



HAL
open science

Processus de génération d'une animation en 3 dimensions à partir d'un texte.

Nicolas Kamennoff

► **To cite this version:**

Nicolas Kamennoff. Processus de génération d'une animation en 3 dimensions à partir d'un texte.. 19es Journées Francophones d'Ingénierie des Connaissances (IC 2008), Jun 2008, Nancy, France. pp.147-159. hal-00416695

HAL Id: hal-00416695

<https://hal.science/hal-00416695>

Submitted on 14 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Processus de génération d'une animation en 3 dimensions à partir d'un texte

Nicolas KAMENNOFF^{1,2}

¹LIPN – UMR 7030
Université Paris13 – CNRS
99, av. J-B Clément – F-93430 Villetaneuse

²L.E.R.I.A – Epitech
24, rue Pasteur
94270 Le Kremlin Bicêtre

Résumé : Cet article présente les hypothèses de travail du projet ASMV (Analyse Sémantique, Modélisation et Visualisation). Il s'agit ici de réaliser un système capable de générer une animation en 3 dimensions à partir d'un texte en langage naturel. Le domaine choisi est celui des accidents de la route. La problématique soulevée concerne les informations implicites du texte et la polysémie des termes. La solution explorée se base sur une représentation des connaissances orientée objet (Van Roy & Haridi, 2004), faisant appel à des raisonnements non monotones (Grégoire, 1990) (Kayser, 1997b) et à l'élaboration d'un système de validation et de correction automatique.

Mots-clés : Compréhension, visualisation, modélisation, connaissances, raisonnement.

1 Introduction

La génération d'images à partir de textes en langage naturel est une tâche peu étudiée. Parmi les quelques projets qui ont tenté l'expérience, on citera notamment NALIG (Ardoni & Di Manzo et al. 1984), WordsEye (Coyne & Sproat, 2001) ou Carsim (Johansson & Berglund et al. 2001).

La génération d'images et plus encore celle d'animations semble toutefois être une approche intéressante de la compréhension automatique du langage. La compréhension est une notion difficile à évaluer ; il n'existe pas à l'heure actuelle de consensus sur la définition des processus qui la composent. Plusieurs approches tentent de la décrire (Le Ny & Kintsch, 1982), (Snow, 2002), (Falardeau, 2003), mais cette notion reste floue. Nous considérons cette tâche comme étant inférentielle. Il est difficile de concevoir un système capable d'avoir une compréhension correcte d'un texte ou d'un discours sans qu'il ait lui-même des connaissances préalables sur le domaine du discours et sur le monde en général.

La compréhension est donc réalisée en fonction des connaissances acquises, cette nature rend l'évaluation de la tâche subjective. Plusieurs interprétations, plus ou moins similaires, d'un même texte sont possibles. Dans le domaine de l'enseignement,

l'évaluation de la compréhension se faisait par des exercices de « lecture silencieuse » introduits par Edward Thorndike en 1917 (Lafontaine, 2003). L'exercice consiste en la lecture d'un texte suivi d'un questionnaire censé valider la compréhension par l'élève. Cette méthode était considérée comme la plus performante avant que l'évaluation de la compréhension en tant que telle ne soit abandonnée dans les cursus.

Cet article introduit la problématique choisie par le projet ASMV-CAR (Constats d'Accidents de la Route), démarré en 2007, actuellement en cours de réalisation. Il vise à la création d'un système de génération d'animation en 3D à partir de textes de constats d'accidents de la route. Ce domaine d'étude offre plusieurs avantages. La représentation des scènes décrites, quoique variée, reste dans des limites assez raisonnables (peu d'actions élémentaires possibles pour une voiture), des travaux sur ce domaine sont déjà réalisés au LIPN (Nouioua, 2007), (MICRAC) et nous disposons d'un corpus de descriptions rédigées par les conducteurs. Nous traitons, dans cet article, de l'une des principales difficultés rencontrées, le problème de la non-exhaustivité des descriptions textuelles.

Après un aperçu des principaux travaux de génération d'image en 3 dimensions à partir d'un texte en langage naturel en section 2, la section 3 développe cette problématique. La section 4 présente ensuite le projet ASMV et les différentes hypothèses suivies pour la résolution de ces problèmes.

2 Travaux connexes

Peu de travaux portent actuellement sur la génération d'une animation à partir d'un texte en langage naturel.

Une approche de cette problématique est présentée par le projet Carsim (Dupuy & Egges, 2001), (Johansson & Berglund, 2005). Ce projet, développé à l'Université de Lund, tente également de créer un système de génération d'animation en 3 dimensions à partir d'un texte narratif. Les auteurs se basent sur un corpus de textes en suédois provenant de journaux et de l'organisme suédois de collecte des données sur les accidents de la route (STRADA). Leur approche est de créer une représentation en utilisant un formalisme tabulaire (défini en XML) permettant de définir l'ensemble des connaissances nécessaire à la génération d'une animation. L'animation est ensuite perçue comme un problème de satisfaction de contraintes afin de respecter les différents éléments décrits dans la scène. Cette dernière étape permet d'obtenir une description géométrique précise qui est ensuite utilisée pour la vidéo. Cette approche est intéressante et représente une première dans la génération de rendu animé à partir d'un texte. Cette technique est toutefois limitée par le formulaire de description au domaine des accidents de la route. De plus l'absence d'un moteur physique et d'une description fine des objets ne permet pas de garantir le réalisme de la scène. Après une validation par un ensemble d'utilisateurs, les auteurs indiquent un taux de 55% des textes générant une animation jugée satisfaisante.

WordsEye (Sproat & Coyne, 2001) est un autre travail notable. Ce système vise à générer une image à partir d'un texte descriptif. Ce texte est toutefois écrit spécifiquement pour la solution, le travail ici est dirigé vers la gestion des relations spatiales entre les objets et les descriptions physiques de ceux-ci (taille, couleur,

transparence...). Ce système est essentiellement basé sur l'analyse syntaxique, permettant de définir les liens entre les différents termes de la phrase.

3 Problématique

Le langage permet la transmission d'informations entre les individus. Les textes ou discours représentent donc un ensemble d'informations. Nous étudions la compréhension comme un processus permettant l'assimilation de ces informations suivant nos connaissances. La valeur sémantique de chaque mot d'une phrase apparaît comme résultant d'une adaptation des connaissances préalables du destinataire concernant cet objet. Le fait que les connaissances de chaque lecteur ou auditeur d'un texte soient basées sur l'interprétation d'un monde réel commun permet au narrateur, d'omettre volontairement certaines informations, soit qu'il les juge inutiles (une description complète, si tant est qu'elle soit possible, serait fastidieuse à écrire, et encombrée de détails qui n'intéressent aucun lecteur), soit qu'il n'en dispose pas lui-même. Parmi ces informations un certain nombre seront inférées par le destinataire du message.

Lors de la conception de systèmes de génération d'animation à partir de textes en langage naturel, les problèmes liés à la non-exhaustivité peuvent apparaître sous trois angles présentés dans cette section.

3.1 Absence d'informations

Un texte présente généralement les informations considérées par l'auteur comme étant les plus importantes pour la description de l'événement. Les acteurs, leurs actions et interactions sont donc décrits par le narrateur. Mais ces termes représentent bien souvent des objets, des actions ou des interactions complexes, répartis dans le temps et l'espace, et chacun d'eux est défini par un grand nombre de paramètres.

Dans le cas des textes de constat d'accidents de la route, un grand nombre de paramètres sont en effet omis par le rédacteur. Ainsi, le texte présenté en figure 1 ne comporte aucune indication concernant les vitesses, la largeur de la route (nombre et dimensions des voies), le marquage au sol, le nombre d'embranchements du carrefour, les voitures impliquées (modèle, couleur, dimensions, nombre de portes), etc...

« Me rendant à Beaumont sur Oise depuis Cergy, je me suis retrouvée à un carrefour juste après la sortie Beaumont sur Oise. J'étais à un stop avec 2 voitures devant moi, tournant à droite vers Mours. Alors que la première voiture passait ce stop je fis mon contrôle à gauche et je démarrais mais je percutais la deuxième voiture qui n'avait pas encore passé le stop. »

Fig. 1 – Texte de constat d'accident de la route.

3.2 Ambiguïté des termes ou polysémie

Cette difficulté apparaît au niveau de la représentation des connaissances et du calcul du sens des expressions. La polysémie regroupe plusieurs phénomènes distincts (Jacquet, Venant et al. 2005). Si toutes ses formes apparaissent dans la problématique générale du traitement automatique des langues (TAL), nous la définissons ici comme le phénomène, particulièrement important pour la génération d'image, qui mène à représenter une même expression d'action par plusieurs ensembles d'actions élémentaires distincts. La génération d'une animation demande un grand degré de précision sur la représentation d'une action. Il est donc nécessaire de choisir la visualisation appropriée.

Prenons l'exemple du dépassement. Cette action est généralement représentée par le schéma en figure 2.a. Mais si à l'état initial, la voiture 1 se trouve déjà sur une file à gauche de la voiture 2, l'action est représentée par la figure 2.b. Si les informations du texte suffisent généralement pour savoir dans quelle configuration on se trouve, elles sont rarement exprimées de manière explicite.

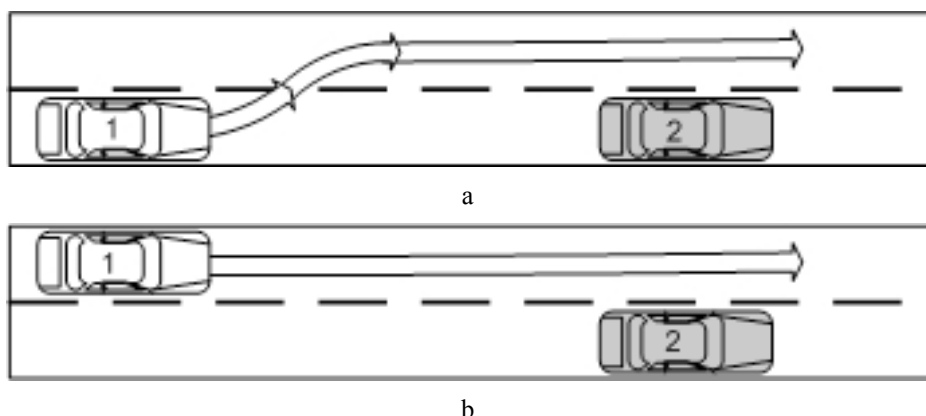


Fig. 2 – Différentes actions de dépassement.

3.3 Informations sur le fonctionnement du monde

D'autres connaissances sont nécessaires pour obtenir une compréhension cohérente d'un texte. En effet les informations contenues dans le texte ne permettent pas la restitution complète du message. Des connaissances plus générales sur le monde sont nécessaires. Pour reprendre l'exemple donné par Daniel Kayser dans (Kayser, 1997a), une compréhension « littérale » des informations du texte présenté en figure 1 fournit le schéma présenté en figure 3.a. Or la compréhension normale de la scène doit mener au schéma 3.b. Le processus permettant la correction de la scène en prenant en compte les connaissances dont on dispose sur le fonctionnement « normal » du monde doit donc être formalisé. Le formalisme ici utilisé est une représentation des connaissances

utilisant des expressions logiques et la définition des processus nécessaires à leur application pour la tâche donnée.

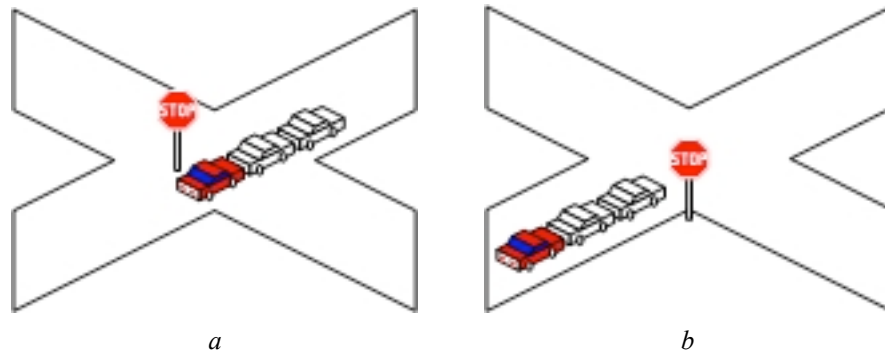


Fig. 3 – Deux compréhensions de l'état initial de la scène décrite.

4 Processus de la compréhension automatique

Cette section présente la solution explorée par notre projet ASMV. Nous exposons tout d'abord l'architecture globale du projet. Ensuite les processus permettant la complétion des informations fournies par les textes pour la génération d'une scène seront présentés en détails. Enfin le système de validation et de correction de l'animation sera esquissé.

4.1 Architecture du projet ASMV

Le projet ASMV vise à la génération d'une animation en 3 dimensions à partir d'un texte descriptif exprimé en français. Le processus de génération adopté par le projet peut être découpé en 4 parties importantes, ces parties et leurs liens sont représentées par la figure 4.

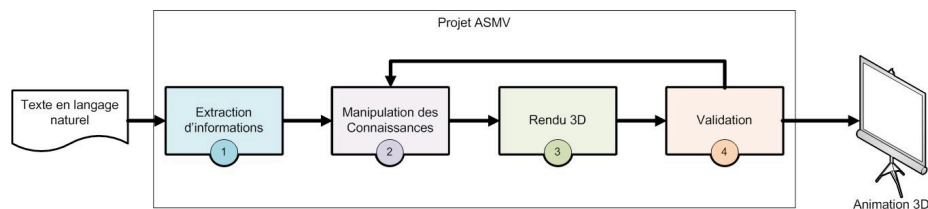


Fig. 4 – Architecture du projet ASMV.

4.1.1 Extraction d'informations (étape 1)

La première étape est d'extraire les informations contenues dans le texte. Cette partie fait principalement appel à des techniques syntaxiques. L'extraction d'information est une tâche déjà bien étudiée en TAL. La tâche d'extraction nécessaire au projet ASMV consiste ici à déterminer l'ensemble des acteurs, des actions et des événements survenus dans le discours sous une forme temporellement ordonnée.

Afin de se focaliser sur la problématique de la génération d'une animation à partir d'un texte, le projet ASMV utilise un système d'extraction d'information précédemment développé. Ici notre choix s'est porté sur les travaux effectués par Farid Nouioua (Nouioua, 2007) sur ce même corpus, l'extraction des informations de textes de constats d'accident de la route, travaux basés sur la logique des défauts (Reiter, 1980). Une partie du travail réalisé par l'auteur correspond aux besoins de notre projet. Le résultat obtenu est alors un ensemble d'informations sous une forme prédicative (figure 5).

texte :

Au feu tricolore, j'étais arrêtée derrière une voiture (Véhicule Nom_De_Véhicule). Le feu passant au vert, j'ai amorcé un démarrage mais le chauffeur du véhicule ci-dessus nommé n'a pas démarré. Je suis donc venue heurter l'arrière de la voiture.

Prédicats :

```
vrai(arreter, auteur, 1) .  
vrai(combine(suivre, veh_A), auteur, 1) .  
vrai(feux_tricolore, auteur, 1) .  
vrai(feux_vert, auteur, 2) .  
vrai(demarrer, auteur, 3) .  
-vrai(demarrer, veh_A, 3) .  
vrai(combine(heurter, veh_A), auteur, 4) .  
vrai(combine(position_choc, arriere), veh_A, 4) .
```

Fig. 5 – Exemple d'informations extraites.

Les prédicats représentent des états, des actions ou des événements. Les acteurs y apparaissent comme des constantes, et la restriction du domaine d'étude aux constats d'accident de la route permet presque toujours de lever les ambiguïtés purement linguistiques. De plus, d'après une étude comparative entre les prédicats extraits et les informations nécessaires extraites manuellement, seules peu d'informations « graphiques » sont omises par ce processus. Ce système est donc satisfaisant pour le début de l'étude. L'architecture modulaire du programme réalisé permettra toutefois à l'avenir de modifier cette étape sans altérer le reste du système.

4.1.2 Manipulation des connaissances (étape 2)

Il s'agit d'une des étapes les plus importantes du processus. Elle permet de faire le lien entre les informations contenues dans le texte et les connaissances spécifiques du système. Le système de représentation des connaissances sera présenté par la suite.

Cette étape intervient une première fois suite à l'extraction des informations. Elle peut ensuite être appelée à plusieurs reprises par le système de validation, permettant alors une correction de la scène générée.

Les informations extraites du texte sont confrontées aux connaissances du domaine afin d'obtenir un nouveau formalisme de la scène. Ce processus se base sur plusieurs tâches. Premièrement la définition des acteurs de la scène, il s'agit des éléments physiques représentés et leur état initial. Ensuite pour chaque étape, la gestion des déplacements spatiaux permet de définir le résultat attendu. La complétion des informations (développée par la section 3.2) génère ensuite les informations nécessaires qui ne sont pas fournies par l'auteur. Enfin un calcul permet d'obtenir les données temporelles (durée de chaque action et événement de la scène). Ces étapes sont résumées par le schéma en figure 6.

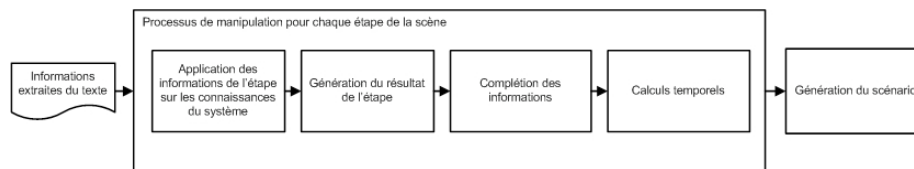


Fig. 6 – Processus de manipulation des connaissances.

Le formalisme obtenu représente alors toutes les connaissances connues et déduites de la scène décrite par le texte. C'est à partir de celui-ci que le scénario de l'animation est créé.

4.1.3 Rendu 3D (étape 3)

Cette partie permet de générer l'animation en 3 dimensions à proprement parler. Cette partie du programme prend en entrée un scénario généré par l'étape précédente. Ensuite le moteur 3D affiche les objets en utilisant les modèles 3D associés (cf. section 3.4.1) et les déplace en utilisant les fonctions associées à chaque prédicat du scénario.

Le choix d'un rendu en 3D présente plusieurs avantages. Premièrement au niveau de la visibilité, plusieurs points de vue peuvent être proposés à l'utilisateur pour l'observation de la scène. Cela peut être des points fixes, ou embarqués sur les acteurs de la scène, ou même un système permettant le déplacement libre de la caméra dans la scène. Ainsi l'utilisateur peut vérifier à sa guise les différents éléments de l'animation. Ensuite, lors du calcul de l'animation, il sera assez simple d'implémenter l'application des connaissances générales, notamment les lois physiques (Bourg, 2002).

Les technologies 3D ont connu un essor important ces dernières années. De nombreux moteurs 3D sont disponibles et proposent des solutions simplifiées pour obtenir des rendus 3D interactifs. Après une étude des différentes possibilités, nous avons choisi l'utilisation du moteur Ogre3D (Ogre3D). Les fonctionnalités

disponibles, la licence OpenSource et la maturité du projet justifient notre sélection parmi l'ensemble des systèmes disponibles.

4.1.4 Validation de la scène (étape 4)

L'étape précédente crée une première animation à partir du texte. Or la section 2.3 montre que l'interprétation littérale du texte peut conduire à des erreurs, si on ne prend pas en compte les connaissances générales sur le fonctionnement «normal» du monde. Cette quatrième étape a donc pour but d'utiliser les possibilités du système de rendu en 3D pour valider la cohérence de l'animation à l'aide de connaissances plus générales. Si ce système détecte une erreur, il appelle de nouveau l'étape de manipulation des connaissances en ajoutant de nouvelles informations décrivant l'erreur détectée.

Ce système est présenté plus en détail dans la section 4.4.

4.2 Complétion des informations

L'une des difficultés de la génération d'animation est le manque d'informations exprimées par le texte. Dans cette section, nous présentons la solution explorée par le projet ASMV. Il s'agit ici de l'utilisation de valeurs par défaut associées aux paramètres. Ces valeurs peuvent toutefois être modifiées par des éléments textuels.

4.2.1 Valeurs par défaut

Comme expliqué précédemment, le but d'un auteur lorsqu'il écrit un texte est de transmettre de l'information en ne fournissant que les éléments qu'il juge essentiels. Or certains éléments non transmis s'avèrent essentiels à la compréhension. L'auteur compte donc sur le fait que les destinataires vont a priori inférer ces informations. Ces informations doivent donc être obtenues à l'aide des connaissances du système. Notre approche propose d'utiliser des valeurs par défaut préalablement entrées dans les connaissances du système. Chaque paramètre apparaissant dans les connaissances du système doit donc posséder une valeur par défaut.

Plusieurs valeurs peuvent toutefois être définies pour un paramètre. Le choix d'une de ces valeurs doit être déterminé par des connaissances contextuelles. La vitesse par défaut d'une voiture sera par exemple de 50 km/h (correspondant à la vitesse limitée en zone urbaine). Toutefois si la scène est située sur une voie rapide (autoroute) alors cette valeur sera de 130 km/h, ou de 90 km/h sur une route ordinaire.

Ces valeurs par défaut sont des valeurs inférées et donc incertaines. La section 4.4 présente une hypothèse permettant la validation de telles informations.

4.2.2 Modificateurs

Dans certains cas, des indications sur ces paramètres non exprimés apparaissent. Ces indices se retrouvent au travers de qualificatifs (adverbes ou adjectifs). Ainsi dans le syntagme « *J'ai freiné brusquement* », le temps de réaction du narrateur n'est pas

donné, ni la puissance du freinage, il signale ici simplement par l'adverbe « *brusquement* » que son freinage est plus rapide et plus fort que ce qu'il juge lui-même comme étant « normal ».

Des connaissances sur ces modificateurs doivent donc être définies. Ces modificateurs ont eux-mêmes des comportements par défaut. Les valeurs de ces actions doivent aussi être fournies au préalable au système. Ainsi on appliquera sur un freinage le modificateur « fort » par une augmentation prédéfinie de la puissance de freinage (dans la limite des capacités du véhicule).

4.3 Gestion de la polysémie

Dans le projet ASMV, la polysémie est gérée au niveau de la représentation des connaissances. Cette section fait une brève présentation du système de représentation mis au point. Le système de génération de mouvement est ensuite détaillé.

4.3.1 Représentation des connaissances

Les processus de manipulation de la langue sont par nature révisables. En effet lorsque l'on lit ou écoute un texte, les informations arrivent et sont intégrées phrase après phrase ; celles qui sont contenues dans une phrase peuvent modifier les inférences précédemment faites. La logique mathématique, qui est monotone, n'est donc pas adaptée pour modéliser ce processus (Kayser, 1997b). Cette nature révisable nous amène ici à étudier le système de représentation des connaissances en utilisant un système de logique non-monotone et plus précisément la logique des défauts de Reiter (Reiter, 1980) (Grégoire, 1990).

Classification des connaissances

Pour la représentation des connaissances, le projet ASMV se base sur une classification des connaissances. La classification apparaît à deux niveaux. Le premier est une différenciation entre les connaissances générales sur une catégorie, que nous appelons ici « concept » et les connaissances spécifiques (objet du monde ou action précise) que nous nommons ici « instance ». Une instance est la spécification d'un concept, et ses paramètres ont des valeurs définies. Ainsi la voiture du narrateur sera une spécification du concept de voiture, qui effectuera une ou plusieurs instances des concepts d'action.

Pour ces deux catégories, plusieurs types sont ensuite définis :

Les objets : représentent des objets du monde physique, il s'agit d'éléments ayant une représentation graphique (voiture, route, panneaux, etc...).

Les actions : représentent des actions « volontaires » effectuées par les objets (accélérer, freiner, tourner, dépasser, etc...)

Les événements : représentent des interactions entre un ou plusieurs objets (chocs).

Structure des connaissances

Chaque connaissance contient un certain nombre d'informations. Ces informations concernent les connaissances elles-mêmes (description, mots (ou lexies) associés, capacités, propriétés) ou des liens entre différents objets (méronymes, holonymes, spécialisations, etc...).

Ces diverses informations peuvent être représentées sous différentes formes, correspondant aux besoins. Ces descriptions peuvent être simples : numériques, textuelles, relationnelles, ou plus complexes (expression de logique du premier ordre, expression en logique non monotone, prédicatives) et même composées.

Pour la représentation des connaissances, le projet ASMV utilise les avantages des technologies informatiques, les langages orientés objet (Van Roy & Haridi, 2004), et les bases de données relationnelles. Chaque connaissance est générée à partir d'une structure (appelée type) et est organisée dans une hiérarchie permettant de définir des relations d'héritage et de coercition calquées sur les langages orientés objets. Ces données sont enregistrées dans une base de données permettant de garantir la pérennité des relations.

4.3.2 Génération des actions

La représentation du mouvement d'acteurs virtuels a été le thème de différents travaux au cours des années. Plusieurs techniques ont été mises au point formant deux grandes familles, les mouvements pré-calculés (animation par frame, animation squelettique, modèles biomécaniques) et les mouvements générés (modèles descriptifs, modèles générateurs) (Kulpa, 2005).

Le problème de polysémie des termes d'action apparaît lors de la génération de l'animation. De plus, les paramètres de distance, de durée et d'angle des actions de la scène étant variables, le recours à une technique d'animation malléable est nécessaire. La génération de mouvements composés à partir de mouvements simples est ici utilisée. Cette technique définit un ensemble de mouvements simples, élémentaires, et décrit les actions plus complexes à partir de ces mouvements simples.

Dans le cas du déplacement d'une voiture, les capacités d'une voiture sont : l'accélération (augmenter la vitesse des roues), le freinage (ralentir la vitesse des roues) et le virage (pivoter l'axe des roues avant). Ces trois actions ne sont pas ambiguës, elles sont alors décrites comme «axiomes» de mouvements. Lors du rendu graphique, des informations concernant ces actions élémentaires sont utilisées pour faciliter et guider la génération. Ces informations décrivent le mouvement apporté aux modèles 3D servant à la représentation de l'objet et l'effet sur les paramètres de l'objet. Le système de rendu anime alors ces objets après l'application des règles apportées par le moteur physique.

Les animations plus complexes, comme le dépassement pris en exemple dans la problématique, sont décrites par les connaissances de l'objet voiture. Ici encore le raisonnement permettant la génération du mouvement est incertain. Les connaissances sur ces mouvements complexes sont donc représentées par des expressions non-monotones. La description des mouvements complexes peut de même faire appel à d'autres mouvements simples, complexes, et aux propriétés des objets.

Afin de garantir l'obtention d'au moins une extension des expressions, celles-ci sont définies en utilisant une forme dite normale (Grégoire, 1990). Dans le cas où le traitement génère différentes interprétations, chaque interprétation est traitée par le système.

4.4 Validation et correction de l'animation

La complétion des informations présentée par la section 3.2 permet d'obtenir suffisamment d'information pour générer l'animation. Toutefois ces informations sont déduites par le système, et restent donc incertaines. De plus, nous avons vu en section 2.3 certains problèmes liés à l'interprétation de la scène. C'est sur ces deux étapes que la validation évoquée doit intervenir.

4.4.1 Utilisation des technologies 3D

Utiliser un rendu en 3 dimensions présente une fois de plus un avantage. En effet, ces technologies offrent un ensemble d'outils, notamment pour la détection des positions et des collisions. Pour la validation des informations, le projet ASMV vise à la mise en place d'un système de validation de l'animation. Cette étape se base sur l'ensemble de règles représentant des connaissances complémentaires nécessaires à la compréhension.

Cette étape permet de vérifier deux points importants, les actions et événements utilisés. Par exemple, quelle que soit son mode de réalisation, l'action de dépasser doit permettre d'identifier un couple d'instant T , T' tels qu'un véhicule A est derrière un véhicule B à l'instant T et qu'il est devant à l'instant T' . Le système vérifie alors que la scène construite vérifie bien cette contrainte. Le second point porte sur les connaissances du domaine. Ainsi pour l'exemple fourni par la figure 3, nous pouvons remarquer qu'une voiture à l'arrêt au milieu d'un carrefour représente une situation anormale.

4.4.2 Sources d'erreurs

On doit disposer de connaissances fines sur ce qui est normal et sur ce qui ne l'est pas ; p.ex. une voiture arrêtée en plein milieu d'un carrefour est une anomalie, mais elle cesse de l'être si on sait qu'elle veut tourner à gauche et que des voitures arrivent en sens inverse.

Une anomalie ne signifie pas une erreur ; après tout, un accident est une anomalie, donc il doit bien y avoir des choses anormales dans la description d'une situation d'accident. Ce qui en revanche est un signe presque certain d'erreur, c'est une anomalie qui n'est pas décrite dans le texte. En effet, comme il a été dit, un texte n'a pas à préciser ce que le lecteur connaissant les normes d'un domaine reconstruit sans effort. Mais il est peu vraisemblable qu'il ne précise pas les éléments anormaux d'une situation. Donc si la visualisation mène à des situations qui violent les normes du domaine sans que le texte mentionne d'une façon ou d'une autre cette violation, c'est très probablement que la compréhension du texte est erronée.

4.4.3 De la validation à la correction

Un spectateur humain sera alerté par son sens commun des erreurs (collision, mauvaise interprétation) intervenant dans une animation. Les connaissances qu'il possède sur l'activité « normale » de la circulation routière lui permettent de faire mentalement les corrections nécessaires et de trouver les modifications nécessaires sur l'ensemble des paramètres définis par le système pour rectifier ce qui ne va pas.

Les informations incertaines du système sont les informations définies par défaut. Le défi maintenant est de définir les stratégies de modification de ces paramètres associés, selon les erreurs détectées.

Lors de la correction de l'animation, il est nécessaire de conserver une trace des modifications apportées aux paramètres. En effet, la modification de certains paramètres peut entraîner l'apparition de nouvelles erreurs. La correction de celles-ci ne doit pas revenir sur les corrections précédemment effectuées. Dans le cas contraire, le processus entrerait dans une boucle infinie.

5 Conclusion

Cet article présente le travail en cours dans le cadre du projet ASMV. Ce système vise à la mise en place d'un système de compréhension automatique de textes décrivant des accidents de la route. Les défis majeurs d'une telle entreprise sont l'insuffisance des informations explicitées dans des textes « réels », la polysémie des termes d'action et d'événements, et les erreurs d'interprétations possibles. Les solutions explorées se basent sur l'utilisation de paramètres par défaut et une description des connaissances du système ayant recours aux systèmes non monotones et particulièrement à la logique des défauts de Reiter (Grégoire, 1990) (Kayser, 1997b). Cette application tire également profit de la représentation graphique, par le biais d'une animation en 3 dimensions, comme système de simulation permettant la manipulation des informations et la correction des erreurs dans le choix des paramètres et l'interprétation du texte.

Nous avons montré ici les principales difficultés auxquelles nous sommes confrontés, les solutions adoptées ou envisagées pour les résoudre. Une première implémentation est en cours, et nous sommes d'ores et déjà capables de traduire certains des prédicats fournis par l'analyse linguistique en une scène animant correctement les objets mentionnés. Les mois qui viennent permettront de traiter l'ensemble de ces prédicats, et de diagnostiquer les anomalies de la situation visualisée.

Dans une phase ultérieure, ces travaux ,actuellement focalisés sur les constats d'accidents de la route, pourraient être développés pour d'autres applications de génération automatique d'animation : confrontation de récits ou de témoignages sur un événement par la génération d'une animation à partir de diverses sources d'informations, outils pédagogiques pour l'apprentissage de la narration, système de conception industrielle aidée par ordinateurs.

6 Remerciements

Je souhaite remercier Daniel KAYSER et Nabil EL-KADHI pour leurs conseils lors de la rédaction de cet article et pour leur intérêt et la confiance qu'ils me portent concernant le projet ASMV.

Je remercie également la compagnie de responsabilité civile MAIF pour la mise à disposition des constats d'accidents de la route du corpus.

Références

- ARDONI G. & DI MANZO M. ET AL. (1984). Natural Language driven Image Generation. *Proceedings of the 22nd annual meeting on Association for Computational Linguistics*. p. 495-500. Stanford.
- BOURG D. (2002). *Physics for game developers*. O'Reilly.
- COYNE B. & SPROAT R. (2001). WordsEye : an automatic text-to-scene conversion system. *{SIGGRAPH} 2001, Computer Graphics Proceedings*. p. 487-496.
- DUPUY S., EGGES A. ET AL. (2001). Generating a 3D simulation of a Car Accident from a Written Description in Natural Language: the CarSim System, *Annual meeting of ACL*, 2001
- FALARDEAU E. (2003). Compréhension et interprétation : deux composantes complémentaires de la lecture littéraire, *Revue des sciences de l'éducation*, Volume 29, numéro 3.
- GRÉGOIRE E. (1990). *Logiques non monotones et intelligence artificielle*. Hermès.
- JACQUET G., VENANT F. ET AL. (2005). Polysémie lexicale. *Sémantique et traitement automatique du langage naturel*. p. 99-132, Hermès-Lavoisier.
- JOHANSSON R., BERGLUND A. ET AL. (2005). Automatic text to scene conversion in the traffic accident domain. *Proceedings of IJCAI*. p. 1073-1078.
- KAYSER D. (1997a). La sémantique lexicale est d'abord inférentielle. *Langue Française*. Numéro 113, p. 92-106.
- KAYSER D. (1997b). *Représentation des connaissances*. Hermès.
- KULPA R. (2005). *Adaptation interactive et performante des mouvements d'humanoïdes synthétiques : aspects cinématique, cinétique et dynamique*. Thèse soutenue à l'INSA de Rennes.
- LAFONTAINE D. (2003). Comment faciliter, développer et évaluer la compréhension des textes aux différentes étapes de la scolarité primaire ? *Conférence de consensus PIREF*.
- LE NY J.F. & KINTSCH W. (eds.) (1982). *Language and Comprehension*. North Holland.
- MICRAC : <http://www.irit.fr/MICRAC/>.
- NOUIOUA F. (2007). *Extraction et utilisation des normes pour un raisonnement causal dans un corpus textuel*. Thèse soutenue au LIPN, Université Paris13.
- OGRE3D : <http://www.ogre3d.org/>.
- REITER R. (1980). A logic for default reasoning. *Artificial Intelligence Journal*. Volume 13 (1-2). p. 81-132.
- SNOW C. (2002). *Reading for Understanding Toward an R&D Program in Reading Comprehension*. RAND.
- VAN ROY P. & HARIDI S. (2004). *Concepts, techniques, and models of computer programming*. M.I.T. Press.