorisent le passage du dioxygène dans le sang.)

- 8th Grade. A.16 : 0.9168

“The alveolar wall makes easier the passage of the oxygen into the blood because the wall is very thin, richly irrigated and with a wide surface.” (La paroi alvéolaire favorise le passage du dioxygène dans le sang car la paroi est très fine, largement irriguée et d’une très grande surface.)

This question was interesting because its answer, to be fully correct, must contain three properties of the alveolar wall: it is thin, it is richly irrigated and exchange surface is wide; only the two last students gave these three properties as explaining factors of easy passage of oxygen into the blood.

We compared the results of this automatic correction by LSA to those from the two biology teachers and four adult judges and we found that correlations were all high and significant and varied from .68 to .82, a
variability that is equivalent to the variability between adults judgments. It must be noted that the correlations were higher with “elaborated” corpora than with “basic” corpora, a trend opposite to the trend obtained with MCQs.

5. Perspectives

5.1. SUMMA

We were able to make use of the different fonctionnalities provided by the SUMMA server, both in semantic space construction and use for similarity measures. Even though SUMMA is a versatile and fully realized LSA solution, it presents some implementation limits that we inherited as we left the SUMMA LSA calculation untouched.

In further works we will try to test the influence of some of the parameters of construction of the semantic spaces (stop words, stemming, entropy, dimensions...) to see the impact on their performances and try to determine if an optimized set of parameters can be found. The performance of a semantic space can be estimated with the number of correct answers to a MCQ test (see figure 7) or with a more sophisticated measure.

On the other hand, we only used cosine as similarity measure. So we could check if the similarity measure has an impact on the “optimized set of parameters” by considering it as a parameter even if it does not intervene in the semantic space construction process.

As the SUMMA server has a “pluginable” architecture, we can customize most of the semantic space construction steps. For instance, we are currently evaluating a segmentation plugin that detects (from a pre-defined list) the common collocations of the corpus and consider them as a single “word” (or token).

5.2. Intelligent Tutoring Systems

Intelligent Tutoring Systems ([23],[24]) make use of knowledge representation and artificial intelligence techniques to provide a person-
Alized learning environment. ITS help students achieve learning objectives by diagnosing the student’s current state of knowledge and presenting the most appropriate learning resources (courses, examples, definitions, exercises, tests, interactive tools).

Learner knowledge acquisition is represented by the student model. The overlay model [25] (see figure 9) is the most common type of student model and represents student knowledge on top of a domain model, a concept network representing the course objectives.

An ITS obtains information about the student, either implicitly, by analysing student behaviour and the learning resources accessed, or explicitly, by proposing some evaluations.

The corpus of learning resources handled by an ITS can be used as the basis for the creation of a semantic space representing the domain knowledge of the courses. LSA can then be applied to various ITS tasks.

Figure 9: Student overlay model.
Some high-level tasks such as student answer evaluation [26] or student summary evaluation [27], have successfully used LSA in order to compare student textual input to some kind of golden standard.

As shown in figure 8, the answers given by a student can also be compared to the answers given by other students. Clustering algorithms can then be applied in order to obtain a partition of several subgroups of students that resemble each other with respect to their skills, so that they can be treated individually by fulfilling their particular needs.

We consider that LSA can be exploited in order to take care of some of the lower level tasks such as indexation and domain model creation ([28],[29]).

5.2.1. Indexation

Student interactions with a resource can be interpreted in terms of skills, thanks to an indexation (see figure 10) mapping learning resources to domain concepts. The interactions with a particular resource can be used to infer the current state of knowledge of the concepts indexing this resource. For instance, if a student answers correctly a particular question, the concepts being evaluated by that question are considered acquired and the student model is updated.

We successfully applied LSA in order to index a set of 30 multiple choice textual questions by a set of 21 concepts having a textual description [14]. The course objectives for the particular lesson evaluated by the set of questions were extracted from the official program provided by the national ministry of education; these 21 objectives were considered as concepts in the domain model. Our analysis platform enabled us to interrogate the semantic spaces on the SUMMA server in order to estimate different intensities.

– The global pertinence of a question to evaluate a given concept. We calculated the LSA similarity (cosinus) between the textual description of each objective and the textual concatenation of the question lemma and all possible choices of the question.

– The intensity of the relationship between a concept and each possible choice proposed in the question. We could have calculated the similarity between the concept description and each choice independently,
or we could have taken into account the concatenation of the lemma and each option. We chose the latter calculation, we considered that the choice by itself would not take into account the context of the question specified by its lemma.

We could then estimate the pertinence of a given question to evaluate a particular concept, and how the choice made by a student, correct or incorrect, was an indication that he had acquired or not this concept. Our student diagnosis framework uses of the Dempster-Shafer belief function theory to represent imperfect information and make inferences under uncertainty and imprecision, and the LSA similarities calculated through the analysis platform allowed us to attach conditional beliefs to the indexations of the questions by the concepts. Finally, the semantic space built from the ressource corpus can be used to identify the concepts that are going to be part of the domain model.

5.2.2. Domain modeling

The task of specifying the domain model on an ITS is cumbersome. Domain experts and pedagogic experts must hand-code this representation, following some kind of structural language allowing for the def-
inition of the domain concepts and their relationships (e.g. semantic networks [30], description logics [31], ontologies [32], etc.). Having a corpus of documents that describes the contents of courses handled by an ITS, and following some of the ideas from the Topic Model [33], we consider that LSA can help alleviate this task, by providing tools allowing to obtain semantic context or gist of the domain knowledge.

We are currently working on some functionalities for our analysis platform that will permit experts to semi-automatically define a neighborhood of words representing the domain knowledge of a given corpus. By giving a starting word and configuring some parameters (depth, stopwords, etc.) a network of related words can be obtained, annotated with the semantic similarities between the nodes. A set of operations are provided so that the experts can edit the resulting graph (nodes can be deleted, nodes can be paired, the graph can be extended by expanding a node further, etc.).

References


Knowledge evaluation based on LSA


