How to Model a Likely Behavior of a Pedagogical Agent from a Real Situation

Mohamedade Farouk NANNE
Assistant Professor, Research Unit “Digital Documents and Interaction”, Dept. Mathematics and Computer Science, Faculty of Sciences and Techniques, University of Sciences Technology and Medicine, Nouakchott - Mauritania

***Abstract*—The aim of this work is to model the behavior verbal and nonverbal behavior of a Pedagogical Agent (PA) can be integrated into an Intelligent Tutoring System. The following research questions were posed: what is the nonverbal component of an educational communication? How to study this component for a computational model of plausible behavior of a virtual agent? What correlations between educational actions and the direction of gaze of a human agent? To carry out exploratory work, a methodological approach based on multi-modal video corpus study was adopted. Within a multidisciplinary team consisting of computer scientists and didactics of mathematics, an educational situation in which a virtual pedagogical agent is likely was developed. Dyadic interactions between teachers and learners late second early third (15-16 years) in a skills assessment interview in mathematics following the resolution of exercises by students with a mathematics software was filmed. A multi-level annotation scheme to annotate the observed behavior was proposed. The multidisciplinary research subject (ITS, Human-Machine Interfaces, Educational sciences, educational, linguistic etc.) due to the development of the coding scheme a delicate but important work given the wealth of knowledge from different disciplines. After a portion of the collected work annotation corpus statistical measures derived from annotations carried out suggest different strategies for teachers in terms of gaze direction depending on the learner profile and pedagogical actions. These measures have enabled to extract rules to control the nonverbal behavior of a PA.**

Keywords—pedagogical agent; nonverbal communication; behavior; corpus analysis

I. Introduction

The general theme of this work is the integration of a pedagogical agent in a Computing Environment for Human Learning (ILE). This theme, very extensive, can itself give rise to many research directions in computer science but also in Education Sciences, in Cognitive Sciences or Psychology. Set in this context requires a research topic to restrict and specialize this vast field of research. It was decided to focus this work on modeling the behavior of a pedagogical agent. This work should provide answers to a number of questions. In particular: in what situations a pedagogical agent can it improve communication between a student and an ITS? How to model these situations and interactions to specify the behavior of an agent?

To answer these questions, an empirical approach based on video corpus analysis compiled from situations of communication between humans was followed. The objective of this analysis is to evidence rules of behavior to model the behavior of the pedagogical agent.

In the field of educational agents, multimodal behavior of the agent is often limited expressiveness and is not based on a detailed analysis of multimodal behaviors of teachers, but rather on general rules derived for example from the literature in sociolinguistics. Although the use of educational videos developed in the field of ITS, they are rarely used as a resource for annotating and understanding of multimodal communication in the educational context. It was decided to follow a methodology that takes into account the sociolinguistic literature but especially to data from real situations. This methodology for annotating videos corpus and statistical analysis of these annotations was developed in the field of Human-Machine Interfaces.

A key objective of this work is to explore the application of this methodology to the field of ILE and specifically pedagogical agents. We were taken for this to be a video corpus based annotation analysis targets, define a schema and a markup annotation protocol, to analyze these annotations to finally provide a set of rules to control certain elements of non-verbal behavior of the agent.

II. Methodology

The multimodal behavior varies depending on the context and therefore must be routed so situated. The nonverbal behavior literature is rich in experimental methods and studies. However, from the perspective of human-machine interface design, it provides little specific enough knowledge to be used in a relevant way in a given situation. Approaches type digital corpus aim to provide such specific and detailed knowledge to a particular situation for building computational models. This methodology recommends starting with defining questions that wants answers using this method and also the definition of basic concepts. Then comes the development of an encoding scheme involving these and these concepts. The next step is to collect video footage from which the behavior will be defined. Once these sequences are recorded, an encoding scheme according to the already determined to be performed. This coding must be validated and after validation, statistical measures should be implemented to draw lessons for defining control rules of the pedagogical agent.

The particular methodology developed for modeling the behavior of a pedagogical agent located based on that general methodology. The main difference is that, according to general methodology, the coding scheme directs corpus of the collection and the concepts which are based on this scheme are fixed in advance, upstream of the collection. In the case of this work, exploratory work communication actual data collection in a learning situation served prior to the definition of many of the concepts. Given the very precise contextualization of this work, some of the underlying concepts to this particular study could not be determined a priori. These were extracted from viewing videos in preparatory work for the establishment of the coding scheme. The steps of the video corpus annotation methodology which underpins our research methodology will be detailed later.

A. Define the research questions and basic concepts

To achieve satisfactory results, it is important to define research questions. Their definition requires the determination of concepts and definitions which they refer.

B. Coding Scheme

A coding scheme is a set of parameters and indicators to be considered when observing sequence to annotate. It contains all the elements that could provide answers to research questions. It is therefore linked to the concepts and definitions which research questions refer. A coding scheme consists of annotation values. An annotation value is the means by which the annotator to describe his observation of an event. Those annotation values can be arranged in sets or groups of sets. These annotations values may be exclusive or not. The development of a coding scheme is a difficult task. [1] proposes to answer the following questions before developing a coding scheme: what are the coding schemes already defined in other studies that are similar or close enough to study to achieve? Do annotations values of these patterns are naturally defined and clearly structured? The links between these annotation values, research questions and the underlying concepts are they outstanding? If the answers to these questions are affirmative, [2] believe that there is a strong chance that the use of such encoding schemes as they are or by making changes would have a positive contribution.

C. Registration and behavior coding

Once the coding scheme defined, the annotator has methodological tools needed to describe his observations a series of practical and technical questions arise. These questions relate to how annotators will work and save their decisions (observations). Managing to combine the human and material resources in order to make observations and record them is the goal of this stage of the method.

[2] propose recording sequences of behavior observed in real time for later annotation. The advantage of video is the ability to perform multiple viewings but also the opportunity to review the event by slowing down the movements. Records can be stored for later use. They can also be seen by different observers and / or with different objectives. For example, different annotators can annotate different behaviors recorded. The same sequences can also be seen by an observer same but in different time for it confirms previous observations or for it to work in a progressive manner. The videos provide a literal observation of behavior , that does not allow the vision in real time. The recording tools are becoming relatively inexpensive, easy to use and information loss risks are minimal.

D. Representation of initial data

Whatever the technical tools used during annotation (paper and pencil tool, recordings involving cameras connected to laptops ...), annotations will ultimately processed by computer.

The use of an approach based on technology often has many advantages. Many computer systems have been developed for this purpose. One example is the software Anvil Mr. Kipp [3], the system developed by The James Long Company [4], the system developed by L. Procoder Noldus [5] and the system ObsWin developed by N. Martin, C. Oliver and S. Hall [6].

Generally, computer-coding systems are used to define the codes, their use and their characteristics. These systems safeguard the codes and the time associated with these codes in files. Annotators can edit these files and change them if they find that the initial codes are wrong. With appropriate equipment, video recordings can be controlled with the mouse. The annotator can then select episodes that appear interesting to him and watching them repeatedly.

The annotated data are stored in files for their computer processing (statistical calculations). Software has been developed can process these data such as SAS and SPSS but that these data are processed by the software, it must be organized according to forms required by that software. Depending on the desired treatment, it may be the other need to develop its own statistical analysis programs.

[7] defined a standard organization of annotation data to make them shareable. This standard is called SDIS.

E. Study of reliability of annotations

The records must be validated. In other words, it is necessary to assess the reliability of decisions made by annotator in his annotation. This step of the method of verifying the implementation of coding scheme by the annotator and prudence in decision-making. Several methods can be used to make this verification. If there is a standard that is supposed to correct, annotators can compare their observations with respect to this standard.

A set of video sequences is selected to serve as support for the validation step. Several annotators annotate these sequences independently of each other. A measure to calculate a coefficient of inter-judge agreement among different observers is then applied. This is the Kappa [8]. It can be applied to an isolated annotation value, or to a set of annotation values. The agreement is considered perfect if the coefficient is 1. If the agreement is greater than 0.75, it is considered excellent. If it is between 0.60 and 0.75, it is considered good. For agreements that are less than 0.60, the agreement is inadequate.

F. Statistical calculation on annotations

Statistical calculations on annotations enable a comprehensive and detailed view (depending on the nature of the calculations) of the studied indicators, their possible occurrences and co-occurrences. This step allows to define the elements constituting the studied behaviors and assembling these elements model the behavior will be defined.

III. Constitution of the Corpus and Annotation

Habert defines a corpus as a collection of language data that are selected and organized along linguistic lines and extralinguistic explicit job to serve as determined sample of a language [9]. Such corpus may be textual, audio, video or multimedia. The corpus can be used in several areas: language, interactional linguistics, ergonomic psychology, sociology, etc. A body is considered well-formed if some elements are taken into account in its constitution. These criteria vary according to the area studied.

A corpus should reach a critical size to be representative [10], but also to the statistical measures are reliable. A well-trained corpus must cover one language and more specifically one variation of this language. It is clear that this condition is applicable only on language use in corpus, because for example if we want to compare the gestures used by the French and those used by Anglo-Saxon in a given circumstance, we are obliged to collect videos two different languages. Two other criteria, like the previous criterion concerning the linguistic corpus. This is the time covered by the text corpus and language register.

For computer applications, we distinguish the corpus for learning and corpus for the test [11]. The corpus for learning are the corpus for study of indicator values (learn what the values of these indicators). The corpus for the test are the corpus collected while indicator values are defined in advance, the purpose of their incorporation is to validate or invalidate these values. In no way do the results or implement models from a training corpus on the same corpus as a test because if the values of the indicators sought are defined by the act of a body back to the same corpus could validate that attributed values are well within the corpus, which would validate the annotation process but would not validate the fact that these values are the most appropriate for the studied indicators.

The constitution of a corpus depends on the objective motivating the constitution both in terms of the situation at the level of collection itself. At the end of this constitution [2], a corpus can be considered sufficient to meet the aims of its constitution or not to be, which can result in a new collection. Out of all the footage, it often happens that all is not relevant in relation to the objectives of the corpus. Filter is then performed to select the data for the annotation. Filtering can also be applied in the case where the annotation means are not equal to the data collected such as the number of annotators or time devoted to the annotation work. Other reasons may lead filtering data collected such as the volume indicators looking. When using hundreds of indicators, the volume of data that can be mark up is different than when a few indicators are used. The indicators used in this work are detailed in the "coding scheme" section. A compromise feasibility and "reliability" appears. More data is annotated many more indicator values are confirmed which increases the reliability of values assigned to the indicators but the question of the feasibility intervenes quickly because it is expensive to annotate a number of important data.

This section describes how the corpus was collected and annotated. It describes the status of the data collection, the data collection, the selection of data on which the work is done, the coding scheme which has been developed and the annotation work.

A. Constitution of corpus

The annotation and corpus analysis are very time consuming. To select the most informative sequences with respect to the research questions, a fairly large number of videos was collected. In addition to this large size of the corpus, the other three criteria for a corpus is well formed (see the introduction to this section) are verified. The interviews are all recorded in French, in the same period (the period that covers all records is less than two months) and all components (textual, audio and visual) are recorded as part of post-resolution of interviews mathematical exercises for school students late third and early second intended to be applied in the same frame.

The communication situation that was filmed is that of interpersonal communication between a teacher and a student. The student before performing the test (which is individually performed on a computer). The cognitive profile and all Student Answer the exercises are then printed and sent to the teacher. The teacher had a few days to study these documents and prepare for the personal interview. At the beginning of each interview, the teacher prepared students to the situation of record so that it is not too tight and to reduce the influence of the presence of cameras so that this presence has no effect its possible interventions. The purpose of this interview is twofold. On the side of the student's goal is to help him overcome his difficulties in algebra and meet their specific questions by explaining its fragility and its levers and advising him to improve his level. The goal is to collect a corpus allowing us to specify the behavior of a pedagogical agent from real situations.

The teacher and student are sitting next to each other, facing a low table on which documents are placed. The table height is an important factor. Indeed, the two players must keep complete freedom of movement, at least as regards the upper body, so as not to force the use of different communication modes. This interview situation that the teacher and student are naturally positioned slightly turned towards each other. Two cameras were used, one positioned in the axis of the teacher and the other in the axis of the pupil, forming a triangle with approximately 60 degrees. Each camera records the entire scene but can effectively capture facial expressions, gaze and human gestures and postures on which it focuses. The height of the cameras have been set so as to properly observe both the facial expressions of people filmed (including when they look at printed material placed on the coffee table) and, with sufficient accuracy, the pointed places these documents (when a deictic gesture to a printed document is used).

Preliminary work viewing was conducted in order to identify a priori, informally, episodes considered particularly rich and interesting, firstly in terms of multimodal communication and, secondly, in terms of instructional events made by the teacher.

B. Coding Scheme

Like any experimental approach, an approach based on the corpus of study begins with the identification of one or more theoretical questions that wants answers. These questions and theoretical objectives and the study of existing work should direct the collection of video data and constitution of the coding scheme.

A multi-level encoding scheme from the collected corpus and theoretical elements from the literature was formed. This scheme has been retouched and validated by a team of researchers who come from many disciplines including linguistics, teaching and IT.

Many taxonomies to manually annotate observed behaviors have been developed at different levels, from the physical signs in different ways to more subjective levels related to the interpretation of such messages related to acts of dialogues or emotions [12] [13] [14] [15].

Before discussing the development of the coding scheme, it would be good to discuss the taxonomy of Pariès and DAMSL scheme whose application areas overlap that of this work.

The work of Pariès focus on the classroom teacher's communication. They are therefore different from this work since it is interested in the communication between two people (teacher-student) and not to the communication between a person and a group of people (teachers and students). This work is nevertheless interesting because most of the functions of the speech defined by Pariès remain relevant in the context of this study. Pariès defines two types of functions in the teacher talk: cognitive function and non-cognitive functions. Cognitive functions are related to the task to solve and mathematical knowledge. Cognitive functions are mentioned by Pariès: Distribution tasks Introduction of a sub-task, balance, Evaluation, Justification, Structuring. The non-cognitive functions are independent of the task in their formulation even though they may have an effect on the resolution. In this function category, Pariès class the following functions: Commitment in the job, Mobilizing students' attention, encouragement, Pooling the student's response.

Pariès taxonomy was developed using a corpus of six sessions recorded in four different classes of fifth, a refresher class in sixth and during a particular lesson in fourth class [16].

In this taxonomy, several functions will not be transposed to this study because the context is different. The communication between a teacher and a class is different from the communication between a teacher and a student. Thus functions such as "Mobilizing the attention of the student" for managing noise in a class and the "Pooling of student response" that allows to share the answer to all of a student by repeating, are not relevant in this situation.

Another taxonomy attracted the attention. This is DAMSL which was proposed by Core and Allen [17]. The first version of DAMSL was developed using the TRAINS corpus which gathers oral dialogues between two participants working together to solve a planning task. This corpus has 18 dialogs for 1524 set annotated by two people. This project was then used as the basis for other such body COCONUT [18], Monroe [19] and SWBD [20].

DAMSL has four main dimensions annotation :

- Communicative status: specifies if the statement is understandable and if it was completed and if the speaker seems able to convey what he meant. For example, a statement can be annotated as "self-talk", meaning that the speaker seems unable to transmit the information contained.
- Information-level: represents the semantic scope of the statement. In this category, four types of statements can be classified: statements that are interested in performing a task, those who speak of the management of a task and those interested in the management of communication. The fourth type includes statements that are not part of the first three types.
- Forward looking functions: defines how the statement will influence the discourse, context, actions of the speakers. In this category are the words of the speaker on the next actions to take, the effects on the statement following the conversation and statements containing requests.
- Backward looking functions: shows the relationship between the statement and previous statements of the speech. Such a statement could have a goal to meet, accept, reject, correct a previous speech. To do this it is necessary to specify the type of function that contains the statement and the previous part of speech to which the statement refers.

These categories are not mutually exclusive.

[21] defines two types of statements of a speaker: the locutionary acts and speech acts. By locutionary acts, it refers to what is said by the speaker, that is to say the fact to produce sounds. It defines the illocutionary act by the social act performed by producing a statement (producing a promise, an application, a statement ...). This work was pursued by linguists such as [22] offering four types of acts (act of utterance, propositional act, and perlocutionary act illocution) or [23] that mention five different functions the speech acts (assertive, the directive, the promissifs, expressive and statements). As part of this work, only the first level of analysis provided by Austin was used. This study focuses on speech acts represented by the speaker's intention, in this case, the teacher.

Taking into account existing studies mentioned above, and the particular situation that is the object of this study, a multi-level annotation scheme was developed. This scheme consists of five main levels: the intention expressed by the teacher, the means used by the teacher to express this intention, the strategies used by the teacher, emotional parameters and the nonverbal behavior. These levels will be detailed later.

Intention

In this level of annotation scheme, the illocutionary goals of the teacher are placed. [24] distinguishes three intentions of the teacher talk. These intentions are: information, evaluation and animation. These intentions have been identified in the corpus that was collected. Annotation values in this category are divided between the three types of intention.

It is inspired by the division made by Pariès into two types of functions: cognitive and non-cognitive. Among the non-cognitive functions, it is classified everything relating to the entertainment. For cognitive function, it is classified everything concerning information and evaluation.

All DAMSL categorization levels are present in the intention category. The Animate part is very close to the level "Information-level" of DAMSL and specifically the sub category "Communication-management". Other DAMSL levels are also present. This is "Forward looking functions" and "Backward looking functions" because some functions refer to functions related to the previous parts of speech (eg, "show cause") and other functions influencing the Following of the conversation (eg, "to ask"). The fourth level DAMSL (Communicative status) is also present in the annotation value "interrupt point"), especially with his "Abandoned" value. For parties to assess and advise on the most appropriate DAMSL level is that of "information-level" in its parts "Task" and "Task management".

Means

In this level of annotation scheme, the means used by the teacher to express an intention are classified. This is essentially linguistic means but also other resources that was identified in the corpus. The annotation values classified at this level three overlapping levels of DAMSL taxonomy. Some values can influence the continuity of the conversation (eg, "ask a question") and this is how they intersect the level "Forward looking functions". Others refer to previous parts (eg, the annotation value "sum") and thus overlapping the level "Backward looking functions". Some values relate to the management of the task and the task itself (eg, "read part of a support" and "focus on educational support") and the management of communication (for example, the annotation value "be humorous." These overlapping level "information-level" of DAMSL.

Strategies

In this category all the strategies used by the teacher to achieve a given objective are classified. Annotations values placed here cut across levels DAMSL. In particular level of duties "Forward looking functions" (eg, incitement affects the continuity of the conversation) and "Backward looking functions" (eg, the value annotation "correct errors" which refers to the past learning) and "information-level" in its part related to the management of the task and the task itself. We also note the similarity between the "Introduction of a sub task" of the taxonomy Pariès and annotation value "cut a complex question into sub-questions" that level of our annotation scheme.

Affective parameters

At this level of annotation scheme, the emotional parameters observed in the corpus are classified. This level intersects the DAMSL schema diagram for the functions "Forward looking functions" (eg annotation value "motivate learners" that would affect the rest of the conversation) and "Backward looking functions" (eg "mitigate speech" which refers to a previous part of speech). The annotation value "Encourage the learner" is found in Pariès in the "Encouragement" function.

Nonverbal behavior

This category includes the direction of gaze, gestures and facial expressions.

The gaze direction and mutual eye contact are very important in human-human interactions. The look not only helps manage practical tasks such as the exchange of turn in the conversation, but it also conveys a broad spectrum of information about the speaker, such as sociability, personality, culture . In Western culture, people often establish eye contact are perceived as more attentive, friendly, cooperative, confident, mature and sincere; while those who avoid the gaze of others are perceived as cold people pessimistic, defensive, evasive, indifferent and submitted. Several experiments have also highlighted the role persuasive than the eye can play [25] conducted a number of studies on human-agent interaction. They showed that the look is very important to make the plausible speech. It was also shown that individuals who cooperate mutually longer look than those in competition with each other [29]. People using eye contact receive more job offers after an interview, more help when they request it, and teachers who watch over the students make them more productive.

Pedagogical agents can use gestures to encourage, blame, empathy, praise (facial expressions, hand gestures), encourage them to ask questions (gesticulate, scratching their heads), etc. They can also show how to perform a physical action in a simulated 3D environment (hand gestures, postures).

Facial expressions are located not only in the geometry of the face, but also in terms of the texture of the face, and particularly the color (e.g., reddening) and brightness (tears, sweat etc.). The first model of blood vessels (including facial color depends) was developed by [26]. In this model, the authors define emotion as a function of two parameters: a parameter to control the muscle model, and a parameter to indicate the face of the texture variations.

IV. Annotation and Validation of Annotations

[27] defines a video annotation by the establishment of a textual description or digital video content, regardless of the part of the document in question. [28] defines the video annotation by the process by which textual or other information is associated with specific segments of documents. This information does not alter the original document, but are simply mapped therewith. Defined as an annotation is a generic term that includes both the addition of information without specific constraints, such as an exchange of e-mail about a video, or addition of information that must meet a defined format.

The objective of the annotation work is to apply statistical calculations on annotations. These annotations must be organized in a structure that allows for these calculations a robust, reliable and systematic. A tool has been chosen to organize these annotations in XML files which were then assembled in order to apply the desired statistical calculations.

Two annotation types can be distinguished: the manual annotation and automatic annotation. Automatic annotation proceeds by computer processing of video or audio signal to extract the elements. The software to use depends on the level of annotation to lead.

For example, the fact of calculating the size or duration of a video does not require complex treatments whereas the detection of specific elements in a video or comparing two images is complex. The manual annotation is necessary when automatic annotation cannot meet the objectives of the annotation work. For example, to extract the educational process initiated by a teacher, there is need to use this type of annotation.

The corpus analytical work was started by a transcription job of the teacher's speech with Praat, sound analysis software. The audio signal from the video has been imported into Praat then was manually segmented on the basis of the auditory content. After completing the transcript of the speech, it was imperative to use annotation software to import the files resulting from the transcription.

The annotations are defined in a hierarchical manner. There is need to use a tool that supports a multi-level coding scheme. According to the study [30], only Anvil tools, Media & Text Editors and Elan support this kind of coding scheme. But among these tools, only Anvil can import Praat transcription files.

The video selected was annotated with the Anvil software. This annotation work began with defining the specification file that contains the coding scheme described in the previous section. The TextGrid files produced by transcription with the Praat software was imported in Anvil. In these files, the teacher's speech is transcribed and is segmented into blocks. The blocks was analyzed one by one, and in the analysis, the annotation values constituting the scheme was checked one by one and each time the analyzed part of speech corresponded to an annotation value, the time to this correspondence and the interval marked on the appropriate track was defined.

The total duration of annotated sequences represents 54 minutes and 55 seconds. The duration annotation work of each part of speech is a function of its density. Part of speech tagged with multiple tags, takes longer to annotate a portion having a single label. It took us the final 54 hours and a half to complete the annotation work, what makes a average time was devoted to annotate a minute of video.

For now this work was limited to the gaze direction with respect nonverbal component of communication. This choice is explained partly by the importance of eyes in human communication, and secondly by a greater simplicity of this annotation component relative to other components of nonverbal communication such as such gestures. Working on the look has allowed carrying out the test work and methodological development which was the first. The annotation of this track was different from the other tracks in the coding scheme because it did not based on the blocks of the speech provided with importing files Praat but only watching videos. Other annotation work including gestures and facial expressions are underway.

The annotation work is a very costly process in terms of time. It was therefore impossible to validate all annotated sequences. It was decided to test the annotations on an extract of three minutes. The choice of this extract was adopted after viewing annotated excerpts. The first criterion was the annotation density values. Other criteria were also taken into account, such as the agreement between the annotation values for each block and the teacher talk.

This extract three minutes is not a lot for an annotated total of 54 minutes, but as in the case of exploratory work and given the available resources, the validation work should be limited. A second annotator took four hours to annotate this extract. This is not surprising given the density of the chosen extract.

To support the testing, this extract has been annotated by the teacher who gave back to the learner. In addition to its expertise in didactics and pedagogy, it is best placed to explain his speech, that is to say through this discourse determine the pedagogical act she wanted to accomplish.

Having entered this excerpt the two annotations were compared using the Kappa statistical measure.

Kappa is a statistical measure proposed by Cohen [8]. The agreement found between qualitative judgments or not, is the sum of a "random" component and a component of agreement "true". The Kappa coefficient, denoted κ, proposes to quantify the intensity or quality of the actual agreement between paired qualitative judgments. It expresses a relative difference between the proportion of observed agreement and the proportion of random agreement that is to say the likelihood of annotators agree luckily that is the expected value under the null hypothesis of independence judgments, divided by the amount available beyond random agreement. k corresponds to the maximum of that agreement that it would be corrected under the simple effect of chance.

The application of the statistical test of interrater agreement kappa gave very good results. The results are between 0.75 and 1, which corresponds according to the classification of [31] a good agreement to excellent. It is evident that this first validation is very limited. In a subsequent phase, these annotations may be validated using several annotators and more meaningful recording times. This work is exploratory work and seeks to highlight control rules that make the probable agent. This initial check shows that the coding process is relevant, if not actually validated.

V. Modeling and Implementation

The annotated corpus was the subject of a study by statistical analysis of annotations. Annotations are stored by ANVIL in XML files. A Java program into taking these XML files and to obtain a share of simple statistical evidence about the annotations (number of hits, average duration, etc.) and, secondly, the probability co-occurrence between the annotations and the direction of gaze was developed. These have allowed studying the correlation between the direction of gaze and other annotation values. In other words a search for every act done by the teacher, what is the probability of focusing the gaze to each of the centers of attention was performed. Thereafter the analysis algorithm implemented will be detailed.

The files from the annotation and whose extension is .anvil are XML files. The parser used to extract data from these files based on the library JAXP (Java API for XML Processing) which is used to parse XML documents through a Java program. After parsing the data are stored in arrays of vectors. Each annotation value is presented by an array of vector, vectors and a hundred paintings was created (number of annotation values encoding scheme). The size of the array is the number of annotation files passed to the parser. Following this step for each value of annotation all occurrences are stored in a table of vectors. The data on the occurrence of an annotation value are the start time, the end time of the period, the index of the event, the name of the annotation track.

Statistical measures applied to these vectors boards are explained below. Following these overall results and detailed calculations are stored in HTML files.

The steps of calculating the annotations are as follows:

Step 1: For each annotation value we calculated the total time of agreement with a single point of focus look. Eg for annotation value "Ask a question" we observe that the teacher focuses his gaze on the student for 40% of the time.

Step 2: The results of the preceding step may not be relevant in case of superposition of strongly linked annotation values. For example, if an annotation value A often appears simultaneously with an annotation value B, and if calculations show that these two annotation values have the same focal point of the eye, we can not distinguish whether this is the result of the performance of A or the completion of B. It is therefore necessary to study the dependence of the two values and it is the objective of this step.

Step 3: It is also possible that the influence is not that of an annotation value to another, but that of a block of annotation values on the value studied. For this we calculated all intersections annotation values were encountered in the annotated episodes. Specifically for each value annotation we studied its intersection with the blocks of annotation values that overlap.

Step 4: We also calculated for each annotation value, the focal point of the corresponding eye if there is no intersection with other annotation values which can be called "exclusive intersection "between a focusing point of view and an annotation value.

We distinguish annotations in three broad categories: those for which the co-occurrence values are stable from one student to another, those that seem to be related to the cognitive level of the student and those for which these values vary independently of the cognitive level of the student. Cognitive profiles of products allow students to place students in relation to the knowledge expected at this level of competence in elementary algebra. We hypothesize that this "level" of the student is an important factor explaining some differences in the analysis of the corpus. This hypothesis will of course be consolidated by a study on a larger panel of students.

VI. Conclusion and Perspectives

In this work the role of the agent to conceive is that of a teacher giving a feedback to a learner on a mathematical evaluation. Not to miss elements that may be essential, it was useful to make use of scenes back and coordinate this information coming from real situations with theoretical knowledge from the scientific literature. Thus these sessions were organized with students the school level at which the application that will host the agent is intended. There is no question here of a noted work but rather a work that helps the student to better understand his knowledge of algebra. These sessions involved two teachers and eight students. Thus corpus of teaching communication situations with strong emotional charge was collected.

Once the corpus is collected, the question that arose was how to study nonverbal component in educational communication for a plausible computational model of a pedagogical agent. To study this component, the context of ILE video corpus analysis methodology developed for the HMI was adopted. This methodology begins with the definition of objectives. To secure these objectives, the screenings of videos from the collection of the corpus and a study of existing taxonomies and annotation schemes were made. A multi-level annotation diagram modeling the actions of the teacher as the general level of ordinary communication at the level of educational communication was developed. Teacher's intentions, linguistic or other means he uses to express themselves, teaching strategies implemented, the teacher's emotional parameters were taken into account in establishing this pattern of annotation. For non-verbal communication, work is limited to the gaze direction. The viewing was carried out by several people from several disciplines: computer science, pedagogy, didactics and linguistics. This work was also used to select the set of episodes judged interesting for the annotation. Following this scheme annotation all selected sections were annotated. Given the time and energy that takes such work, validation could be made only on a limited sample. But this symbolic validation has given excellent results.

The next stage of work was to provide answers to the question what correlations between verbal communicative behavior and the viewing direction of a human agent. The communicative behavior in its verbal aspect was annotated, and the direction of gaze of the teacher. Just make calculations to study the correlation between this component of nonverbal behavior and different values constituting verbal behavior. A Java program was conducted for this purpose.

At the end of this step we get for each annotation value the probability that the light of the teaching points towards one of the focal points of the eyes. Thus, a model has been proposed to control the pedagogical agent. A working implementation of the rules defined in this research is underway on the agent litebody [32]. After setting these rules and those that will result from the annotation of gestures and facial expressions, a validation work will be done and guidance to struggling learners is planned.

ACKNOWLEDGMENT

Most of this work was conducted as part of a doctoral thesis at the University Paris 8 in collaboration with Jean-Hugues Réty, Elisabeth Delozanne, Jean-Claude Martin, Brigitte Grugeon, Catherine Pelachaud and Nelly Bensimon.

REFERENCES

[1] BAKEMAN R. Behavioral Observations and Coding. In H. T. Reis & C. K. Judd (Eds.), Handbook of research methods in social psychology (pp. 138-159). New York: Cambridge University Press, 2000.

[2] BAKEMAN R., DECKNER D.F, Analysis for Behavioral Streams, In Teti, D. M. (Ed.), Handbook of Research Methods in Developmental Psychology. Oxford, UK: Blackwell Publishers, 2004..

[3] KIPP M. Gesture Generation by Imitation. From Human Behavior to Computer Character Animation. Florida, Boca Raton, Dissertation.com, 2004. http://www.dfki.de/~kipp/dissertation.html

[4] LONG J. Video coding system reference guide. Caroga Lake, NY: James, Long Company, 1996..

[5] NOLDUS L.P.J., TRIENES R. J. H., HENRIKSEN A. H. M., JANSEN H., et JANSEN R. G. The Observer Video-Pro : New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. Behavior Research Methods, Instruments, and Computers, 32, 197-206, 2000.

[6] MARTIN N., OLIVIER C., et HALL S. ObsWin: Observational data collection and analysis for Windows. CTI Psychology Software News, 9, 14-16, 1999 .

[7] BAKEMAN R., and QUERA V. Analyzing Interaction: Sequential Analysis with SDIS and GSEQ. New York: Cambridge University Press, 1995.

[8] COHEN J., (1960). A coefficient of agreement for nominal scales", Educ. Psychol. Meas.: 20, 27-46.

[9] HABERT, B. 'Representative corpus, what,why and how ?', Des corpus représentatifs : de quoi, pour quoi, comment ?, in M. Bilger (éd.),Linguistique sur corpus. Études et réflexions, Perpignan, Presses Universitaires de Perpignan, pp.11-58, 2000.

[10] HABERT B., FABRE C. et ISAAC F. 'From the writing to digital: build, standardize and exploit the electronic corpus', De l'écrit au numérique : constituer, normaliser et exploiter les corpus électroniques, Paris, InterEditions, 1998.

[11] RASTIER F. 'Epistemological stakes of corpus linguistics', Enjeux épistémologiques de la linguistique de corpus. In G. Williams, Acte des deuxièmes Journées de Linguistique de Corpus, Presses Universitaires de Renne, Lorient, 2002.

[12] BAKEMAN R., et GOTTMAN J. M., (1997). Observing Interaction. An introduction to sequential analysis. Second edition., Cambridge University Press 1997.

[13] POGGI I. Mind Markers. In M. Rector, I. Poggi and N. Trigo. Meaning and Use. Oporto, Portugal University Fernando Pessoa Press: 119-132. 2002

[14] HARRIGAN J.A., ROSENTHAL R. and SCHERER K. The new handbook of methods in nonverbal behavior research, Oxford University Press, 2005.

[15] MARTIN, J. C. Multimodal Human-Computer Interfaces and Individual Differences. Annotation, perception, representation and generation of situated multimodal behaviors. Habilitation à diriger des recherches en Informatique. Université Paris XI, 2006.

[16] PARIES C. M. 'Comparing practices of teachers of mathematical relationships between teachers and speech potential student activities', Comparaison de pratiques d'enseignants de mathématiques relations entre discours des professeurs et activités potentielles des élèves. In Recherches en didactique des mathématiques : vol. 24, no2-3, pages 251-284, 2004.

[17] CORE M. G. and ALLEN J. F. (1997). Coding Dialogues with the DAMSL Annotation Scheme. AAAI Fall Symposium on Communicative Action in Humans and Machines, Menlo Park, California,American Association for Artificial Intelligence. citeseer.ist.psu.edu/ core97 coding. Html

[18] EUGENIO B.D., JORDAN P., et PYLKKANEN L. The coconut project: dialogue annotation manual. Technical report, University of Pittsburgh, 1998.

[19] STENT A. The monroe corpus. Technical Report 728, University of Rochester. 2000.

[20] JURAFSKY D., BATES R., COCCARO N., MARTIN R., METEER M., RIES K., SHRIBERG E., STOLCKE A., TAYLOR P., et ESS-DYKEMA C.V. Switchboard discourse language modeling project final report. Rapport Technique Summer Research Workshop Technical Reports 30, Johns, Hopkins University, Baltimore, 1997.

[21] AUSTIN J.L. How to Do Things with Words. Cambridge (Mass.) 1962.

[22] SEARLE J. Speech Acts. Cambridge University Press. 1969.

[23] SEARLE J. et VANDERVEKEN D. Foundations of Illocutionary Logic. Cambridge: Cambridge University Press, 1985.

[24] DABENE, L. 'For a taxonomy of meta-communicative operations in foreign language classroom', Pour une taxinomie des opérations métacommunicatives en classe de langue étrangère. COSTE, D. (éd.) Interactions et enseignement/apprentissage des langues étrangères, Etudes de Linguistiques Appliquée, 55, 39-46, 1984.

[25] CASSELL J., THORISSON K.R. The power of a nod and a glance: Envelope versus emotional feedback in animated conversational agents. Applied Artificial Intelligence, 13, 519-538, 1999

[26] KALRA P., MAGNENAT-THALMANN N. Modeling of Vascular Expressions in facial Animation, Computer Animation, pp. 50-58, 1994.

[27] ASSFALG J., BERTINI M., COLOMBO C., DELBIMBO A. Semantic Annotation of Sports Videos. IEEE MultiMedia 9(2): 52-60, 2002.

[28] EMOND B., BARFURTH M.A., COMEAU G., BROOKS M. 'Video annotation technologies and pedagogical applications' , Technologies d'annotation vidéo et leurs applications à la pédagogie du Piano. In : Recherche en Education Musicale. 2006.

[29] ARGYLE M. Bodily Communication. New York:Methuen & Co. Ltd. 1988.

[30] ROHLFING K., LOEHR D., DUNCAN S., BROWN A., FRANKLIN A., KIMBARA I., MILDE J.-T., PARRILL F., ROSE T., SCHMIDT T., SLOETJES H., THIES A., WELLINGHOFF S. Comparison of multimodal annotation tools – workshop report. Second Congress of the International Society for Gesture Studies, Lyon. 15-18 Juin 2005 http://www.gespraechsforschung-ozs.de/heft2006/tb-rohlfing.pdf

[31] LANDIS J. R., et KOCH G. G. The measurement of observer agreement for categorical data. Biometrics 33: 159–174, 1977.

[32] Bickmore, T.W., Schulman D., Shaw, G., DTask & LiteBody: Open Source, Standards-based Tools for Building Web-deployed Embodied Conversational Agents, Intelligent Virtual Agents, PP. 425-431 2009.