

# Modélisation de dialogues narratifs pour la conception d'un ACA narrateur

Alexandre Pauchet \* François Rioult † Émilie Chanoni ‡ Zacharie Alès \* Ovidiu Șerban \*

\* INSA Rouen - LITIS, {prenom.nom}@insa-rouen.fr

† Université de Caen - Greyc, francois.rioult@unicaen.fr

‡ Université de Rouen - Psy-NCA, emilie.chanoni@univ-rouen.fr

## Résumé :

Dans l'optique de la conception d'un Agent Conversationnel Animé (ACA) narratif et affectif, cet article démontre l'importance de l'interaction dans le processus de narration à des enfants. Un corpus de 30 dialogues de narration entre des parents et leur enfant de 3 à 5 ans, a été annoté à l'aide d'une grille dite "mentaliste". Deux méthodes d'extraction de connaissances ont été appliquées aux annotations des dialogues afin de les modéliser. Celles-ci révèlent des régularités dans les explications données par les parents dans la description des émotions des personnages. Ces travaux fournissent une ligne directrice pour la conception du modèle dialogique d'un ACA narrateur affectif.

**Mots-clés :** Modélisation du dialogue ; Extraction de connaissances ; Agent Conversationnel Animé

## 1 Introduction

La conception d'un modèle de dialogue est une tâche difficile et souvent pluridisciplinaire. Qu'il soit dédié à la narration interactive ou non, il implique de nombreux mécanismes : traitement de signaux multimodaux (parole, gestes, regards, etc.), reconnaissance et génération de langage naturel, gestion du dialogue, modélisation des émotions, prosodie et comportement non verbal. En particulier, la gestion de la multimodalité et des émotions dans le dialogue reste à ce jour insuffisante au sein des Agents Conversationnels Animés (ACA) [Cassell et al., 2000], bien que ces aspects soient essentiels pour des interactions efficaces [Swartout et al., 2006].

Avec l'émergence des environnements numériques, et plus particulièrement de systèmes de narration participative, les situations d'interactions enfant-agents humanoïdes sont de plus en plus fréquentes. Le comportement d'un ACA dédié à la narration interactive, devrait correspondre aux standards d'interaction adulte-enfant afin de faciliter la compréhension de ce dernier. En particulier, le modèle dialogique de l'agent devrait être adapté aux compétences socio-cognitives et langagières de l'enfant.

Le but de cet article est de proposer une méthode de modélisation du dialogue appliquée à

des dialogues de narration entre parents et enfants. Nous souhaitons fournir des outils permettant de faciliter l'extraction de régularités à partir de dialogues interactifs. Pour ce faire, nous proposons, d'une part d'extraire des motifs dialogiques à partir d'un corpus, et d'autre part une méthode prédictive permettant de guider une session de narration interactive, comme modèle de dialogue pour ACA narrateurs.

Un état de l'art sur le dialogue homme-machine est présenté section 2. La méthode ainsi que le corpus de dialogues sont décrits section 3. Les détails techniques des deux procédures sont décrits section 4 et le modèle extrait est analysé et expliqué section 5. L'article se termine par une courte conclusion ainsi qu'un exposé de nos futurs travaux.

## 2 Modèles du dialogue pour ACA narrateur : un (court) état de l'art

Les ACA sont des interfaces autonomes et anthropomorphiques, incarnés par des personnages animés aux compétences multi-modales : langage naturel, expressions du visage, regards, attitudes et gestes [Cassell et al., 2000]. Les ACA peuvent être catégorisés selon leur expressivité : systèmes de type présentateurs télé, interactions face à face (un ACA avec un humain) et conversations multipartites (plusieurs ACA et utilisateurs) [André and Pelachaud, 2010]. Les projets de recherche récents se focalisent sur l'interactivité des ACA en perfectionnant les expressions faciales et le comportement non verbal afin d'améliorer la qualité générale de l'agent [Cassell et al., 2000, Pelachaud, 2009]. Greta [Pelachaud, 2009], MARC [Courgeon et al., 2009] et le projet européen SEMAINE [Schröder, 2010] sont de bons exemples des capacités actuelles des ACA.

Dans le domaine de la narration, les agents virtuels intelligents, qu'ils soient incarnés ou non, peuvent être utilisés en tant que personnages

expressifs (ex : [Seif El-Nasr and Wei, 2008]) ou en tant que narrateurs. Le projet GV-LEX (Gesture and voice for an expressive reading), par exemple, a pour but de fournir aux robots Nao [Gouaillier et al., 2009] et à l'ACA Greta [Pelachaud, 2009] la capacité de lire du texte sans ennuyer l'auditeur [Gelin et al., 2010]. Ce projet propose d'utiliser une intonation expressive ainsi que des gestes tout en parlant afin de produire des narrations crédibles.

En ce qui concerne les systèmes et modèles du dialogue pouvant être intégrés dans les ACA, plusieurs approches existent.

**L'approche à états finis** (voir par exemple [McTear, 2004]) qui représente la structure du dialogue par un automate à états finis dans lequel chaque énoncé conduit à un nouvel état. En pratique, cette approche est limitée aux systèmes de dialogue directifs.

**L'approche par formulaire** représente le dialogue comme un processus de remplissage de formulaire contenant des entrées prédéfinies (voir par exemple [Aust et al., 1995]). Les contributions possibles sont fixées à l'avance.

**L'approche par planification** (exemple : [Allen and Perrault, 1980]) combine la reconnaissance de plans et la théorie des Actes de Langage [Searle, 1969]. Cette approche est complexe du point de vue calculatoire et requiert des composants très avancés de TAL afin d'inférer les intentions du locuteur.

**Le framework ISU** (*Information State Update*) [Larsson and Traum, 2000] utilise une représentation formelle du terrain commun, l'*état d'information*, ainsi qu'une structure gérant le raisonnement de l'agent.

**L'approche logique** représente le dialogue et son contexte par un formalisme logique et utilise des mécanismes tels que l'inférence et les jeux de dialogue (voir par exemple [Hulstijn, 2000]). La plupart des travaux concernant les approches logiques ne sont actuellement qu'au stade de la théorie.

**Les approches par apprentissage** proposent des techniques telles que l'apprentissage par renforcement [Frampton and Lemon, 2009] afin de modéliser le dialogue via des processus de Markov. Ces approches requièrent un travail d'annotation considérable.

En raison de la complexité des systèmes de dialogue complet, la plupart des ACA existants n'intègrent que des processus basiques de gestion du dialogue, telles que la détection de mots-clés au sein d'une approche de type états finis ou frame-based (ex : le pro-

jet SEMAINE [Schröder, 2010]). La gestion du dialogue reste donc inefficace dans les ACA actuels [Swartout et al., 2006]. Les approches précédentes utilisent comme représentation des structures de données régulières (ex : l'automate [McTear, 2004]), extraites manuellement ou apprises automatiquement à partir d'un corpus de dialogues, de traces ou de fichiers log. Ces structures de données ne permettent de représenter que des motifs d'interactions linéaires. Cependant la gestion du dialogue implique plusieurs dimensions et non une seule [Bunt, 2011]. Le modèle d'interaction d'un ACA nécessite la gestion de tous les aspects associés aux interactions humaines (gestion de tâches individuelles et collectives, feedbacks, aspects affectifs, obligations sociales, etc.), exprimés suivant différentes modalités (sémantique, prosodie, gestes, expressions, etc.).

En comparaison, le modèle de dialogue proposé dans la suite de cet article combine planification au service de la résolution de la tâche - prédiction et planification des interventions de l'enfant - et une gestion plus réactive par jeux de dialogue pour les conventions/motifs dialogiques. La représentation matricielle encodant le dialogue permet en plus de tenir compte du caractère multidimensionnel du dialogue.

### 3 Modélisation de dialogues

La méthode de modélisation du dialogue proposée est présentée Figure 1 :

1. *collecte et numérisation* d'un corpus de dialogues au format audio ou vidéo. Le corpus que nous considérons dans cet article est composé d'histoires enfantines racontées par des parents à leur enfant ;
2. l'étape *transcription et codage* consiste à produire des données brutes à divers niveaux de détails (tours de parole, énoncés, onomatopées, pauses, etc.) selon les caractéristiques que l'on souhaite exhiber ;
3. une phase d'*extraction de connaissances*, suivant un schéma de codage spécifique, est ensuite appliquée aux énoncés encodés afin d'obtenir une description précise des comportements dialogiques. Les dialogues sont alors considérés comme annotés.
4. une phase d'*extraction de régularités* (modélisation) est appliquée aux annotations.
5. le modèle peut alors être *exploité*<sup>1</sup>.

1. Cette étape n'est pas présentée dans cet article.

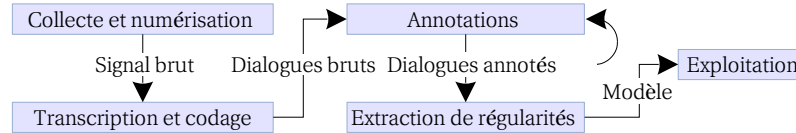


FIGURE 1 – Analyse du dialogue

Ligne	Locuteur	Énoncé	Annotations				
25	P	T'inquiète pas	A	P	E	-	-
26	P	Donc là ils se cachent	A	P	B	-	-
27	P	Ils cherchent	A	-	F	-	-
28	P	qui pourrait avoir pris la couronne.	Q	-	F	-	-
29	E	Elle est dedans, elle est dedans la couronne.	A	-	F	-	-
30	P	Donc là ils suspectent plein de monde, Cornélius, Céleste, la vieille dame...	A	P	Y	C	J
31	P	Qui a bien pu prendre la couronne ?	Q	-	F	-	-
32	E	La couronne elle est dedans.	A	-	F	-	-
33	P	Tu crois ? !	Q	H	K	-	-
34	E	Oui.	A	-	F	-	-
35	P	Mais Babar il ne sait pas qu'elle est dedans.	A	P	N	O	J
36	P	Donc il se dit que c'est une bombe, la couronne	A	P	N	C	J
37	P	ou je ne sais quoi.	A	R	N	-	-

TABLE 1 – Représentation matricielle des annotations d'un dialogue narratif parent-enfant

Dans les sous-sections suivantes, nous présentons l'expérimentation réalisée pour collecter le corpus de dialogues de narration d'histoires enfantines (étape 1) ainsi que le schéma de codage utilisé pour obtenir une représentation matricielle des dialogues (étape 2).

### 3.1 Corpus de dialogues narratifs

La narration d'histoires d'un parent à son enfant est une situation classique participant au développement de l'enfant. Les contextes sociaux et langagiers apportés par l'adulte, sont nécessaires à l'enfant dans son processus d'apprentissage des compétences socio-communicatives, cognitives et morales. Les enfants développent une *théorie de l'esprit* [Astington and Baird, 2005] durant leurs premières années et deviennent ainsi capables d'assimiler le fait qu'une personne est déterminée par ses propres intentions, émotions et états mentaux. Ce développement n'est possible qu'au travers des situations sociales de dialogue. Le discours des adultes concernant les états mentaux, en particulier, se révèle être un médiateur d'apprentissage du concept de cognition sociale - grâce à une participation active au dialogue et à des interactions dynamiques.

Dans cette étude, nous utilisons un corpus de 90 dialogues entre parents et enfants âgés de 3, 4 et 5 ans, filmés en situation de récit d'histoires enfantines (10 enfants par tranche d'âge x 3 histoires différentes). Ces enregistrements

sont retranscrits et annotés suivant une *grille mentaliste* [Chanoni, 2009] afin de faire ressortir les informations relatives aux états mentaux (croyances, volition, émotions, etc.) contenues dans les énoncés. La longueur moyenne des dialogues est de 89,3 énoncés.

### 3.2 Représentation des dialogues

Comme le souligne Bunt, la gestion du dialogue est multi-niveaux [Bunt, 2011]. Afin de concevoir un modèle du dialogue multi-dimensionnel, les annotations sont représentées matriciellement. Chaque énoncé est caractérisé par un vecteur d'annotations dont les composantes correspondent aux différentes dimensions de codage : une ligne par énoncé et une colonne par espace/dimension de codage.

Le Tableau 1 présente un exemple de dialogue provenant du corpus collecté. Chaque énoncé est caractérisé par un numéro de ligne, un locuteur (P : parent, E : enfant), une transcription et des annotations encodées suivant 5 dimensions :

- la première colonne caractérise la nature de l'énoncé : une (A)ffirmation, une (Q)uestion, une demande d'attention - générale (G) ou concernant l'histoire (D) ;
- la seconde colonne définit la référence de l'énoncé. Il peut se référer à un personnage (P), à l'auditeur (H) ou au narrateur (R) ;
- la troisième colonne est dédiée aux états mentaux. Les interlocuteurs peuvent exprimer une

(E)motion, une (V)olition, une cognition observable (B) ou non (N), une déclaration épistémique (K), une hypothèse (Y) ou une (S)urprise. La surprise se distingue des autres émotions de par son lien avec les croyances ;  
 – les deux dernières colonnes représentent les explications par (C)ause/conséquence, (O)pposition ou empathie (M), qui peuvent être utilisées, soit pour expliquer l’histoire (J), soit pour préciser une situation par l’évocation d’un contexte personnel (F).

Par exemple, la ligne 35 est encodée ainsi : l’énoncé est une affirmation (A) portant sur un état mental se référant à un personnage - “*Babar*” - (P) ; l’état mental correspondant - “*sait*” - se réfère à une cognition non observable (N) ; “*Mais*” dénote une justification par opposition (O) ; enfin, l’énoncé se réfère à l’histoire (J).

La construction de cette représentation matricielle nécessite un processus d’annotation manuel, semi-automatique et/ou automatique - un pour chaque dimension/colonne. Les matrices obtenues peuvent également être vues comme des séquences de vecteurs d’annotations. Une fois les dialogues annotés, l’étape d’extraction de régularités est appliquée sur les matrices.

## 4 Extraction de régularités pour la modélisation du dialogue

Nous proposons deux approches d’extraction de régularités : un calcul de similarité par programmation dynamique permettant de collecter des motifs dialogiques et une méthode de prédiction d’événements se concentrant sur la caractérisation des interactions de l’enfant.

### 4.1 Extraction de motifs dialogiques

Avec notre représentation matricielle, un *motif dialogique* est défini comme un ensemble d’annotations dont la disposition apparaît - de manière exacte ou approchée - dans plusieurs dialogues. Un motif peut contenir des annotations non adjacentes en ligne ou en colonne (i.e. il peut avoir des trous), et deux instances d’un même motif peuvent contenir des insertions, des suppressions ou des substitutions de caractères. Deux motifs sont donc considérés comme similaires si leur distance d’édition est faible.

La figure 2 présente la méthode utilisée pour extraire un ensemble de motifs dialogiques perti-

nents. Elle est composée d’une extraction de régularités basée sur un alignement de matrices par programmation dynamique, permettant de collecter un ensemble de paires de motifs similaires, suivi d’une étape de clustering afin de regrouper les motifs dialogiques récurrents.

La méthode d’extraction de motifs en deux dimensions s’apparente à l’alignement de matrices. Il s’agit d’une généralisation de la distance d’édition locale entre deux vecteurs de caractères. La distance d’édition  $ed$  (ou distance de Levenshtein) entre deux vecteurs de caractères  $s_1$  et  $s_2$  correspond au coût minimal des trois opérations d’édition élémentaires (insertion et suppression de caractères, ainsi que la substitution d’un caractère par un autre) permettant de transformer  $s_1$  en  $s_2$ . Un alignement local de deux matrices de caractères  $s_1$  et  $s_2$ , de tailles respectives  $m_1 \times n_1$  et  $m_2 \times n_2$ , consiste à chercher les portions de  $s_1$  et  $s_2$  qui sont les plus similaires (parmi toutes les portions de  $s_1$  et  $s_2$ ). Pour ce faire, une table à 4 dimensions  $T$  de taille  $(m_1 + 1) \times (n_1 + 1) \times (m_2 + 1) \times (n_2 + 1)$  est calculée, de telle sorte que  $T[i][j][k][l]$  soit égal à la distance d’édition locale entre  $S_1[0..i - 1][0..j - 1]$  et  $S_2[0..k - 1][0..l - 1]$ ,  $\forall i \in \llbracket 1, m_1 - 1 \rrbracket$ ,  $j \in \llbracket 1, n_1 - 1 \rrbracket$ ,  $k \in \llbracket 1, m_2 - 1 \rrbracket$  et  $l \in \llbracket 1, n_2 - 1 \rrbracket$ . Dans notre méthode, le calcul de  $T$  est obtenu par minimisation d’une formule de récurrence. Une fois  $T$  calculée, le meilleur alignement local est obtenu en effectuant un algorithme de tracé arrière à partir de la position où  $T$  atteint sa valeur maximale. Ce tracé arrière permet d’inférer les caractères faisant partie de l’alignement. La figure 3, commentée en section 5, présente un exemple d’alignement issu de notre corpus. Pour de plus amples informations concernant l’extraction de motifs en deux dimensions, se référer à [Lecroq et al., 2012].

L’alignement de matrices permet d’extraire les motifs par paires. Nous les regroupons à l’aide d’algorithmes heuristiques de clustering (voir tableau 2). L’idée sous-jacente est que les clusters les plus conséquents représentent les comportements les plus communs, tandis que les petits clusters reflètent des comportements plus marginaux. Une matrice de similarité entre les différents motifs est calculée grâce à une distance d’édition globale appliquée aux paires de motifs détectés. Cette matrice est utilisée comme entrée des algorithmes de clustering.

Cette méthode a été testée sur le corpus de dialogues de narration. Durant la phase d’extraction, 1740 motifs dialogiques ont été collectés,

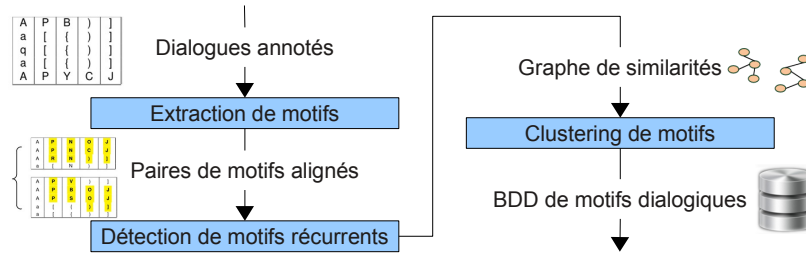


FIGURE 2 – Extraction de motifs dialogiques.

Méthode	Référence	Nombre de clusters trouvés					
		5	20	50	80	116	150
Single-Link	[Florek et al., 1951]	41	97	183	270	320	360
CHAMELEON	[Karypis et al., 1999]	458	605	<b>628</b>	-	-	-
ROCK	[Guha et al., 2000]	520	600	621	626	629	630
Spectral clustering non normalisé	[Von Luxburg, 2007]	277	<b>658</b>	563	155	194	226
Spectral clustering selon Shi and Malik	[Von Luxburg, 2007]	524	615	<b>628</b>	<b>631</b>	631	<b>632</b>
Spectral clustering selon Jordan and Weiss	[Von Luxburg, 2007]	<b>555</b>	616	<b>628</b>	630	631	<b>632</b>
Propagation d’affinité	[Frey and Dueck, 2007]	-	-	-	-	<b>632</b>	-

TABLE 2 – Indice de Dunn en fonction du nombre de clusters pour les heuristiques implémentées. Le caractère ‘-’ est utilisé lorsqu’une solution n’est pas produite pour un nombre de clusters donné. Les valeurs surlignées correspondent aux meilleur(s) résultat(s) pour chaque colonne.

de taille variant entre 10 et 124 énoncés pour une moyenne de 28,9.

Le nombre de solutions des méthodes de clustering étant trop élevé pour une comparaison au cas par cas, l’indice de Dunn [Dunn, 1973] a été utilisé afin d’évaluer les méthodes. Si  $s_{ij}$  mesure la similarité entre deux motifs  $i$  et  $j$ , et  $c(i)$  est le nombre de clusters contenant  $i$  pour une solution donnée, l’indice de Dunn est égal à

$$\frac{\min_{c(i) \neq c(j)} s_{ij}}{\max_{c(k)=c(l)} s_{kl}}$$

Ainsi, les solutions comportant un indice de Dunn élevé sont susceptibles d’être pertinentes, car composées de clusters compacts et clairement séparés. Le nombre de clusters étant lui-même inconnu, les méthodes ont été testées sur un grand nombre de valeurs. Le tableau 2 présente une partie représentative des résultats de l’indice de Dunn. Les meilleures méthodes semblent être la propagation d’affinité et les méthodes de type spectral clustering.

## 4.2 Prédiction des interventions de l’enfant

Cette section est consacrée à la définition d’un modèle du dialogue permettant de stimuler l’interaction. Dans cette optique, les interventions de l’enfant doivent être finement modélisées en se concentrant sur la *prédiction d’événement*.

Nous recherchons des séquences d’événements dialogiques entraînant une interaction particulière afin de s’en servir comme plan pour générer les interventions d’un ACA narrateur.

Les contributions majeures sur l’extraction de connaissances à partir de séquences sont principalement consacrées aux épisodes et à la classification de séquences. La prédiction d’événement sur des données discrètes n’y est que très peu abordée [Antunes and Oliveira, 2001]. Nous proposons de découper les données en *tours de parole*, caractérisés par un ensemble d’énoncés successifs provenant d’une seule personne (ici le parent ou l’enfant). Le problème revient à prévoir la fin du tour. Dans ce but, nous considérons des séquences de séries de vecteurs d’annotations (une série de vecteurs d’annotations par tour de parole) se terminant par une intervention de l’enfant. Les séquences du tableau 1 sont :  $\langle (APE)(APB)(AF)(QF) \rangle$ ,  $\langle (APYCJ)(QF) \rangle$ ,  $\langle (QHK) \rangle$ , ...

Pour extraire les régularités menant à la fin des séquences, les épisodes sont explorés par projections récursives grâce à un algorithme glouton. Dans l’exemple ci-dessus, l’algorithme débute avec l’épisode  $\langle (Q) \rangle$ , commun à toutes les fins de séquences. L’algorithme est ensuite appelé une nouvelle fois sur les séquences projetées  $\langle (APE)(APB)(AF) \rangle$ ,  $\langle (APYCJ) \rangle$ . (A) est ajouté à l’épisode qui devient  $\langle (Q)(A) \rangle$ , qui est lui même projeté une nouvelle fois : les

séquences résultantes sont  $\langle (APE)(APB) \rangle, \dots$ . L'explosion combinatoire est limitée par deux contraintes anti-monotones : la fréquence d'apparition et la longueur moyenne des séquences et la distance moyenne en nombre d'énoncés à la fin de la séquence.

Au cours du traitement - dans lequel la séquence est parcourue de la fin vers le début - les épisodes obtenus ne sont pas nécessairement tous appropriés à la prédiction de la fin du tour de parole. Supposons, par exemple, que chaque séquence commence et termine par une (Q)uestion, l'algorithme décrit précédemment donnera comme prédicteur de fin  $\langle (Q) \rangle$ , bien qu'il soit aussi un bon prédicteur de début. Pour éviter ce type de résultats défavorables, la distance moyenne de chaque épisode au début de la séquence doit être prise en compte. Si cette dernière est trop faible, l'épisode n'est pas conservé. Ce processus assure que les régularités extraites sont pertinentes.

Le processus d'extraction fournit un très grand nombre d'épisodes. Afin que l'expert puisse manuellement les évaluer, il est nécessaire d'en limiter le nombre. Dans cette optique, une approche par clustering de trajectoire [Lee et al., 2007] a été adoptée, en considérant que les épisodes sont des séquences de déplacements entre deux ensembles de vecteurs d'annotations. Les déplacements sont classés et un représentant est obtenu pour chaque classe. Ce regroupement permet de passer de plusieurs centaines d'épisodes à seulement quelques dizaines de représentants.

## 5 Analyse des modèles obtenus

L'évaluation des modèles calculés montre qu'un agent narrateur doit être interactif avec l'enfant (au travers de questions et de demandes d'attention) et ce d'autant plus avec les enfants en bas âge. Il apparaît essentiel de solliciter les enfants afin qu'ils interagissent. De plus, la compréhension des émotions et états mentaux des personnages peut être améliorée par une explication du comportement entraîné par l'état mental.

Dans la suite de cette section, l'évaluation des modèles réalisée par une psychologue spécialiste des interactions parents-enfant est détaillée. Ces modèles devraient permettre d'expliquer les comportements relatifs aux états mentaux observés et, à terme, d'être intégrés dans un ACA narrateur afin d'en guider le comportement.

## 5.1 Motifs dialogiques

L'extraction de motifs dialogiques a permis de collecter un ensemble de motifs et de les regrouper selon leur score de similarité calculé par programmation dynamique. La figure 3 présente un exemple d'alignement de motifs. Ce motif montre que les parents parlent, tout d'abord, des causes ou des conséquences du comportement du personnage (P, C, J) sans référence à l'état mental. Après quelques affirmations ou questions, les parents insistent sur la justification du comportement du personnage (ligne 6), puis le mettent en relation directe avec l'état mental du personnage (ligne 7). Enfin, le parent vérifie que l'enfant a compris en posant des questions ou en demandant son attention (ligne 8).

Ligne	Dialogue B3 (4 ans)				Dialogue C8 (5 ans)					
0	A	-	-	-	A	-	-	-		
1	A	P	E	C	J	A	P	-	C	J
2	Q	-	-	-	-	A	-	-	-	-
3	A	-	-	-	-	A	-	-	-	-
4	A	-	-	-	-	A	-	-	-	-
5	A	-	-	-	-	A	-	-	-	-
6	A	P	-	C	J	A	P	-	C	J
7	A	P	E	-	-	Q	P	E	-	-
8	Q	-	-	-	-	D	-	-	-	-
9	A	-	-	-	-	A	P	E	-	-

FIGURE 3 – Alignement de motif dialogique

Ce motif démontre parfaitement qu'il n'est pas suffisant de nommer un état mental pour l'expliquer. En effet, le développement narratif implique une démonstration pratique de la théorie de l'esprit. Le motif décrit le lien entre le comportement du personnage et l'état mental, le second expliquant le premier.

## 5.2 Prédiction d'interaction

Nous décrivons ici les conditions nécessaires à une amélioration significative de la narration interactive, en fonction de l'âge de l'enfant (3, 4 ou 5 ans). Le tableau 4 résume les modèles des interactions de l'enfant (voir section 3.2). Pour chaque âge, les modèles sont caractérisés par :

- leur *longueur moyenne*, qui correspond au nombre moyen d'énoncés entre le modèle et l'interaction de l'enfant. Plus une séquence est courte, moins il y a d'énoncés entre elle et l'intervention de l'enfant ;
- le *modèle*, qui décrit une séquence d'annotations. Par exemple la séquence E-Q symbolise une annotation E suivi, plus ou moins tard, d'une annotation Q. Les annotations peuvent ne pas être dans la même dimension ;

3 ans			4 ans			5 ans		
longueur	modèle	fréquence	longueur	modèle	fréquence	longueur	modèle	fréquence
3,2	E-Q	10,4%	2,1	D-Q	14,9%	1,9	Q	35,4%
3,4	D-Q	16,8%	<b>2,2</b>	<b>E-Q</b>	<b>7,5%</b>	<b>2,2</b>	<b>E-E</b>	<b>9,1%</b>
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 4 – Longueurs moyennes et fréquences de séquences en fonction de l'âge.

– la *fréquence*, qui est le pourcentage de fois où le modèle apparaît.

Les demandes d'attention (codées D, par exemple "regarde !" ou "tu as vu ?"), essentielles pour la narration interactive, sont présentes pour tous les âges. Plus l'enfant est âgé, plus sa réaction à la demande d'attention est rapide. Plus l'enfant est jeune, plus les demandes d'attention doivent être répétées ou ponctuées de questions. Ceci peut s'observer dans des séquences telles que D-D ou D-Q ou D-Q-Q.

110 séquences contenant une demande d'attention ont été recensées dans les dialogues des enfants de 3 ans, 58 pour les enfants de 4 ans. Les enfants de 3 ans interagissent en effet après un nombre d'énoncés moyen compris en entre 3,5 et 4,6. Les enfants de quatre ans réagissent plus rapidement (entre 2,2 et 2,8 énoncés) : l'efficacité du modèle s'améliore avec l'âge. Par contre, les parents se comportent différemment avec les enfants de 5 ans se comportent : les demandes d'attention sont moins fréquentes et soit associées à des états mentaux (D-E ou E-D) soit à des justifications (D-C ou C-D). Nous n'avons dénombré que 21 séquences comportant des demandes d'attention, rapidement suivies d'une interaction de l'enfant (entre 2,6 et 2,7 énoncés). Les séquences comprenant des justifications (codées J, par exemple "puis, Leo casse le château !") sont essentielles au processus d'interaction émotionnelle narrative.

En conclusion, nous souhaitons mettre l'accent sur certains points notables :

- quel que soit l'âge, les séquences contenant des justifications sont fréquemment associées à divers indices (émotion, demande d'attention ou question). Dans ce contexte, l'interaction de l'enfant ne survient qu'entre 3,1 et 4,3 énoncés après le modèle ;
- la longueur des interactions décroît avec l'âge, de 3,2 à 1,9 énoncés ;

- le nombre d'énoncés auxquels sont associées des émotions est quasiment équivalent pour tous les âges. Néanmoins, plus l'enfant est âgé, plus les séquences d'émotions sont variées. Les séquences complexes (émotions et justifications : J-E ou E-J) n'apparaissent qu'avec les enfants les plus âgés ;
- à l'exception des demandes d'attention, les modèles les plus efficaces (en rouge et gras dans le tableau 4) contiennent toujours des émotions (E-Q ou E-E).

## 6 Conclusion

Nous avons montré, dans cet article, les raisons pour lesquelles la narration nécessite de nombreuses interactions et émotions. Nous avons présenté une méthodologie et des outils permettant d'améliorer la modélisation du dialogue. Ces modèles sont dédiés à la narration interactive et peuvent être intégrés dans des ACA narrateurs. La méthodologie proposée consiste, d'une part, à extraire des motifs dialogiques et à les classifier afin d'encoder les conventions dialogiques et, d'autre part, à une prédiction d'événements afin d'encourager l'interaction avec l'auditeur. Une représentation matricielle de l'interaction est utilisée afin d'encoder les aspects multidimensionnels du dialogue. Nos algorithmes ont été appliqués à un corpus de dialogues de narration parents-enfants.

Les interactions enfant-agent peuvent différer des interactions parent-enfant, non seulement par les capacités dialogiques de l'agent, mais aussi en raison de la représentation numérique de l'agent. Afin d'évaluer l'impact de l'incarnation, une seconde expérience a été menée. Des dialogues enfant-adulte ainsi que des dialogues enfant-avatar - durant une expérience de type Wizard of Oz (WOz) - ont été enregistrés via un système de vidéoconférence. Ce corpus de dialogues est actuellement en cours d'an-

notation. Nous prévoyons d'appliquer les algorithmes présentés précédemment afin de comparer les modèles de dialogue obtenus par visioconférence (interaction enfant-adulte) et ceux du WOz (interaction enfant-avatar).

Enfin, nos travaux futurs seront dédiés à l'intégration effective des modèles de dialogue obtenus au sein d'un ACA et à son évaluation en situation de narration interactive. Pour ce faire, nous proposons d'utiliser, comme Hulstijn [Hulstijn, 2000], des jeux de dialogues afin de gérer les conventions représentées par les motifs dialogiques.

## Remerciements

Ce travail a bénéficié du soutien du projet CNRS PEPS INS2I-INSHS « ACAMODIA ».

## Références

- [Allen and Perrault, 1980] Allen, J. and Perrault, C. (1980). Analyzing intention in utterances. *Artificial Intelligence*, 15(3) :143–178.
- [André and Pelachaud, 2010] André, E. and Pelachaud, C. (2010). Interacting with embodied conversational agents. *Speech technology*, pages 123–149.
- [Antunes and Oliveira, 2001] Antunes, C. M. and Oliveira, A. L. (2001). Temporal data mining : An overview.
- [Astington and Baird, 2005] Astington, J. W. and Baird, J. (2005). *Why language matters for theory of mind*. Oxford University Press, New York.
- [Aust et al., 1995] 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, 2011] Bunt, H. (2011). Multifunctionality in dialogue. *Computer Speech and Language*, 25(2) :222–245.
- [Cassell et al., 2000] Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsón, 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, 2009] Chanoni, E. (2009). Comment les mères racontent une histoire de fausses croyances à leur enfant de 3 à 5 ans ? *Enfance*, (2) :181–189.
- [Courgeon et al., 2009] Courgeon, M., Clavel, C., and Martin, J.-C. (2009). Appraising emotional events during a real-time interactive game. In *AFFINE'09*, pages 7 :1–7 :5, New York, NY, USA. ACM.
- [Dunn, 1973] Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3) :32–57.
- [Florek et al., 1951] Florek, K., Lukaszewicz, J., Perkal, J., Steinhaus, H., and Zubrzycki, S. (1951). Sur la liaison et la division des points d'un ensemble fini. In *Colloquium Mathematicum*, volume 2, pages 282–285.
- [Frampton and Lemon, 2009] Frampton, M. and Lemon, O. (2009). Recent research advances in reinforcement learning in spoken dialogue systems. *Knowledge Engineering Review*, 24(04) :375–408.
- [Frey and Dueck, 2007] Frey, B. and Dueck, D. (2007). Clustering by passing messages between data points. *Science*, 315(5814) :972.
- [Gelin et al., 2010] Gelin, R., d'Alessandro, C., Le, Q. A., Deroo, O., Doukhan, D., Martin, J.-C., Pelachaud, C., Rilliard, A., and Rosset, S. (2010). Towards a storytelling humanoid robot. In *AAAI Fall Symposium Series*.
- [Gouaillier et al., 2009] Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., Monceaux, J., Lafourcade, P., Marnier, B., Serre, J., and Maisonnier, B. (2009). Mechatronic design of nao humanoid. *Proc. of the Int. Conf. on Robotics and Automation*, pages 769–774.
- [Guha et al., 2000] Guha, S., Rastogi, R., and Shim, K. (2000). Rock : A robust clustering algorithm for categorical attributes. *Information Systems*, 25(5) :345–366.
- [Hulstijn, 2000] Hulstijn, J. (2000). Dialogue games are recipes for joint action. In *Proc. of Gotalog'00*.
- [Karypis et al., 1999] Karypis, G., Han, E., and Kumar, V. (1999). Chameleon : Hierarchical clustering using dynamic modeling. *Computer*, 32(8) :68–75.
- [Larsson and Traum, 2000] Larsson, S. and Traum, D. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering*, 6(3&4) :323–340.
- [Lecroq et al., 2012] Lecroq, T., Pauchet, A., Chanoni, E., and Solano, G. A. (2012). Pattern discovery in annotated dialogues using dynamic programming. *Int. J. of Intelligent Information and Database Systems*, 6(6) :603–618.
- [Lee et al., 2007] Lee, J.-G., Han, J., and Whang, K.-Y. (2007). Trajectory clustering : a partition-and-group framework. In *Int. conf. on Management of data*, pages 593–604. ACM.
- [McTear, 2004] McTear, M. (2004). *Spoken dialogue technology : toward the conversational user interface*. Springer-Verlag New York Inc.
- [Pelachaud, 2009] Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Trans. of the Royal Society B : Biological Sciences*, 364(1535).
- [Schröder, 2010] Schröder, M. (2010). The SEMAINE API : towards a standards-based framework for building emotion-oriented systems. *Advances in HCI*, 2010 :2–2.
- [Searle, 1969] Searle, J. (1969). *Speech Acts : An Essay in the Philosophy of Language*. Cambridge University.
- [Seif El-Nasr and Wei, 2008] Seif El-Nasr, M. and Wei, H. (2008). Exploring non-verbal behavior models for believable characters. In *Interactive Storytelling*, volume 5334 of *Lecture Notes in Computer Science*, pages 71–82.
- [Swartout et al., 2006] 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.
- [Von Luxburg, 2007] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4) :395–416.