Interactive Narration Requires Interaction and Emotion

A. Pauchet[1], F. Rioult[2], E. Chanoni[3], Z. Ales[1] and O. Şerban[1]
[1]*INSA Rouen, LITIS, Rouen, France*
[2]*Université de Caen, Greyc, Caen, France*
[3]*Université de Rouen, Psy-NCA, Rouen, France*

Keywords: Dialogue Modelling, Knowledge Extraction, Narrative Embodied Conversational Agent.

Abstract: This paper shows how interaction is essential for storytelling with a child. A corpus of narrative dialogues between parents and their children was coded with a *mentalist grid*. The results of two modelling methods were analysed by an expert in parent-child dialogue analysis. The extraction of dialogue patterns reveals regularities explaining the character's emotion. Results showed that the most efficient models contain at least one request for attention and/or emotion.

1 INTRODUCTION

Designing a dialogue model is a difficult and often multidisciplinary task. It involves many algorithms: multi-modal inputs, natural language understanding and generation, dialogue management, emotion, ... In particular, multi-modal and affective dialogue management remains inefficient in Embodied Conversational Agents (ECA) (Cassell et al., 2000), even though this aspect is essential for interaction (Swartout et al., 2006). With the emergence of participative digital storytelling systems, child - humanoid agent interaction situations are increasing. The dialogue model embedded into an ECA, when dedicated to interactive storytelling, should be designed according to the child's socio-cognitive and language skills.

We aim at designing a method for modelling the dialogue between parents and children and tools to ease the extraction of regularities from interactive dialogues. We therefore propose hints to guide an interactive narrative session as well as dialogue patterns extracted from the corpora as a model of dialogue for narrative ECAs.

2 METHOD AND MATERIAL

The method proposed to model the dialogue is described in (Ales et al., 2012). It consists in: 1) *collecting and digitizing* a corpus of dialogues; 2) the *transcription and coding* step produces raw data with various levels of details (speaking slots, utterances, onomatopoeia, pauses, ...) depending on the phenomena and characteristics which the model must exhibit; 3) *knowledge and regularity extractions* are applied, through the utterances coded during the previous step. This knowledge consists in the regularities of the dialogues and constitutes the model; 4) finally, the model is *exploited* for interactive storytelling.

Corpus of Narrative Dialogues

Modelling storytelling dialogues requires knowledge extraction from a corpus of non mediated parent-child storytelling dialogues. In this study, 30 dialogues between children and parents (ages: 3, 4 and 5) were recorded during emotional story telling situations. These records have been transcribed and coded with a *mentalist grid* (Chanoni, 2009) to capture information about the mental states (beliefs, emotions, ...) contained by the various utterances.

Matrix Representation of Dialogues

As outlined by Bunt, dialogue management involves multilevel aspects (Bunt, 2011). To design a dialogue model that supports multidimensionality, a matrix representation is chosen for annotations. Each utterance is characterised by an annotation vector, whose components match the different coding dimensions. A dialogue is represented by a matrix: one row by utterance, one column by coding dimension.

To illustrate this two-dimensional representation, Table 1 presents an example of encoded parent-child dialogue, extracted from the collected corpus. Each utterance is characterised by a line number, a speaker (P: parent, C: child), a transcription and its encoding

Pauchet A., Rioult F., Chanoni E., Ales Z. and Şerban O..
Interactive Narration Requires Interaction and Emotion.
DOI: 10.5220/0004259605270530
In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence* (ICAART-2013), pages 527-530
ISBN: 978-989-8565-39-6

Table 1: 2D annotations of a narrative dialogue between a parent and his/her child.

Line	Speaker	Utterance	Annotations				
31	P	Who could have stolen the crown?	Q	-	F	-	-
32	C	The crown, it's in!	A	-	F	-	-
33	P	Do you believe it?	Q	H	K	-	-
34	C	Yes	A	-	F	-	-
35	P	But Babar doesn't know that it's in	A	P	N	O	J
36	P	So he says that the crown is a bomb	A	P	N	C	J

annotations.

The annotation grid corresponds to the following coding scheme: *Column 1:* an (A)ffirmation, a (Q)uery, a request for paying attention to the story (D), or a demand for general attention (G). *Column 2:* reference of the utterance. It can refer to the character (P), the interlocutor (H) or the speaker (R). *Column 3:* an (E)motion, a (V)olition, an observable or a non-observable cognition (B or N), an epistemic statement (K), an assumption (Y) or a (S)urprise. The surprise is distinguished from other emotions because of its link with the incidental belief. *Columns 4 and 5:* explanations with cause / (C)onsequence, (O)pposition or empathy (M), which can be applied either to explain the story (J), or to precise a situation with a personal context (F).

3 KNOWLEDGE EXTRACTION

Dialogue Pattern Extraction

With our matrix representation, a *dialogue pattern* is defined as a set of annotations which occurs in several dialogues. The method designed to extract significant dialogue patterns consists in a regularity extraction step based on matrix alignment using dynamic programming and a clustering step using machine learning heuristics to group and select the recurrent dialogue patterns. The clustering process is applied on a similarity graph computed during the matrix alignment.

The method for extracting two-dimensional patterns is a generalisation of the local string edition distance. The edit distance between two string S_1 and S_2 corresponds to the minimal cost of the three elementary edit operations (insertion, deletion and substitution of characters) for converting S_1 to S_2. Two-dimensional pattern extraction corresponds to matrix alignment. A local alignment of two matrices M_1 and M_2, of size $m_1 \times n_1$ and $m_2 \times n_2$ respectively, consists in finding the portion of M_1 and M_2 which are the most similar. To this end, a four-dimensional matrix T of size $(m_1+1)\times(n_1+1)\times(m_2+1)\times(n_2+1)$ is computed, such that $T[i][j][k][l]$ is equal to the local edition distance between $S_1[0..i-1][0..j-1]$ and $S_2[0..k-1][0..l-1]$ for all $i \in [\![1, m_1-1]\!]$, $j \in [\![1, n_1-1]\!]$, $k \in [\![1, m_2-1]\!]$ and $l \in [\![1, n_2-1]\!]$. In our heuristic, the calculation of T is obtained by the minimisation of a recurrence formula. Once T is computed, the best local alignment is found from the position of the maximal value in T, through a trace-back algorithm to infer the characters which are part of the alignment. Figure 1, commented in Section 4, presents an example of alignment extracted from the corpus. Details about the two dimensional pattern extraction algorithm can be found in (Lecroq et al., 2012).

The matrix alignment algorithm extracts the patterns in pairs. To determine the importance of each pattern, we group them using various standard clustering heuristics The idea is that large clusters of patterns represent behaviours which are commonly adopted by humans, whereas small clusters tend to contain marginal patterns. A matrix of similarities between patterns is computed through a global edition distance applied on all pairs of selected patterns. This similarity matrix is used as input for the clustering heuristics.

The method has been tested on the corpus of narrative dialogues. During the extraction phase, 1740 dialogue patterns have been collected.

Predicting the Interaction of the Child

As our goal is to build a dialogue model dedicated to narrative ECAs that stimulates child interaction, we have to model the arising of the child's interaction, focusing on *event prediction*. In other words, we look for sequences of dialogue events leading to child's interaction. We split the data over each turn of utterance, in other words over each sequence of parents assertion or question and child's interaction. The problem consists therefore in predicting the end of each turn.

The matrices that encode dialogues are considered as sequences of features, each sequence ending with the child's interaction. For instance, the sequences corresponding to Table 1, are: $<(QF)>, <(QHK)>, <(APNOJ)(APNOJ)>$

The algorithm, without candidate enumeration, mines the episodes with recursive projections, in a greedy manner. The combinatorial explosion is limited with two anti-monotone constraints: the support of the currently computed episode (the number of se-

quences it appears in) and the average distances (in utterances) to the end of the sequences supporting it.

As the mining process is directed from the end of the sequences to their beginning, not all the computed episodes are relevant for the prediction of the end. For instance, if each sequence begins and ends with a (Q)uestion, the above algorithm will give $< (Q) >$ as an end predictor, while it is also a good predictor for the beginning. To avoid these bad cases, the average length of each computed episode has to be taken into account. If it is too small, the episode is not kept. This process ensures that the mined regularities are relevant for the prediction of the child's interaction.

4 ANALYSIS OF THE MODELS

Patterns of Dialogue

Figure 1 presents an example of one pattern alignment appearing in two dialogues. This pattern shows that parents firstly talked about the cause or the consequence of the character's behaviour (P, C, J), without reference to any mental state. After assertions or descriptive questions, parents insisted on the justification of the character's behaviour (line 6), directly with relation to the emotional state of the character (line 7). Finally, the parent checked if the child understood, with asking questions or by requesting his/her attention (line 8).

	Dialogue B3 (4 years old)					Dialogue C8 (5 years old)				
	A	x	x	x	x	A	x	x	x	x
1	A	P	E	C	J	A	P	x	C	J
2	Q	x	x	x	x	A	x	x	x	x
3	A	x	x	x	x	A	x	x	x	x
4	A	x	x	x	x	A	x	x	x	x
5	A	x	x	x	x	A	x	x	x	x
6	A	P	x	C	J	A	P	x	C	J
7	A	P	E	x	x	Q	P	E	x	X
8	Q	x	x	x	x	D	x	x	x	X
9	A	x	x	x	x	A	P	E	x	x

Figure 1: Example of dialogue pattern alignment.

This pattern perfectly demonstrates that an emotion cannot only be named to be explained. The pattern described the link between the character's behaviour and the mental state, the later explaining the former.

Prediction of Interaction

Table 2 sums up the models for the child's interactions. For each age, the models are characterised by the *efficiency* (average number of sentences between the model and the child's interaction. The more a sequence is efficient, the fewer sentences before child's interaction) and the *support* (the percentage of times the model appears).

We highlight the important following facts : 1) Regardless of the age, sequences with justifications were frequently associated with various indexes (emotion, request for attention or question). The child's interaction came after 3.1 to 4.3 sentences after the model. 2) the quickness of the interaction decreased with the age, from 3.2 to 1.9 sentences. 3) the number of sequences with emotion is merely equivalent for all ages. Nevertheless, the older the child is, the more different the sequences of emotion are. Complex sequences (emotion and justification: J-E or E-J) appear only with the oldest children. 4) except with request for attention, the most efficient models (in red in Table 2) contained emotion (E-Q or E-E).

5 DIALOGUE MODELLING FOR NARRATIVE ECAS

ECAs are autonomous and anthropomorphic animated characters with multi-modal communication skills (Cassell et al., 2000). Greta (Pelachaud, 2009) and the European project SEMAINE (Schroder, 2010) provide good examples of current capabilities of ECAs. With regard to dialogue systems and models, which could be integrated into ECAs, several approaches exist. The *finite-state* approach (ex: (McTear, 2004)) which represents the structure of the dialogue as a finite-state automaton where each utterance leads to a new state. The *frame-based* approach represents the dialogue as a process of filling in a form containing slots (ex: (Aust et al., 1995)). The *plan-based* approach (Allen and Perrault, 1980) combines plan recognition with the Speech Act theory (Searle, 1969). The *logic-based* approach represents the dialogue and its context in some logical formalism (ex: (Hulstijn, 2000)). Finally, the *machine learning* approach proposes techniques such as reinforcement learning (Frampton and Lemon, 2009) to model the dialogue with Markov Decision Processes.

Most of the existing ECAs only integrates basic dialogue management processes, such as a keyword spotter within a finite-state or a frame-based approach (ex: the SEMAINE project (Schroder, 2010)). They uses regular structures that can only represent linear interaction patterns, whereas dialogue management involves multi-dimensional levels (Bunt, 2011). The model we propose combines planning management for the task resolution (prediction of the child interaction) and a more reactive management when dealing with dialogical conventions (dialogue patterns), all along multidimensionality management through the matrix coding of dialogues.

3 years old			4 years old			5 years old		
efficiency	model	support	efficiency	model	support	efficiency	model	support
3.2	E-Q	10,4%	2.1	D-Q	14,9%	1.9	Q	35,4%
3.4	D-Q	16,8%	2.2	E-Q	7,5%	2.2	E-E	9,1%
3.5	J-Q	9,6%	2.2	Q-Q	12,7%	2.6	J-D	8,1%
3.5	D-Q-Q	9,6%	2.6	D-E	10,4%	2.7	E-D	6,1%
3.5	E-J	8,8%	2.8	D-D	11,2%	3.1	J-E	8,1%
4.3	D-E	12,8%	3.5	J	14,9%	3.4	V	13,1%
4.3	D-J	8,0%	3.8	B	7,5%	3.7	D-E	7,1%
5.4	B	10,4%	4.0	E-E	7,5%	3.8	J-J	6,1%
5.6	V	13,6%	4.1	E-D	6,7%	4.1	E-J	7,1%
			4.3	V	6,7%			

Figure 2: Efficiency and quantity of all sequences for each age.

6 CONCLUSIONS

We have presented a methodology and tools designed to improve dialogue models that could be integrated in narrative ECAs. The proposed methodology consists in extracting dialogue patterns, clustering to encode dialogical conventions and an event prediction approach to plan the interactions of the listener. A matrix representation of interactions is used to encode the multidimensional aspects of dialogues. The algorithms were applied to a corpus of child-parent interactions during narration. Finally, we have shown why interactive narration requires prominent interactions and emotion.

REFERENCES

Ales, Z., Duplessis, G. D., ¸Serban, O., and Pauchet, A. (2012). A methodology to design human-like embodied conversational agents based on dialogue analysis. In *Proc. of HAIDM@AAMAS*, pages 34–49, Valencia, Spain.

Allen, J. and Perrault, C. (1980). Analyzing intention in utterances. *AI magazine*, 15(3):143–178.

Aust, H., Oerder, M., Seide, F., and Steinbiss, V. (1995). The philips automatic train timetable information system. *Speech Communication*, 17(3-4):249–262.

Bunt, H. (2011). Multifunctionality in dialogue. *Computer Speech and Language*, 25(2):222–245.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., and Yan, H. (2000). Embodied conversational agents. chapter Human conversation as a system framework: designing embodied conversational agents, pages 29–63. MIT Press.

Chanoni, E. (2009). Comment les mères racontent une histoire de fausses croyances à leur enfant de 3 à 5 ans ? Number 2, pages 181–189.

Frampton, M. and Lemon, O. (2009). Recent research advances in reinforcement learning in spoken dialogue systems. *Knowledge Engineering Review*, 24(04):375–408.

Hulstijn, J. (2000). Dialogue games are recipes for joint action. In *Proc. of Gotalog'00*.

Lecroq, T., Pauchet, A., and Solano, E. C. E. G. A. (2012). Pattern discovery in annotated dialogues using dynamic programming. In *IJIIDS*, volume 6, pages 603–618.

McTear, M. (2004). *Spoken dialogue technology: toward the conversational user interface*. Springer-Verlag New York Inc.

Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Trans. of the Royal Society B: Biological Sciences*, 364(1535).

Schroder, M. (2010). The SEMAINE API: towards a standards-based framework for building emotion-oriented systems. *Advances in HCI*, 2010:2–2.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University.

Swartout, W. R., Gratch, J., Jr., R. W. H., Hovy, E. H., Marsella, S., Rickel, J., and Traum, D. R. (2006). Toward virtual humans. *AI Magazine*, 27(2):96–108.