Chapitre 9

# L'indexation conceptuelle de documents multilingues et multimédias

## 9.1. Introduction

Cet article décrit le rôle que peut jouer le traitement automatique du langage (TAL) pour l'indexation conceptuelle de documents multimédias et multilingues, et ainsi permettre une recherche intelligente au sein d'archives digitales de tels documents. Nous allons ici prendre comme exemple les techniques utilisées par les systèmes d'extraction de l'information (EI)[1] et donc nous limiter à l'analyse linguistique dite superficielle (*shallow analysis*)[2] de documents, qui est favorisée par les systèmes EI.

Dans un premier temps nous allons décrire les techniques de base utilisées dans un système d'extraction de l'information, une discipline d'ingénierie linguistique visant à identifier, rassembler et normaliser les informations pertinentes pour des utilisateurs ou des applications spécifiques. L'information extraite est typiquement représentée sous formes de formulaires (*templates*) pré-définis et remplis par les résultats de l'analyse linguistique des documents concernés.

---

[1] Pour plus de détails, voir l'introduction générale proposée par [APP99] et les actes des conférences MUC (Message Understanding Conference) [MUC95] et [MUC98], ainsi que [GRI96].

[2] Le mot « superficiel » est ici à entendre au sens de « surface ». Il s'agit d'une analyse qui, par exemple, ne cherche pas à résoudre tous les problèmes d'attachements de fragments de texte. Cela peut se faire à des niveaux ultérieurs du traitement du document.

Les systèmes d'extraction d'information sont généralement constitués de deux composants principaux. D'une part un ensemble d'outils d'analyse linguistique superficielle et/ou partielle[3], définissant le noyau stable des systèmes unilingues, et d'autre part des représentations conceptuelles hiérarchisées offrant un modèle des domaines d'application définis. En charnière entre ces deux composants sont définis un lexique (ou une terminologie) propre au domaine d'application et certaines sous-grammaires spécialisées qui servent essentiellement à identifier ce qu'on appelle les « entités nommées » (*Named Entities*: noms de personnes ou d'institutions, lieux géographiques, dates etc.) qui jouent un rôle important dans de nombreux domaines d'application, ainsi que cela est amplement montré dans l'histoire des conférences MUC.

A titre d'exemple des techniques de l'extraction d'information, nous présentons le système SMES développé au DFKI pour l'allemand[4]. La partie linguistique de ce système doit faire face à des phénomènes liés à la langue en question. Ainsi, l'allemand ayant une morphologie relativement riche, le système devra disposer d'une analyse morphologique performante. En revanche, les représentations conceptuelles hiérarchisées (ou ontologies) intégrées dans le système sont indépendantes de la langue particulière, et modèlent les domaines d'application du système EI. Les hiérarchies conceptuelles sont à même de constituer en quelque sorte une relation abstraite entre différents systèmes EI unilingues pour une application donnée et l'extension d'un système d'extraction d'une langue à une autre visera surtout à établir une relation entre les expressions clés dans une langue à leur équivalent conceptuel dans l'ontologie. Cette hiérarchie conceptuelle fait donc en quelque sorte office d'*interlingua* dans le cadre d'une application donnée, facilitant par exemple la traduction automatique de parties pertinentes de documents.

Dans un deuxième temps, nous allons adresser le thème du multimédia (en abordant aussi le multimodal), dont certaines applications jouent un rôle croissant dans notre vie quotidienne comme, par exemple dans l'éducation, l'organisation du temps libre, le commerce ou encore les transports. L'utilisation de systèmes multimédias est motivée, d'une part, par les avantages procurés par un média particulier pour résoudre un problème donné, et d'autre part, par le fait que dans certains cas, l'utilisation de certains médias en parallèle peut procurer une plus grande flexibilité et efficacité pour l'utilisateur. Une question centrale qui se pose au

---

[3] Sur la base du principe « Un peu d'information est plus souhaitable que pas d'information du tout », les systèmes EI ne réclament pas des analyses linguistiques complètes. Par ailleurs, des fragments de documents ne pouvant être analysés à ce niveau, peuvent être intégrés dans des processus d'analyse à un niveau ultérieur. L'essentiel est que l'analyse linguistique du système EI ne soit pas interrompue par des messages d'erreurs : cette analyse linguistique doit être robuste.

[4] Pour une présentation plus détaillée, voir [NEU97], [NEU00] ou encore [PISK00]

développement du multimédia/multimodal concerne l'interface d'utilisation, et il s'avère que le langage naturel restant le moyen de communication largement favorisé par l'être humain, les techniques d'analyses automatiques du langage doivent constituer une partie intégrante de tels systèmes, si ceux-ci veulent vraiment offrir une interface d'utilisation interactive.

Ensuite nous décrirons plus en détail un projet en cours (MUMIS[5]) portant sur l'indexation conceptuelle automatique de documents multimédias, se basant sur les résultats de l'extraction multilingue d'information, offrant ainsi des possibilités plus sophistiquées de recherche au sein d'archives multimédias, dans la mesure où l'indexation sera également motivée par des analyses linguistiques poussées, incluant la détection d'expressions coréférentielles. Ce projet offre, par ses buts, une concrétisation adéquate des discussions proposées lors des chapitres précédents. Le projet MUMIS est également intéressant, dans le sens où il fait usage de multiples sources d'information en différentes langues pour offrir une indexation aussi complète que possible. Cette multiplicité constitue un challenge pour les systèmes d'extraction d'information concernés. Ainsi, cette diversité de médias et de langues réclame l'intervention de techniques visant à assurer la consistance de l'information extraite et à en éliminer les redondances.

## 9.2. L'extraction d'information

La quantité d'information disponible sous forme électronique n'ayant cessé de croître au cours des dernières années, il a fallu envisager le développement d'outils informatiques permettant l'accès ciblé et efficace à cette masse d'information. Plusieurs technologies sont concernées par ce problème, parmi elles la recherche d'information (*Information Retrieval*) et l'extraction d'information (*Information Extraction*).

### 9.2.1. *Recherche d'information et extraction d'information*

Là où la tâche principale de la recherche d'information consiste à identifier, au sein d'un ensemble de documents, ceux qui sont pertinents pour un utilisateur, l'extraction d'information vise à extraire l'information pertinente au sein des documents mêmes, et à la présenter sous une forme normalisée à l'utilisateur. Dans le premier cas, l'utilisateur fera usage de requêtes plus ou moins complexes en langage naturel et le système calculera la pertinence des documents par rapport à la

[5] MUMIS (*Multimedia Indexing and Searching Environment*) est un projet financé par l'Union Européenne dans le cadre du programme « *Information Society Program* (IST) », section « *Human Language Technology* (HLT) ». Voir aussi pour plus d'information http://parlevink.cs.utwente.nl/projects/mumis/ ou encore [DEC01]

requête formulée, utilisant pour ce faire, des techniques d'indexation des documents ne réclamant pas d'analyse linguistique poussée. Dans le second cas, l'utilisateur devra décrire plus en détail l'information pertinente pour lui et le domaine d'application, et cela sous formes d'entités, de relations entre ces entités et d'événements impliquant celle-ci. Pour ce faire, l'utilisation de formulaires (*templates*) s'est imposée: chaque champ d'un formulaire représente un type d'information désiré pour le domaine d'application, et l'analyse linguistique du document devra essayer de remplir ce champ. Ainsi, par exemple, les changements de postes au sein de grandes entreprises: Pour détecter les informations pertinentes, if faut identifier les personnes, les entreprises et les liens entres personnes et entreprises, ainsi que les évènements impliquant ces entités (départ, arrivée et remplacement de personnes pour un certain poste dans une certaine entreprise à une certaine date). La plupart des systèmes EI organisent ces formulaires également sous forme de hiérarchies conceptuelles (ontologies), permettant ainsi des opérations d'inférence sur des formulaires partiellement remplis par l'analyse linguistique, ou encore de vérifier la consistance de l'information extraite sur plusieurs formulaires. L'utilisation de formalismes basés sur l'unification de structures de traits typées[6] pour représenter la hiérarchie conceptuelle propre à un domaine d'application permet en outre de formuler des contraintes sur la valeur possible de certains champs des formulaires ou d'établir le partage de certaines valeurs au sein de formulaires.

Il est clair que l'extraction d'information réclame une analyse linguistique bien plus sophistiquée que dans le cas de la recherche d'information. Les systèmes EI doivent être à même de reconnaître toutes sortes d'entités et de relations pertinentes décrites dans les documents, ce qui dans de nombreux cas exige, par exemple, la détection de fonctions grammaticales (sujet, objet direct etc.) au sein d'une phrase, ou encore la reconnaissance d'expressions coréférentielles à travers tout le document. [CUN99] a pertinemment situé l'extraction d'information entre la recherche d'information et la compréhension (automatique) des textes, dans la mesure où l'extraction d'information doit se baser sur des analyses linguistiques poussées, mais ne requiert pas une analyse complète des documents, et échappe ainsi aux difficultés rencontrées par les systèmes d'analyse linguistique dite « profonde » (*deep analysis*).

Mais il ne faut pas perdre de vue que les deux technologies mentionnées sont souvent appelées à collaborer dans le cadre de systèmes complexes : ainsi la requête d'information peut faire office de filtre pour l'extraction d'information et *vice versa*.

---

6 Tel par exemple TDL (Type Description Formalism), décrit dans [KRI94].

### 9.2.2. *Les tâches spécifiques de l'extraction d'information*

Les conférences MUC, déjà mentionnées, ont au fil des différentes éditions, contribué à l'élaboration de certains standards pour l'extraction d'information et ont ainsi également facilité la définition de critères pour l'évaluation des technologies impliquées. Nous donnons ci-dessous la liste des différentes tâches de l'extraction d'information, telles qu'elles ont été formulées dans le cadre de MUC-7:

– les entités nommées (*Named Entities*, NE) : il s'agit ici de reconnaître et de marquer dans le texte toutes les séquences de mots qui représentent une personne, une organisation, une location, une date ou un espace de temps, une devise ou une quantité, etc. (la classification des entités nommées pertinentes varie en fonction des domaines sous considération) ;

– les éléments des formulaires (*Template Element task*, TE) : cette tâche est concernée par l'extraction de l'information de base qui peut être mise en relation avec les organisations, les personnes et autres entités concrètes reconnues par la tâche NE. Cette information peut se trouver à n'importe quel endroit du texte analysé ;

– les relations au sein de formulaires (*Template Relation task*, TR) : ici les systèmes doivent extraire des informations relationnelles du type „employé de“, „fabrique de“, „lieu où“ etc. Ce sont essentiellement les entités nommées qui sont mises en relation ;

– les scénarios au sein de formulaires (*Scenario Template task*, ST) : reconnaît et extrait les informations concernant des événements préspécifiés en fonction du domaine d'application et indique le rôle joué par les différentes entités nommées reconnues auparavant ;

– la coréférence (*Coreference task*, CO) : ce processus capture l'information sur les différentes expressions dans le texte partageant une seule et unique référence. En fait, il ne s'agit pas vraiment d'un module séparé : la détection d'expressions coréférentielles est distribuée sur toutes les tâches particulières de l'EI.

En observant cette description des différentes tâches qui constituent l'extraction d'information, on peut remarquer que la linguistique joue un rôle à des degrés divers : pas ou peu pour la tâche NE (des grammaires pour expressions régulières peuvent suffire), jusqu'à des calculs syntaxiques assez complexes pour la détection de la coréférence, en passant par la reconnaissance de fonctions grammaticales pour la tâche ST.

### 9.2.3. *L'extraction d'information et la génération d'annotations*

L'extraction d'information ne doit pas se limiter à la génération et à l'instantiation de formulaires. Les informations gagnées peuvent également être utilisées pour

produire des annotations qui peuvent servir à marquer directement les documents. Ceci est important car les systèmes EI peuvent proposer, en plus des traditionnelles balises morpho-syntaxiques, des balises conceptuelles, même si celles-ci sont limitées au domaine d'application. En plus de cette possible contribution de l'EI pour la linguistique de corpus, qui a par ailleurs fait l'objet d'une journée de discussion lors d'une conference LREC[7], ce marquage conceptuel permet également de supporter la tâche de la désambiguïsation lexico-sémantique.

Ce ne sont pas seulement les corpus textuels qui peuvent profiter des annotations conceptuelles produites par les systèmes EI : les archives multimédias également gagnent à être décorées de ce type d'annotation, ce qui permet pour des requêtes d'un certain niveau d'abstraction de trouver les documents audiovisuels aux contenus désirés, ou même d'extraire d'un document audiovisuel les séquences contenant les contenus pertinents pour l'utilisateur.

### 9.3. SMES - Un système d'extraction d'information pour l'Allemand

SMES est un système d'extraction d'information pour l'allemand, qui est conçu pour une configuration rapide et aisée en vue de l'application à de nouveaux domaines. Pour ce faire, le système opère une séparation entre le traitement systématique des connaissances linguistiques générales d'un coté et la représentation des connaissances spécifiques à un domaine de l'autre, car il s'est avéré que les seuls outils d'analyse linguistique n'offrent qu'un degré limité d'adaptation. Pour le premier composant, le système utilise un processeur de textes rapide et robuste (permettant uniquement au besoin une analyse partielle du texte). Ce processeur, indépendant des domaines d'application représente le texte analysé sous forme d'une séquence de descriptions fonctionnelles partielles et sous-spécifiées. Pour le second composant, SMES utilise un formalisme d'unification de structures de traits typées, appelé TDL[8].

SMES a développé un modèle pour l 'intégration, en combinant à un niveau abstrait les différents types de représentations (syntagmes, domaines, fonctions grammaticales), des sources de connaissances linguistiques générales et les sources de connaissances spécifiques à un domaine. Cette intégration s'opère de manière purement déclarative. La stratégie adoptée pour permettre l'adaptation rapide a de nouveaux domaines a fait ses preuves. Il est possible d'adapter le système EI à un nouveau domaine en moins de deux mois (une personne/mois) sur la base de la seule description conceptuelle du nouveau domaine et de l'établissement d'une liste de termes clés. Ceci comporte aussi l'immense avantage de permettre d'associer

[7] Voir [MCN00] et [DEC00]
[8] Voir [KRI94]

différents systèmes EI unilingues à un domaine d'application, le rapport s'établissant *via* la représentation abstraite et conceptuelle du domaine sous considération[9].

### 9.3.1. *Les outils d'analyse linguistique*

La chaîne des outils linguistiques consiste en un segmenteur (décrivant environ 60 classes de segments), un analyseur morphologique (incluant une analyse des composés), un étiqueteur, un détecteur d'entités nommées et un analyseur de fragments (syntagmes nominaux et prépositionnels, groupes verbaux), ainsi qu'un analyseur de dépendance sur la base duquel une heuristique pour la détection de fonctions grammaticales a été définie. Un algorithme pour la détection d'expressions coréférentielles intervient à plusieurs niveaux de cette chaîne de traitement. La figure 9.1 donne un exemple de la sortie de l'analyse linguistique. Nous appelons cette structure une « description fonctionnelle sous-spécifiée ».

[*NP_subj* La situation] [*pp* pour l'équipe belge] [*NP_pron_obj* se] [*VG* complique], [*SUBORD* après que [*NP_appos-subj* son gardien de but, Philippe De Wilde,] [*VG_pass* ait reçut] [*NP_obj* la carte rouge] [*NP_time* (80.)].]

***Figure 1**.Résultats de l'analyse de fragments du parseur (superficiel) de dépendance, incluant la détection de fonctions grammaticales*

Les résultats de l'analyse linguistique sont également disponibles en format XML.

### *9.3.2. La modélisation du domaine d'application*

Les connaissances spécifiques au domaine sont modelées à l'aide du formalisme TDL (*Type Description Language*)[10]. Il s'agit d'un formalisme d'unification de structures de traits typées, supportant toutes sortes d'opérations sur ce genre de structures. Ainsi, il est possible de modeler de manière déclarative le domaine d'application sous la forme d'une hiérarchie conceptuelle (ontologie) de structures de traits non spécifiées qui représentent en fait les formulaires qui doivent être remplis. On peut interpréter ce mécanisme comme la possibilité de décrire des contraintes sémantiques sur le résultat de l'analyse syntaxique (voir la cadre droit de la partie gauche de la figure 9.2).

---

[9] Cet exercice d'extension multinlinguale de l'extraction d'information est actuellement en cours dans le projet MUMIS, décrit plus en détail postérieurement.
[10] Voir [KRI94]

Étant donné que le système utilise un formalisme linguistique de haut niveau pour la représentation conceptuelle du domaine, il n'est pas difficile d'établir les rapports entre l'analyse superficielle du texte, le lexique spécialisé (terminologie) et l'ontologie du domaine, ainsi que cela est montré dans la figure 9.2, où le lecteur peut voir que cette relation est définie de manière déclarative à un haut niveau d'abstraction.

Le modèle du domaine d'application peut être intégré dans une structure ontologique lexicale plus générale (comme WordNet). Ainsi un « attaquant » peut être décrit comme étant un « joueur », lui-même étant un « être humain » etc. Ceci peut contribuer à une désambiguïsation lexico-sémantique.

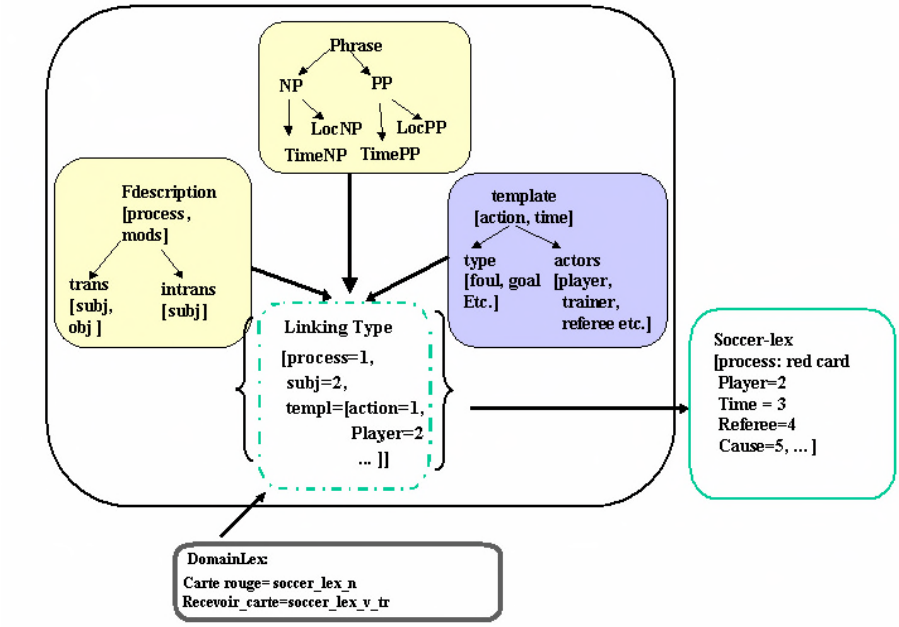


**Figure 2**.*Toute la chaîne de traitement menant au « remplissage du formulaire »*

## 9.4. Rôle du TAL dans les applications multimédia

[AND00a] et [AND00b] présentent une vue d'ensemble du rôle joué par le langage naturel dans les systèmes multimédias et multimodaux. Nous résumons d'abord quelques points centraux de cette analyse et considérons plus particulièrement le

rôle que la technologie particulière de l'extraction d'information peut jouer dans ce contexte.

### 9.4.1. *Systèmes multimédias et multimodaux*

Les termes multimédia et multimodal sont souvent source de confusion. Nous adoptons ici la définition proposée par [MAY99] qui différencie « médium », « modalité » et « code ». Le terme modalité réfère au type de perception concernée, qu'il s'agisse par exemple, de perception visuelle, auditive ou olfactive etc. Le terme médium réfère au moyen de diffusion de l'information (CD-ROM, papier etc.), aux matériels utilisés (microphone, écran, haut-parleur etc.) ainsi qu'aux différents types d'information (texte, séquences audio ou vidéo). Le terme code réfère au moyens particuliers utilisés pour codifier l'information (langage de signes, langage pictural). On peut parler d'un système multimédia ou multimodal si celui-ci permet soit de générer ou d'analyser de l'information multimédia/multimodale, soit d'accéder à des archives digitales constituées de plusieurs médias.

Souvent dans les applications existantes, le processus d'analyse ne s'applique qu'à des données multimodales[11], alors que le processus de génération est concerné par la production d'information multimédia[12].

### 9.4.2. *Intégration de modalités*

Dans le cas de l'analyse, on parle d'un processus d'intégration de modalités dans la mesure où toutes les données modales doivent être fusionnées à un niveau plus abstrait afin de tirer un profit maximum des différentes modalités engagées[13]. Des formalismes de représentation tels qu'ils sont définis dans certaines applications du TAL peuvent jouer un rôle important dans ce processus d'intégration d'informations délivrées par plusieurs modalités. Ainsi les techniques d'unification de structures de traits typées, en combinaison avec un analyseur tablé, sont utilisées dans un système multimodal décrit par [JOH98]. A l'aide de ce formalisme il est possible de construire une représentation sémantique commune à toutes les modalités

---

[11] Mais l'analyse peut aussi s'appliquer à des données multimédias, ainsi que la description du projet MUMIS va le montrer par la suite.

[12] La génération d'information multimodale ne sera pas possible avant d'avoir résolu le problème de la génération automatique de gestes ou de mimiques.

[13] Cette fusion concerne des données synchrones et permet eventuellement d'opérer une désambiguïsation au niveau d'une certaine modalité, en la mettant en rapport avec la somme des autres modalités impliquées. Le project MUMIS, décrit plus loin, cherche à offrir une fusion, et également une désambiguïsation, au niveau des médias, en consultant également des données asynchrones.

concernées sur la base de l'unification des structures de traits représentant la contribution sémantique des modalités particulières.

### 9.4.3. *Coordination de médias*

Dans le cas de la génération multimédia, on parlera d'un processus de coordination de médias. En effet, il ne suffit pas de fusionner différents médias pour obtenir une présentation cohérente de données multimédias : les différentes informations contenues dans les différents médias doivent être précieusement mises en rapport entre elles afin d'obtenir une réelle complémentarité des médias, rendant justice à l'apport particulier de chacun d'entre eux. Les expériences gagnées par les systèmes de génération du langage naturel peuvent s'avérer fort utiles pour l'élaboration de systèmes automatiques de présentation multimédia, dans la mesure où les systèmes de génération en langage naturel sont depuis toujours confrontés à la tâche de sélection et d'organisation du contenu, aussi appelée planification, précédant la réalisation textuelle. Essentiellement deux concepts issus de la génération automatique du langage naturel sont utilisés dans les systèmes de présentation multimédia : celui de schéma (voir [MCK85]) et celui d'opérateur (voir [MOO85]). Le premier concept décrit une série de patrons standards du discours à l'aide de prédicats qui indiquent les relations existant entre les différentes parts de la présentation. Le second concept insiste plus particulièrement sur les contributions particulières des différentes parts de la présentation et permet des révisions purement locales lors du processus de génération.

### 9.4.4. *Accès en langage naturel aux archives digitales multimédias*

Enfin, concernant l'accès à des archives digitales multimédias, il s'avère que le langage naturel peut jouer à plusieurs égards un rôle important. Tout d'abord il est plus facile d'accéder à l'information contenue dans ces archives en adressant des requêtes aux séquences audio (en fait à la transcription de celles-ci) ou aux sous-titres éventuellement associés aux séquences vidéo que d'analyser les images elles-mêmes. Le traitement automatique du langage naturel est en effet bien plus avancé que celui du contenu d'images. Ensuite il s'avère qu'il est bien plus commode d'accéder à des données visuelles à l'aide du langage naturel, car celui-ci autorise une formulation plus flexible et efficace de requêtes. Et enfin, le langage naturel offre une bonne possibilité de condensation de l'information visuelle.

Afin d'accéder à l'information visuelle, certains systèmes font usage d'une analyse simple des données linguistiques associées aux images, ainsi la transcription des commentaires ou encore les sous-titres. Dans la plus grande partie des cas, cela

suffit pour proposer une classification (et une indexation) des données visuelles[14]. Certains systèmes vont plus loin et utilisent les analyses linguistiques plus complexes propres à l'extraction d'information. A l'aide par exemple de la reconnaissance des entités nommées et de séquence linguistiques standards, les systèmes d'accès à l'information multimédia peuvent filtrer des séquences non pertinentes (par exemple les introductions des commentateurs). Ceci est le cas du système « Broadcast News Navigator » développé par MITRE (voir [MER97]) .

Le projet MUMIS, décrit par la suite, franchit une étape supplémentaire, dans la mesure où une multiplicité de documents dits collatéraux à des séquences vidéos[15] (un ensembles de textes hétéroclites consacrés à une rencontre de football) est traitée par des systèmes EI et les annotations conceptuelles unifiées produites par l'analyse sont utilisées pour indexer le matériel vidéo, autorisant donc une complexe requête sur l'ensemble des archives multimédias.

### 9.5. Challenges pour le TAL

Nous avons vu dans ce chapitre que des méthodes et techniques propres au TAL peuvent être étendues afin d'être utiles aussi pour des applications multimédia. Ainsi, les grammaires d'unification peuvent servir de base pour l'agencement et l'analyse de données multimodales et multimédias. La planification du contenu à délivrer propre à la génération automatique du langage naturel peut être appliquée pour la sélection et la structuration du contenu dans le cadre de la génération automatique de présentations multimédia. Et enfin l'utilisation de techniques d'analyse propres à la requête de documents ou encore à l'extraction d'information permet un accès plus complexe et flexible aux archives digitales multimédias.

Toutefois le TAL doit de son côté aussi s'adapter aux nouvelles technologies et s'ouvrir au nouvelles formes de communication offertes par les développements dans le domaine du multimédia et du multimodal. Une dimension nouvelle nous apparaît dans le rôle que peut, et doit jouer, le TAL : il ne s'agit plus seulement d'analyser ou de générer des textes, mais d'offrir un niveau d'analyses tel que les résultats peuvent être utilisés pour l'agencement d'un complexe multimédia autour du langage naturel.

---

[14] Voir ici par exemple les projets européens Olive et Pop-Eye, [JON98] et [JON00], ou encore [DJO98].

[15] Un exemple d'un tel usage de documents collatéraux pour l'indexation de vidéo est également donné par [SAL98], qui indexe ainsi des séquences de scènes de ballet.

Nous pensons que ceci est exactement le point que le projet MUMIS adresse, dans la mesure où un de ses buts principaux réside dans la génération d'annotations formelles, à divers degré d'abstraction, autour desquels une organisation de données multimédias peut se faire, autant pour la sauvegarde de ces données que pour leur accès ultérieur, que soit à l'aide d'hiérarchies conceptuelles ou directement par l'usage du langage naturel.

## 9.6. MUMIS -- un environnement pour l'indexation et la recherche de données multimédia

Le projet MUMIS a comme but principal de développer et d'intégrer des technologies de bases qui supportent l'indexation conceptuelle automatique de données vidéo et permettent la recherche de contenu dans des archives digitales multimédias. Le projet examine de manière plus précise le rôle que des annotations résultant d'analyses linguistiques poussées, combinées avec des informations spécifiques au domaine d'application (ici le football) peut jouer pour indexer de longues séquences vidéo (une rencontre de football). Il s'agit de montrer qu'une extension des techniques de l'extraction d'information ainsi que des technologies innovatrices, peuvent opérer sur des données multilingues, multisources et multimédias.

MUMIS est composé de deux composants : l'un « hors ligne », qui est responsable de la génération automatique d'annotations formelles[16] pour l'indexation conceptuelle du matériel vidéo. Cette indexation se fait sur la base d'information temporelle[17] extraite des multiples documents analysés par les systèmes EI. L'autre « en ligne » (voir figure 9.3) qui est responsable de l'accès en temps réel aux archives multimédias annotées par le premier composant. Ici, nous nous concentrons sur le partie hors ligne.

---

[16] Formelles, au sens où il est fait abstraction de la réalisation textuelle concrète d'évènements pertinents.

[17] Cette information temporelle concerne essentiellement les moments importants d'une rencontre.

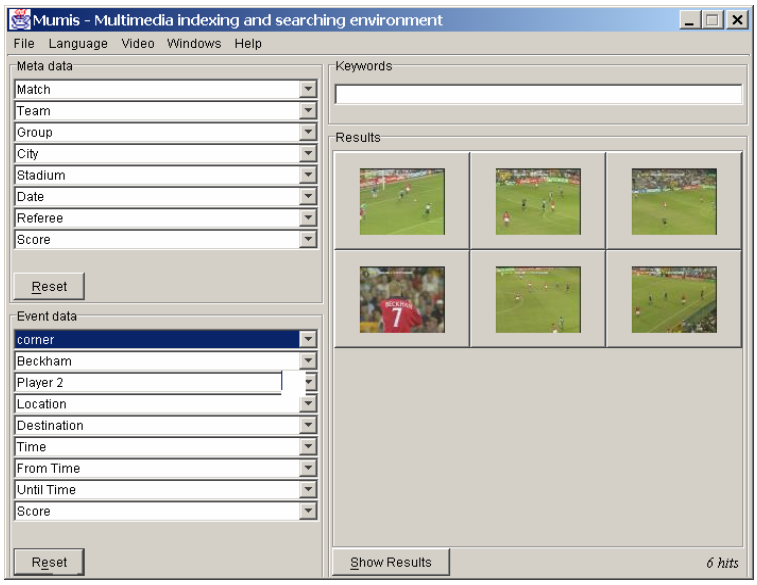


**Figure 3**. *L'interface d'utilisation de MUMIS, ou l'on peut voir que l'action intéressante est le corner, et que l'utilisateur peut visionner tous les corners donnés par le joueur Beckham. L'utilisateur peut voir la séquence vidéo correspondante en clickant sur un résumé d'image. Toutes les annotations (actions, joueurs, temps etc.) pour l'indexation du film sont générées par les systèmes EI*

### *9.6.1. Données multilingues, multi-sources et multimédia*

Afin d'autoriser une indexation aussi complète que possible des séquences vidéos, il est fait usage de données provenant de plusieurs médias (documents textuels, commentaires radio et TV[18]) en différentes langues (Néerlandais, Anglais et Allemand), ce qui permet la construction d'un large ensemble de lexiques unilingues spécialisés et d'une ontologie vraiment représentative du domaine en question.

Etant donné que le projet traite plusieurs médias en plusieurs langues, un nouveau type d'outil s'avère être nécessaire pour assurer la fusion de ces informations et de combiner correctement les annotations qui entretiennent des relations sémantiques. Cet outil doit détecter autant que possible les inconsistances et les redondances.

[18] Les commentaires sont digitalisés et disponibles sous forme de transcriptions.

Les données que MUMIS doit traiter peuvent être classifiées de la manière suivante :

– reportages provenant de journaux quotidiens en ligne, que nous considérons comme textes libres ;

– tickers, sous-titres et énumérations formelles d'événements, que nous considérons comme formant des textes semi-formels[19] ;

– Des descriptions formelles de rencontres particulières, que nous considérons comme étant des textes formels[20] ;

– Matériel audio provenant de programmes radio ou TV ;

– Matériel vidéo provenant de programmes TV.

Les quatre premières sources sont utilisées pour la génération automatique d'annotations qui servent à indexer la cinquième source.

Dans la mesure où les informations contenues dans les textes formels peuvent être considérées comme validées, elles jouent un rôle important pour les différents systèmes d'extraction et pour le contrôle de la fusion de l'information extraite. Mais ces textes ne contiennent que très peu d'information (les buts, les substitutions, les cartons distribués etc.). Ce type de texte n'offre donc qu'un nombre limité de repères temporels pour l'indexation, ce qui limite très fort la liberté de requête de l'utilisateur. Les textes semi-formels, quant à eux, offrent bien plus de repères temporels, qui sont aussi liés à une plus grande diversité d'évènements (incluant, par exemple, les remises en touche, les buts refusés, les fautes commises etc.). Mais la qualité des textes est assez pauvre, particulièrement dans le cas des textes vraiment produits en ligne (dans le cas, par exemple, des tickers). Les textes libres offrent une bien meilleure qualité, mais l'extraction de repères temporels est plus difficile. Ces textes, en général, offrent surtout des informations complémentaires aux actions sur le terrain, qui peuvent être intéressantes pour les passionnés de football (l'âge du joueur, sa valeur financière, son état de santé etc.).

Il reste encore à examiner le rôle concret que l'analyse des transcriptions des commentaires peut jouer dans la fusion de l'information.

### 9.6.2. *Extension des technologies d'extraction de l'information*

Les systèmes d'extraction d'information engagés dans MUMIS[21] doivent subir certaines extensions afin de pouvoir répondre aux nouvelles exigences. L'extension

---

[19] Ces textes contiennent encore d'authentiques expressions en langage naturel.

[20] Ces documents sont des résumés très courts des rencontres et ne contiennent plus de phrases entières.

multilingue se fait essentiellement sur la base du mécanisme décrit dans les sections d'introduction et sur SMES. L'outil de fusion examine alors si les différentes annotations produites par les systèmes peuvent être unifiées en un formulaire, ou en une table d'événements qui sera la base de l'indexation du matériel vidéo. En cas d'inconsistance, l'outil de fusion doit prendre des décisions et se base pour cela ,entre autres, sur le modèle du domaine d'application qui, combiné à un mécanisme d'inférences, permet de déclarer hors-la-loi des annotations générées. La fusion fonctionne ainsi comme un filtre. L'outil peut également contrôler l'origine des annotations, et accordera plus d'importance à une annotation validée par un texte formel[22] .

L'extraction d'information dépasse de fait maintenant les limites du document. Ceci offre également une chance pour l'amélioration des performances sur certaines tâches, telle celle de la coréférence. En effet, les modules responsables pour la détection d'expressions coréférentielles peuvent maintenant avoir accès à des banques de données dynamiques créées lors de l'analyse des différents textes, et surtout lors de l'analyse des textes dits formels.. Ceci est particulièrement vrai dans le domaine sportif où souvent les journalistes se réfèrent aux athlètes à l'aide d'expressions poétiques telles « la perle noire » ou « le Kaiser »[23] pour lesquelles il est difficile de calculer la référence au sein même du document dans lequel cette expression est utilisée. Mais en sachant, par exemple, sur la base de l'analyse d'un texte formel, que Beckenbauer a marqué un goal à la trentième minute, il pourra être possible de résoudre la référence sous-jacente à une expression du type « Le Kaiser (30.) ne laisse aucune chance au gardien hollandais. », et cela quelque soit la langue du document considéré.

### *9.6.3. Synchronisation de la séquence vidéo et des annotations formelles*

Les séquences vidéos sont digitalisées et disponibles dans le format standard MPEG-2[24]. MUMIS est particulièrement intéressé par MPEG-7, qui codifie en

---

[21] Un descendant de SMES, décrit plus haut, pour l'allemand et le néerlandais, et le système de l'Université de Sheffield pour l'Anglais ([HUM98]).

[22] Où l'on voit que les systèmes IE doivent garder traces des types de documents traités, et ainsi délivrer des méta-données, ce qui est également nouveau dans le contexte de l'extraction d'information.

[23] Dans ce cas il est intéressant de noter que ce genre de reportage utilise fréquemment des expressions en une langue étrangère.

[24] MPEG (*Moving Picture Coding Expert Group*) est un groupe de travail dans le domaine des standards ISO. MPEG-1 et MPEG-2 sont des standards pour la sauvegarde d'images animées et de matériel audio, ainsi que pour l'accès à cette information. MPEG-4 et MPEG-7 sont des extensions conceptuelles, dans la mesure où ces standards permettent la codification

XML les informations propres aux images, et qui a défini une interface pour l'intégration d'annotations textuelles.

Pour toutes les annotations produites par l'extraction d'information et filtrées par l'outil de fusion, if faudra vérifier si les données temporelles peuvent être synchronisées avec celles de la séquence vidéo digitalisée. Il semble clair qu'une certaine déviation devra être acceptée, car les informations temporelles incluses dans les annotations ne peuvent être aussi précises que celles présentes dans la codification MPEG de la séquence vidéo. MUMIS devra apporter une réponse à la question de savoir quelle déviation est tolérable. MUMIS va également examiner si le standard MPEG-7 peut apporter une aide substantielle à la résolution du problème de la mise en relation d'annotations textuelles et conceptuelles et la codification des séquences d'images. Ici une coopération ultérieure avec des projets et des groupes de recherche concernés par la segmentation automatique de vidéos devrait apporter des améliorations substantielles. Nous pensons ici particulièrement au projet ASSAVID (*Automatic Segmentation and Semantic Annotation of Sports Videos*), qui traite du même domaine que MUMIS et qui extrait des informations sémantiques sur la base de l'analyse de caractéristiques de l'image (voir [ASSF02]).

## 9.7. Conclusions

Nous avons discuté le rôle que le traitement automatique du langage peut jouer, à divers degrés, dans le cadre de l'annotation conceptuelle de données multimédia. En guise de conclusion, nous énumérons les points où le TAL en général et MUMIS en particulier[25], offrent des contributions à l'extraction et à l'accès de contenu multimédia :

– traitement de sources d'information multimédias et multilingues ;

– indexation conceptuelle par l'application de systèmes EI à des domaines bien définis et par l'usage d'information déjà analysée comme contrainte

– utilisation et extension de technologies TAL innovatrices pour la génération automatique d' annotations formelles correspondant à du contenu multimédia ;

– fusion d'informations en provenance de multiples sources pour l'amélioration de la qualité des annotations formelles ;

---

du contenu d'images sous forme d'objet (MPEG-4) et autorisent la recherche de contenu dans les images (MPEG-7). Voir ici [DAY01].

[25] Si le but principal de MUMIS, comme exemple concret de ce type d'application, réside dans l'indexation de séquences vidéo, il n'en reste pas moins que l'annotation conceptuelle comme résultat de l'extraction multilingues d'information et du processus de fusion, peut être appliquées à tous les documents disponibles.

– définition d'une structure d'annotation complexe, qui est codifiée en XML et qui est susceptibe d'être intégrée dans le standard MPEG-7 ;

– intégration de nouvelles méthodes pour l'interface d'utilisation qui sera aussi guidée par les connaissances sur le domaine.

# Bibliographie

[AND00a] André E., *Natural Language in Multimedia/Multimodal Systems*, in Mitkov R. (ed.), Handbook of Computational Linguistics, Oxford, 2000.

[AND00b] André E., *The Generation of Multimedia Presentations,* in Handbook of Natural Language Processing, Marcel Dekker, 2000.

[APP99] Appelt D.E., *An Introduction to Information Extraction*, in AI Communications (vol. 12), 1999.

[ASSF02] Assfalg J, M. Bertini, C. Colombo, A. Del Bimbo, *Semantic Annotation of Sports Videos*, IEEE Multimedia, 9(2):52-60 April-June2002.

[CUN99] *Information Extraction: A user Guide*, Rapport de recherche CS-99-07, Department of Computer Science, University of Sheffield, May 1999.

[DAY01] Day N, *MPEG-7 Applications: Multimedia Search and Retrieval*, in Proceedings of the First International Workshop on Multimedia Annotation, MMA-2001, 2001.

[DEC00] Declerck T., Neumann G., *Using a parameterisable and domain-adaptive information extraction system for annotating large-scale corpora?* , in Proceedings of the Workshop Information Extraction meets Corpus Linguistics, LREC-2000, 2000.

[DEC01] Declerck T., Wittenburg P., Cunningham H., *The Automatic Generation of Formal Annotations in a Multimedia Indexing and Searching Environment*, in Proceedings of the Workshop on Human Language Technology and Knowledge Management, ACL-2001, 2001.

[DJO98] Djoerd H., de Jong F., Netter K. (Eds), *14th Twente Workshop on Language Technology, Language Technology in Multimedia Information Retrieval*, TWLT 14, Enschede, Universiteit Twente, 1998.

[GRI96] Grishman R., Sundheim B., *Message Understanding Conference -- 6: A Brief History*, In Proceedings of the 16th International Conference on Computational Linguistics, COLING-96, 1996.

[HUM98] Humphreys K., Gaizauskas R., Azzam S., Huyck C., Mitchell B., Cunningham H., Wilks Y., *University of Sheffield: Description of the LaSIE-II System as used for MUC-7*, in *Seventh Message Understanding Conference (MUC-7)*, http://www.muc.saic.com/, SAIC Information Extraction, 1998.

[JOH98] Johnston M., *Unification-based Multimodal Parsing*, in Proceedings of the 17th International Conference on Computational Linguistics, COLING-98}, 1998.

[JON98] de Jong F., Netter K., *Olive: Speech-Based Video Retrieval*, in Hiemstra D., de Jong F., Netter K. (Eds), *Language Technology in Multimedia Information Retrieval* (Proceedings of the 14th Twente Workshop on Language Technology, TWLT 14), Enschede, Universiteit Twente, 1998.

[JON00] de Jong F., Gauvin J., Hiemstra D., Netter K., *Language-Based Multimedia Information Retrieval*, in Proceedings of the 6th Conference on Recherche d'Information Assistee par Ordinateur, RIAO-2000, 2000.

[KRI94] Krieger H.-U., Schaefer U., *TDL -- a type description language for constraint-based grammars*, in Proceedings of the 15th International Conference on Computational Linguistics, COLING-94, 1994.

[MAY99] Maybury M., *Multimedia Interaction for the New Millenium*, in Proceedings of Eurospeech 99, 1999.

[MCK85] McKeown K., *Text generation*, Cambridge University Press, 1985.

[MCN00] McNaugh J. (Ed), Information Extraction meets Corpus Linguistics, LREC-2000, 2000.

[MER97] Merlino A., Morey D., Maybury M., *Broadcast News Navigation using Story Segments,* ACM International Multimedia Conference, 1997.

[MOO89] Moore J., Paris C., *Planning Text for Advisory Dialogues*, in Proceedings of the 27th ACL, Vancouver, 1989.

[MUC95] Sixth Message Understanding Conference (MUC-6), Morgan Kaufmann, 1995.

[MUC98] *Seventh Message Understanding Conference (MUC-7)*, http://www.muc.saic.com/, SAIC Information Extraction, 1998.

[NEU97] Neumann G., Backofen R., Baur J., Becker M., Braun C., *An Information Extraction Core System for Real World German Text Processing*, in Proceedings of the 5th Conference on Applied Natural Language Processing, ANLP-97, 1997.

[NEU00] Neumann G., Braun C., Piskorski J., *A Divide-and-Conquer Strategy for Shallow Parsing of German Free Texts*, in Proceedings of the 6th Conference on Applied Natural Language Processing, ANLP-00, 2000.

[PISK00] Piskorski J., Neumann G., *An Intelligent Text Extraction and Navigation System*, in Proceedings of the 6th Conference on Recherche d'Information Assistee par Ordinateur, RIAO-2000, 2000.

[SAL98] Salway A., *Talking Pictures: Indexing and Representing Video with Collateral Texts*, in Hiemstra D., de Jong F., Netter K. (Eds), *Language Technology in Multimedia Information Retrieval* (Proceedings of the 14th Twente Workshop on Language Technology, TWLT 14), Enschede, Universiteit Twente, 1998.

# Index