Of the Use of Natural Dialogue to Hide MCQs in Serious Games

Franck Dernoncourt[1]
(1) LIP6, 4 place Jussieu, 75005 Paris
franck.dernoncourt@gmail.com

ABSTRACT

Of the Use of Natural Dialogue to Hide MCQs in Serious Games

A major weakness of serious games at the moment is that they often incorporate multiple choice questionnaires (MCQs). However, no study has demonstrated that MCQs can accurately assess the level of understanding of a learner. On the contrary, some studies have experimentally shown that allowing the learner to input a free-text answer in the program instead of just selecting one answer in an MCQ allows a much finer evaluation of the learner's skills. We therefore propose to design a conversational agent that can understand statements in natural language within a narrow semantic context corresponding to the area of competence on which we assess the learner. This feature is intended to allow a natural dialogue with the learner, especially in the context of serious games. Such interaction in natural language aims to hide the underlying MCQs. This paper presents our approach.

KEYWORDS : Educational conversational agent, artificial intelligence, serious game, multiple-choice questionnaire, automatic assessment of free-text answer.

1 Introduction

We will define in this first part the key concepts of the article, namely the context of serious games and conversational agents, which are the solution we are exploring to address the problem of masking multiple choice questions.

1.1 Serious games

Serious games are a learning approach on fun. Learning can take place in the context of training, awareness or communication (Thomas, 2004). The serious games market has increased exponentially: up to $ 1 billion in 2004 (Sawyer, 2004), experts estimated it at about $ 10 billion in 2010.

Interacting through dialogue with a virtual agent helps to maintain focus and motivation of the player in a serious game. Currently, this dialogue, whether in the serious games or video games such as narrative video games (storytelling) and in most environments for human learning, consists of multiple choice: the player interacts with the game with multiple choice questions, which serve as dialogue.

The dialogue is subsequently very constrained, reducing the learning of the player, who can simply click on one of the possibilities without necessarily thinking in-depth. We believe that more flexible dialogue systems can be a relevant answer to this problem.

1.2 Conversational agents

A dialog is a verbal activity which involves at least two interlocutors and is used to accomplish a task or simply exchange words in a given situation. It is a coordinated sequence of actions (linguistic and non-linguistic) (Vernant, 1992).

The idea of human-computer interaction based on natural language is not new: it emerged in the 1950s with the Turing test. Nevertheless, this issue, at the conceptual and practical level, remains topical. There are for example annual competitions like the Loebner Prize (Loebner, 2003) or the Chatterbox Challenge to take a Turing test by imitating human verbal interaction, but no program is managed so far to reach the level of a human (Floridi et al., 2009).

To define performance criteria for conversational agents, we will consider the following four criteria pre-conditioning the development of an intelligent dialogue system proposed by (Rastier, 2001):

1. learning: temporary integration of information about the user,
2. question: request for clarification from the system,
3. rectification: suggestion of rectifications of questions, if necessary,
4. explanation: explanation by the system of a reply given previously.

Conversational agents fall into two main classes:

- conversational agents for non-task-oriented, i.e. conversation with the user on any topic with a friendly relationship often as ALICE (Wallace, 2009);
- task-oriented conversational agents, which have a goal assigned to them in their design.

The task-oriented conversational agents themselves are usually classified into two categories:

- service-oriented conversational agents, such as providing a consultancy service on a website, such as the virtual assistant Sarahde PayPal[1],
- educational conversational agents, whose goal is to help the user learn.

Our work focuses on educational conversational agents (tutor bots).

2 State of the art

After explicating the basic definitions in the previous section, we discuss briefly the state of the art of the architecture of conversational agents as well as the evaluation systems of free responses in more detail.

2.1 Architecture of conversational agents

Figure 1 shows an example of architecture of a conversational agent. The user enters a phrase that conversational agent converts into an abstract language, here Artificial Intelligence Markup Language (AIML): this translation is used to analyze the content of the sentence and make requests via a search engine in a database knowledge. The response is generated through an abstract language, AIML also here, the need to bring natural language before presenting it to the user.

However, this architecture is very rudimentary and rigid. For example, often must update the knowledge base to include knowledge about the user, particularly in the context of a tutoring business that requires monitoring of the achievements of the user as well as his motivation. A number of educational conversational agents have been designed and implemented, such as (Zhang et al., 2009), (De Pietro et al., 2005) (Core et al., 2006), (Pilato et al. 2008) or (Fonte et al., 2009).

Various architectures have been developed, here are the elements common to most of them:

- a knowledge base inherent to field,
- an answer manager,
- storage structures for exchanges in the form of trees especially in the educational conversational agents designed within a video game.

[1] https://www.paypal-virtualchat.com/

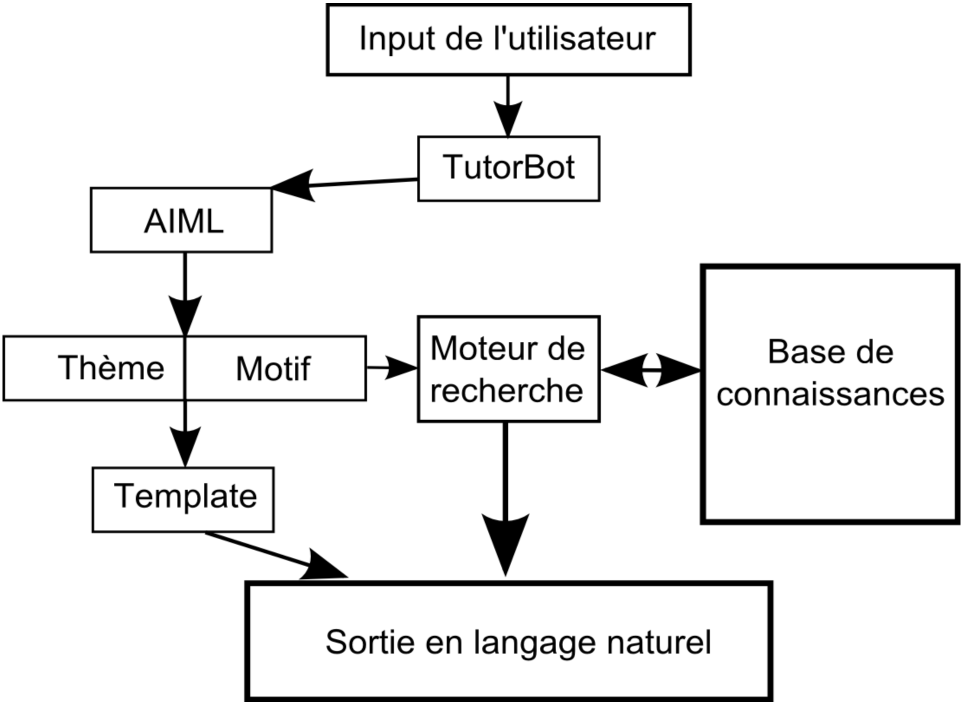


FIGURE 1 – Example of architecture of conversational agent (TutorBot)

Source: (De Pietro et al., 2005).

Although its simplicity and the relatively good performance of the conversational agents using it make it attractive, AIML is very limited and can be summarized in a simple pattern matching. Patterns of inputs (predefined sentences of the user) and outputs (responses of the conversational agent) is defined largely by expansion and a priori.

2.2 Evaluation systems for free-text answers

In parallel to the research on conversational agents, much work focused on evaluating free-text answers, that is to say, answers written in natural language that are given by the learners. This work is motivated by experimental results showing the boundaries of the MCQ as a tool for assessing the knowledge of learners (Whittington and Hunt, 1999), and its complementarity with free-text responses (Anbar, 1991). By knowledge we mean here and in the rest of the article not only the ability to recreate the information previously learned, but also the ability to make basic reasoning showing understanding of the subject.

For example, (Anbar, 1991) showed that students who excel in oral examinations will tend to have poor performance in the MCQ. Conversely, the MCQ results do not predict well the performance of the learner in the oral examinations.

Notwithstanding these well-known limitations of the MCQ, they still represent the most used tool to assess learners. This paradox can be explained by the much higher cost of alternative methods: while it is trivial to automatically correct MCQ, this does not also apply to other methods, which require, given current techniques, human interventions that are long and therefore costly.

The automatic evaluation of free-text response has its opponents, who point out that the fact that assessing a text is a task that is inherently complex and subjective. However, given that this subjectivity which results in a significant variation in scores among human raters, the system of automatic evaluation can at least be consistent in its subjectivity.

Early research on the automatic evaluation systems appeared fifty years ago. One of the notable projects was the Project Essay Grade, led by Ellis Batten Page at Duke University (Page, 1968). His work was based on the use of stylistic features of the response of the learner, such as word size and number of prepositions, to predict the human note of correction. In his later experiments (Page, 1995), this system seems to predict the human note of correction more precisely than some human correctors.

In the late 1980s, a new technique was developed to better understand the underlying concepts in a text: the latent semantic analysis (LSA) (Deerwester et al., 1988; Deerwester et al., 1990). This technique was initially used in the field of information retrieval and it was only later applied to the evaluation of free-text responses. The LSA would be easy to achieve if a word corresponded to only one concept, and vice versa. However, since in natural languages a word can have different meanings, a word may subsequently refer to different concepts, thus showing a wide ambiguity of the word. The LSA uses the context in which the word is used to remove the ambiguity, in other words to understand to which concept the word refers to in the given context. Figure 2 illustrates the purpose of the LSA.

The LSA does not take into account word order, syntactic or logical relations. In addition, it can be quite expensive from a computational point of view. Despite this, experiments have shown that the overall quality scores of a test given by experts are less accurate than the score resulting from LSA (Landauer, 1998). This surprising result is nevertheless to be put in perspective given the limitations of the LSA previously mentioned and obviously depends on the conditions of the experiment.

A completely different approach to the LSA was adopted by the Educational Testing Service (ETS). ETS is the largest private nonprofit organization for educational measurement and evaluation in the world. Organizing over 20 million exams annually (TOEFL, GRE, GMAT, etc..), ETS can have access to considerable corpus. For over twenty years, its R & D department has been working on solutions to automatically grade the candidates' answers. After trying to use the LSA to classify the responses (Burstein et al., 1996), ETS decided to move away and develop the technology c-rater (Leacock et al., 2003), C for content, which focuses on responses ranging from a few to a hundred words. C-rater is based on a preprocessing of the response shown in Figure 3. This preprocessing allows to show the answer in various linguistic features such as POS tags, lemmas of each word or the presence of negation. These linguistic features are then used to compare the candidate's response with a response model using a concept detection algorithm named Goldmap. Initially, Goldmap was based on a set of filtering rules determined binarily. Although this allowed easy understanding of the decisions, the binary rules induced a significant lack of flexibility. To address this problem, Goldmap now adopts a probabilistic approach based on the principle of maximum entropy for the detection of concepts and integrates a dozen ad hoc rules. The results look promising, according to the authors (Leacock et al., 2003). However, to our knowledge there is currently no standardized performance test to compare the different systems of automatic evaluation: it is therefore difficult to effectively compare different systems.

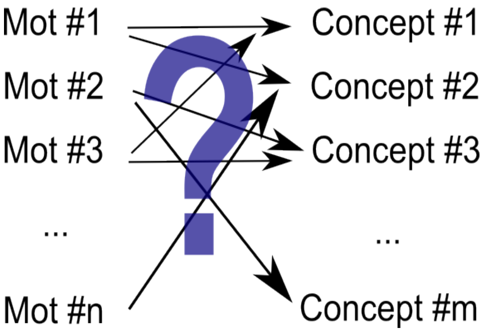


FIGURE 2 – Objective of the LSA: find concepts to which the words are associated.

In addition to the LSA and c-rater, it is interesting to note that many papers highlight the potential contribution of machine translation to the evaluation of free-text responses. A prime example is the method BLEU (Papineni et al., 2001). Originally designed to evaluate and rank the machine translation systems, BLEU method was successfully applied to the evaluation of free-text responses. The method relies on the comparison between the text and a set of candidate models texts. Applied to translation, the candidate text corresponds to the output of machine translation system, and text models correspond to the translations done by human experts. The score given by BLUE to the candidate text based on the number of N-grams in common between the candidate text and the texts models, which turns out to be an effective despite its simplicity. However it is very sensitive to the forms of models in the texts. When BLUE is applied to the evaluation of free-text responses, the candidate text corresponds to the answer of the learner, and text models correspond to typical answers given by teachers. Nevertheless, BLUE has important limitations, such as mismanagement of negations: a sentence denying a fact A would for example have almost the same score as a sentence affirming A.

Beyond the BLUE method, it is interesting to note that the field of translation and evaluation are in quest for the same ideal: finding a formalism in which the facts could be expressed independently of any natural language.

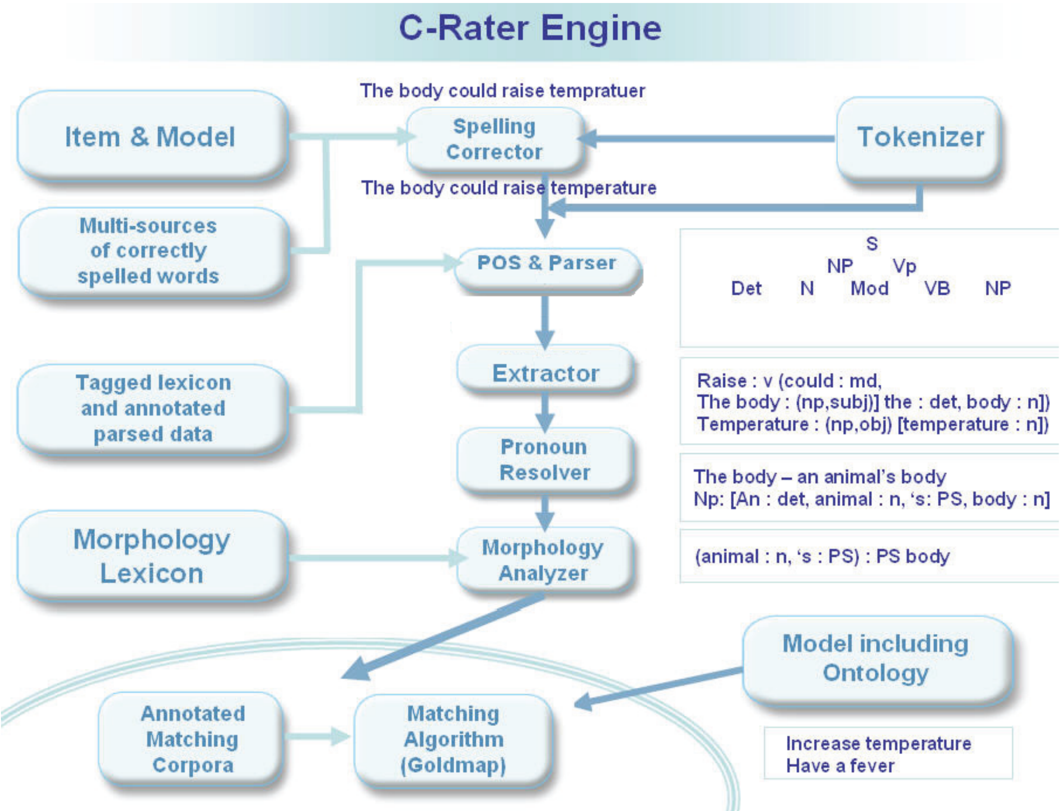

C-Rater Engine
The body could raise tempratuer
Spelling Corrector
The body could raise temperature
Item & Model
Tokenizer
Multi-sources of correctly spelled words
POS & Parser
S
NP
Vp
Det
N
Mod
VB
NP
Tagged lexicon and annotated parsed data
Extractor
Raise : v (could : md,
The body : (np,subj)] the : det, body : n])
Temperature : (np,obj) [temperature : n])
Pronoun Resolver
The body – an animal's body
Np: [An : det, animal : n, 's: PS, body : n]
Morphology Lexicon
Morphology Analyzer
(animal : n, 's : PS) : PS body
Model including Ontology
Annotated Matching Corpora
Matching Algorithm (Goldmap)
Increase temperature
Have a fever

FIGURE 3 – Architecture of c-rater. Source: Sukkarieh et al., 2009.

3 Approach

We saw in the previous section that much work focused on systems for assessing free-text responses. In this section we will highlight the features of our approach, in particular the characteristics relating to the assessment of free-text responses underlying with regards to MCQ and the environment of serious games.

3.1 MCQ particularities

Our work aims to give a grade to the answer of the learner. In our approach, we differ from traditional evaluation systems free-text answers from two main points:

- The answer of the learner is not rated compared to model answers, but is connected to an underlying MCQ,
- Interaction is possible with the learner, because the system has the form of a conversational agent.

Thus, research has focused on the evaluation of free-text responses but to our knowledge none have sought to evaluate a free-text response in terms of an underlying MCQ. We will therefore develop alternatives to the usual techniques (LSA, BLEU and c-rater) to adapt them to the use of MCQ.

The interest to reduce the user's response to an MCQ is multiple. On the one hand, many assessment tests now are in the form of multiple choice: we could therefore rely directly on the existing tests. On the other hand, the literature on automatic generation of MCQs from ontology is rich (Papasalouros et al., 2008): we could thus eventually have a comprehensive evaluation system directly from ontologies or even course books. The MCQ allows to bridge the gap between the knowledge base that provides the courses and tests given to the learner.

In a MCQ, the learner chooses one or more answers. In addition to the correct choices, there are also a number of incorrect choices. These incorrect choices can detect the presence of errors in the learner actively, that is to say by checking directly if the response contains no incorrect choice. The active detection of errors is absent from most assessment systems free-text answers because they are based only on comparison with model sentences. We can therefore identify these errors, while the conventional systems tend to ignore them.

The fact that the system is in the form of a conversational agent naturally allows us to respond more easily to situations where the answer of the learner fails to be directly evaluated by the system via the conversational agent, a new question may be posed to the learner to invite him to rephrase or clarify his answer. This interaction with the conversational agent can be compared to the oral tests with a human examiner and therefore avoids the disadvantages from traditional written examinations that are inherently static.

3.2 Insertion into a playful, serious environment

The simulation of a natural dialogue with the player in a video game dates back thirty years. The adventure game *King's Quest I: Quest for the Crown* who was developed by Sierra On-Line and published in 1984 is among the pioneers in the genre. It is only recently that the conversational mode was used for educational purposes, notably in the game Façade (Mateas et al., 2005), which we will briefly introduce in the next paragraph.

In Façade, the player is invited to a dinner during which takes place a marital conflict: the player's objective is to reconcile the couple. For this purpose, the player types sentences and both members of couples respond verbally. Figure 4 shows a screenshot in which the player asks the woman, Grace, if she feels upset vis-à-vis her husband Trip. By interacting with the couple, the player learns to better understand relationships.

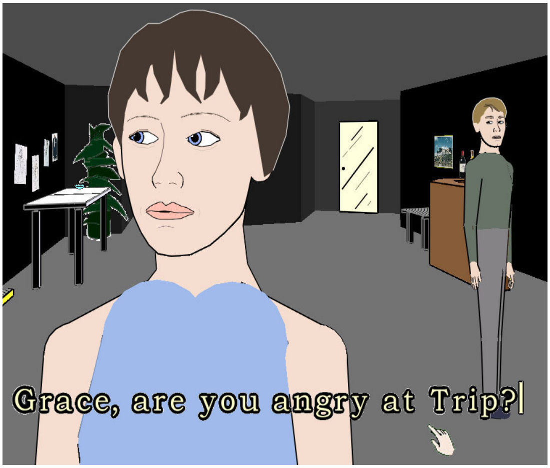


FIGURE 4 – Screenshot of game Façade. The player interacts with the couple.

However, until now, this kind of dialogue system based primarily on the identification of keywords by which the game's storyline fits and does not rely on underlying MCQ. In order to focus on the aspects of conversational agents and of MCQ, we integrate our system within the Learning Adventure platform[2] (Carron, 2010).

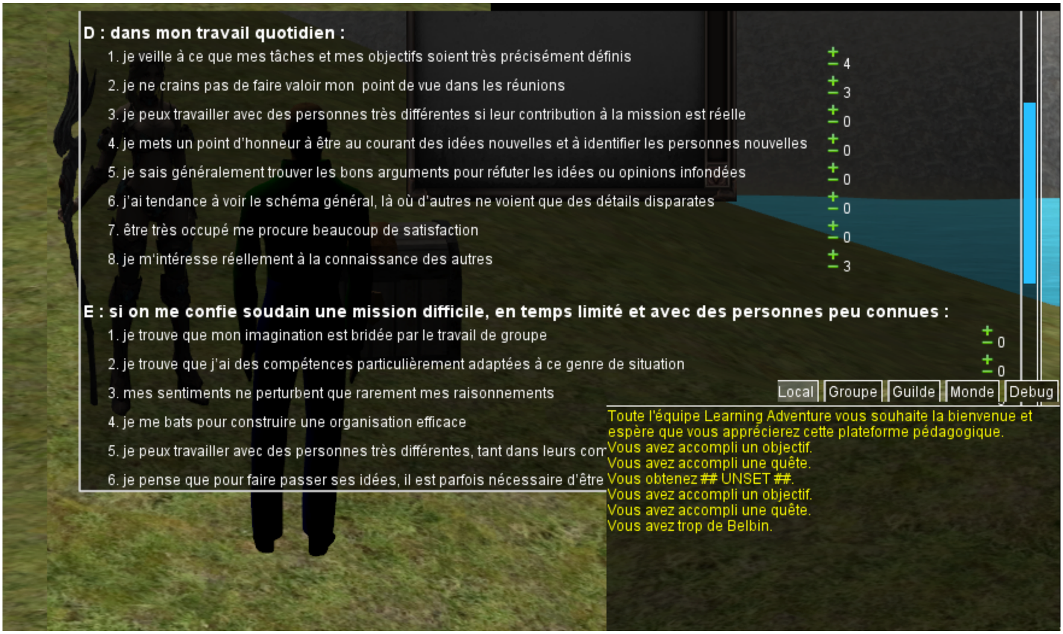
Learning Adventure is an open 3D environment with online multiplayer and where the learner must perform quests through numerous activities through which he interacts with the environment and other players. Emphasis is placed on the immersive nature of the game, like the current popular MMORPGs. Interaction with other players, i.e. with other learners, is an important aspect of the game as it contributes greatly to the motivation of the player: the MCQ is not a solitary game, but a social game, in which then enters the traditional mechanisms of the peer motivation (Dickey, 2007) (Kim et al., 2009).

Besides motivation resulting from the collaboration and competition between learners, the multiplayer aspect can also provide an opportunity for a human tutor to intervene in the game. Such intervention may have several objectives: to assist learners in tasks considered difficult, strengthen teacher-student relations by sharing a playful moment, etc..

The modality of online games has many other interests, particularly to ensure that the educational content is current, easily track the progress of individual learners and facilitate the deployment of new content.

[2] http://learning-adventure.eu

Figure 5 illustrates a MCQ that appears in the Game. Figure 6 presents the scenario editor, which includes the ability to easily add and edit MCQ without having any special computer skills. Our system aims to eventually make the MCQ invisible and use the editor scenarios to allow the teacher to include MCQ and other elements of the learning scenario.



D : dans mon travail quotidien :

1. je veille à ce que mes tâches et mes objectifs soient très précisément définis ± 4
2. je ne crains pas de faire valoir mon point de vue dans les réunions ± 3
3. je peux travailler avec des personnes très différentes si leur contribution à la mission est réelle ± 0
4. je mets un point d'honneur à être au courant des idées nouvelles et à identifier les personnes nouvelles ± 0
5. je sais généralement trouver les bons arguments pour réfuter les idées ou opinions infondées ± 0
6. j'ai tendance à voir le schéma général, là où d'autres ne voient que des détails disparates ± 0
7. être très occupé me procure beaucoup de satisfaction ± 0
8. je m'intéresse réellement à la connaissance des autres ± 3

E : si on me confie soudain une mission difficile, en temps limité et avec des personnes peu connues :

1. je trouve que mon imagination est bridée par le travail de groupe ± 0
2. je trouve que j'ai des compétences particulièrement adaptées à ce genre de situation ± 0
3. mes sentiments ne perturbent que rarement mes raisonnements
4. je me bats pour construire une organisation efficace
5. je peux travailler avec des personnes très différentes, tant dans leurs com
6. je pense que pour faire passer ses idées, il est parfois nécessaire d'être

Toute l'équipe Learning Adventure vous souhaite la bienvenue et espère que vous apprécierez cette plateforme pédagogique.
Vous avez accompli un objectif.
Vous avez accompli une quête.
Vous obtenez ## UNSET ##.
Vous avez accompli un objectif.
Vous avez accompli une quête.
Vous avez trop de Belbin.

FIGURE 5 – Capture d'écran de la plate-forme Learning Adventure (Carron, 2010)

Like c-rater ETS, we operate a pretreatment on the sentence of the learner, as shown in Figure 3: after an introductory phase corresponding to correction surface (spelling, punctuation, etc.), the response is analyzed lexically and grammatically tagged, then parsed, before proceeding to the step of referential disambiguation and morphological analysis. As these processes are assumed to be known by the reader, we will not detail them here. Technically, these steps (except the first that uses the spellchecker Hunspell) are based primarily on open-source, free library and NLTK (Bird et al. 2009) which offers many features of language processing.

Once the answer has been preprocessed, each word of the response is described by a set of attributes. We can then compare the different responses of the MCQ. The extracted data in the pretreatment contain many essential information for the comparison, such as the presence of negation.

To increase the robustness of the system, we plan to consolidate the results of the comparison based on the preprocessed response with statistical algorithms LSA and BLEU. This multiple comparison and the results will be discussed in detail in a future article, the purpose of this paper is to present the overall objectives, the context of the problem and the general structure of our system.

Conventional conversational agents use very little the usual tools of NLP, as shown in this following example from the official website of Cleverbot.com, which is nevertheless one of the most famous conversational agents:

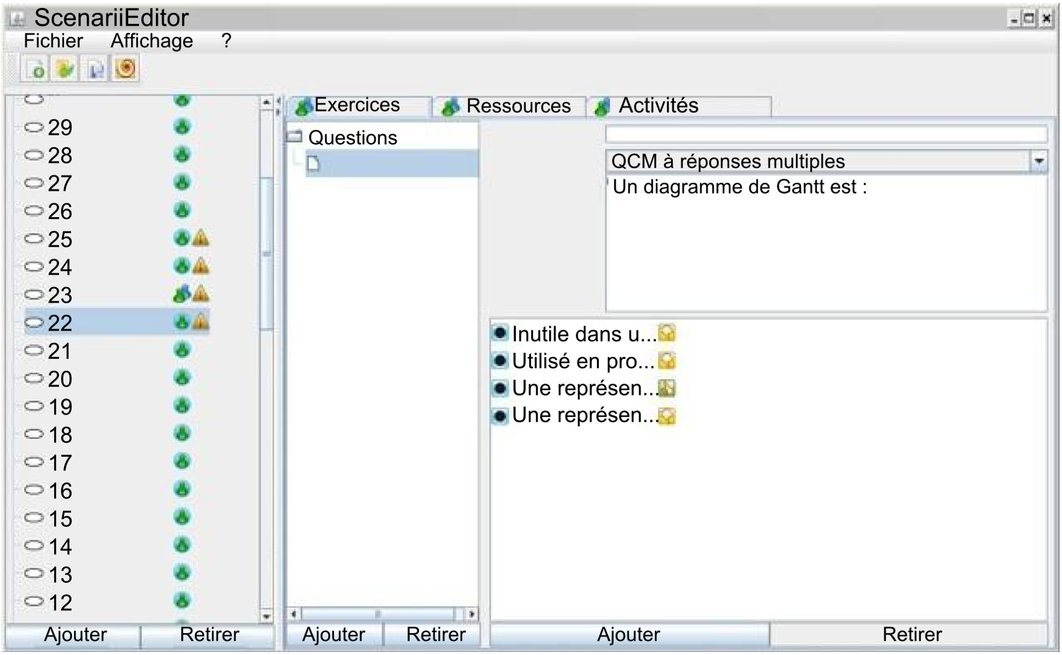
Human: Hi.

Cleverbot: Are you a boy?

Human: Yes I am a boys

Cleverbot: Are you a girl?

In this example, we see that the typo of the user who wrote "boys" instead of "boy" has clearly prevented the conversational agent Cleverbot to understand the sentence and therefore to propose a coherent dialogue.



QCM à réponses multiples

Un diagramme de Gantt est :

● Inutile dans u...
● Utilisé en pro...
● Une représen...
● Une représen...

FIGURE 6 – The scenario editor for Learning Adventure

By restricting the semantic field and stating its purpose, and we can integrate the usual NLP techniques in our conversational agent to make transparent the MCQ vis-à-vis the learner.

Finally, as shown by (D'Mello et al., 2010), the educational conversational agent is enhanced when the modality is oral and not written. Therefore, we use Dragon NaturallySpeaking 11, which is the leader in speech recognition and published by the company Nuance, and the software AT & T Natural Voices ® Text-to-speech to transmit the responses of the conversational agent in oral form. Note that these two applications are not free.

4 Conclusions and perspectives

This paper presented a novel approach to assess learners based on MCQ masked by a conversational agent in a serious game. The interactive nature of dialogue can make to the evaluation system a new dimension, allowing in particular to requests for clarification (Purver et al., 2003).

One challenge in research evaluation systems of free-text response is the absence of benchmarks, a lack which can be explained by intellectual property (Sukkarieh and Blackmore, 2009). Whatever the reasons, this gap is a problem for research in the field.

In recent months, three major initiatives MITX, and Coursera Udacity were launched. Their goal is to provide users with free online courses, which have already attracted more than 100,000 students. All three are based in large part (in addition to tests in which the programming code of the student is evaluated on a set of tests) on MCQ to evaluate learners, in the absence of more efficient systems. However, these MCQ are criticized as a limitation of this kind of online course where evaluation is fully automatic in order to ensure free access for a large number of learners. The application of masking MCQ is very important and will continue to increase along with the number of online courses.

Beyond educational contexts, such a system could also be used in other areas such as individual assistance, like the one provided by call centers is generally very scripted, i.e. following very inflexible scenarios, corresponding to a sequence of MCQ.

5 Acknowledgments

I want to thank my supervisor Jean-Marc Labat for his invaluable advice and the DGA for its financial support. I also thank Thibault Carron for his many ideas and assistance on Learning Adventure, for which he is one of the initiators.

6 References

ALHADEFF, E. (2008). Reconciling Serious Games Market Size Different Estimates. *In Futurlab Business & Games Magazine* - Numéro du 9 avril 2008.

BIRD, S., KLEIN, E. et LOPER, E. (2009). Natural Language Processing with Python. O'Reilly Media.

BURSTEIN, J., KAPLAN, R., WOLFF, S. et LU, C. (1996). Using Lexical Semantic Techniques to Classify Free-Responses. *In Proceedings of SIGLEX 1996 Workshop, Annual Meeting of the Association of Computational Linguistics*, University of California, Santa Cruz.

CARRON T., MARTY JC. et TALBOT S. (2010). Interactive Widgets for Regulation in Learning Games. *The 10th IEEE Conference on Advanced Learning Technologies*, Sousse, Tunisia.

CORE, M., TRAUM, D., LANE, H. C., SWARTOUT, W., GRATCH, J., LENT, M. V. et MARSELLA.

S. (2006). Teaching negotiation skills through practice and reflection with virtual humans. *Simulation* 82(11):685–701, 2006.

D'MELLO, S., GRAESSER, A. et KING, B. (2010). Toward Spoken Human-Computer Tutorial Dialogues. *Human-Computer Interaction,* (4):289--323.

DE PIETRO, O., M. DE ROSE et G. FRONTERA. (2005). Automatic Update of AIML Knowledge Base in E-Learning Environment. *In Proceedings of Computers and Advanced Technology in Education.*, Oranjestad, Aruba, August (2005): 29–31.

DEERWESTER, S., DUMAIS, S., FURNAS, G., LANDAUER, T., HARSHMAN, R., LOCHBAUM K. et STREETER, L. (1988). Brevet (US Patent 4,839,853).

DEERWESTER, S., DUMAIS, S., FURNAS, G., LANDAUER, T. et HARSHMAN, R., Indexing by Latent Semantic Analysis. *In Journal of the Society for Information Science*, vol. 41, no 6, 1990, p. 391-407.

DICKEY, M. D. (2007). Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research and Development*, 55(3), 253–273.

FLORIDI, L., TADDEO, M. et TURILLI, M. (2009). Turing's Imitation Game: Still an Impossible Challenge for All Machines and Some Judges—An Evaluation of the 2008 Loebner Contest. *Minds and Machines*. Springer.

LANDAUER, T.K., LAHAM, D., REHDER, B. et SCHREINER, M.E. (1997). How Well can Passage Meaning be Derived Without Using Word Order? A Comparison of Latent Semantic Analysis and Humans, *in Proceedings of the 19th Annual Conference of the Cognitive Science Society.*

LEACOCK, C. et CHODOROW, M. (2003). C-rater: Automated Scoring of Short-Answer Questions. *Computers and Humanities.* pp. 389-40.

LOEBNER, H. (2003). Home Page of the Loebner Prize - The First Turing Test. http://www.loebner.net/Prizef/loebner-prize.html [consultée le 03/03/2012].

KIM, B., PARK, H. et BAEK, Y. (2009). Not just fun, but serious strategies: Using meta-cognitive strategies in game based learning. *Computers & Education*, 52(4), 800-810. doi:10.1016/j.compedu.2008.12.004.

MATEAS, M. et STERN, A. (2005). Structuring Content in the Façade Interactive Drama Architecture. *AIIDE.*

PAGE, E.B. (1968). The Use of the Computer in Analyzing Student Essays. *International Review of Education*, 14, 210-224.

PAGE, E.B. (1995). The Computer Moves into Essay Grading: Updating the Ancient Test, *Phi Delta Kappan*, 76(Mar), 561-565.

PAPASALOUROS, A., KOTIS, K. et KANARIS, K. (2008). Automatic generation of multiple-choice questionsfrom domain ontologies. *IADIS e-Learning*, Amsterdam.

PAPINENI, K., ROUKOS, S., WARD T. et ZHU, W. (2001). BLEU: a method for automatic evaluation of machine translation. *In Proceedings of the 40th annual meeting on Association* *for Computational Linguistics*. 311—318.

PEREZ, D., ALFONSECA, E. et RODRIGUEZ, P. (2004). Application of the BLEU method for evaluating free-text answers in an e-learning environment. *In Proceedings of the Language Resources and Evaluation Conference* (LREC).

PILATO, G., ROBERTO P. et RICCARDO R. (2008). A kst-based system for student tutoring. *Applied Artificial Intelligence 22*, no. 4: 283-308.

PURVER, M., GINZBURG, J. et HEALEY, P. (2003). On the means for clarification in dialogue. *Current and new directions in discourse and dialogue*. Springer. 235—255.

RASTIER, F. (2001). Sémantique et recherches cognitives, *PUF* (2e éd).

SUKKARIEH, J. Z. et BLACKMORE, J. (2009). c-rater: Automatic content scoring for short-constructed responses. *Florida Artificial Intelligence Research Society (FLAIRS) Conference*, Sanibel, FL.

SAWYER, B. (2004). Serious Games Market Size. *Serious Games initiative Forum 01-04-2004.*

THOMAS, P., éditeurs (2010). *Actes de RJC EIAH 2010 (Rencontres Jeunes Chercheurs en Environnements Informatiques pour l'Apprentissage Humain)*, Lyon. ATIEF.

VERNANT, D. (1992). Modèle projectif et structure actionnelle du dialogue informatif. *In Du dialogue, Recherches sur la philosophie du langage*, Vrin éd., Paris, n°14, p. 295-314.

WALLACE. , S. (2009). Parsing the Turing Test, The Anatomy of A.L.I.C.E. Springer.

WHITTINGTON, D. et HUNT, H. (1999). Approaches to the computerized assessment of free text responses. *In Danson, M. (Ed.), Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough, UK.

ZHANG, H. L., Z. SHEN, X. TAO, C. MIAO et B. Li. (2009). Emotional agent in serious game (DINO). *In Proceedings of The 8th International Conference on Autonomous Agents and Multi-agent* Systems-Volume 2, 1385–1386.