

**UNE MÉTHODE GUIDÉE PAR LA SÉMANTIQUE POUR LA COMPRÉHENSION  
AUTOMATIQUE DES ÉNONCÉS ORAUX ARABES**

---

**Lamia Hadrich Belguith,**

Maître Assistant Habilité en Informatique

[l.belguith@fsegs.rnu.tn](mailto:l.belguith@fsegs.rnu.tn) , +216 74 278 777

**Younès Bahou,**

Doctorant en Informatique

[bahou\\_younes@yahoo.fr](mailto:bahou_younes@yahoo.fr) , +216 74 278 777

**Abdelmajid Ben Hamadou,**

Professeur en Informatique

[abdelmajid.benhamadou@isimsf.rnu.tn](mailto:abdelmajid.benhamadou@isimsf.rnu.tn) , +216 74 278 777

**Adresse professionnelle**

Laboratoire LARIS – MIRACL

Faculté des Sciences Economiques et de Gestion de Sfax

B.P. 1088, 3018 - Sfax – TUNISIE

**Résumé** : La compréhension automatique de la parole vise à extraire le sens utile des énoncés oraux. Dans ce papier, nous proposons une méthode originale pour la compréhension automatique de la parole arabe. La méthode proposée est fortement guidée par la sémantique. Elle a l'avantage d'être robuste face aux disfluences comme les hésitations, les répétitions et les autocorrections. Aussi, nous présentons le système SARF, un serveur vocal arabe interactif pour des renseignements sur le transport ferroviaire tunisien. Le module de la compréhension de la parole de SARF est basé sur la méthode proposée. Les résultats de l'évaluation de ce module sont très encourageants. En effet, le taux obtenu de *F-Measure* est de 71.79%.

**Mots clés** : Compréhension de la parole, analyse sémantique, disfluences, traitement de la parole arabe, dialogue Homme-machine, serveur vocal arabe interactif.

**Summary** : The automatic speech understanding aims to extract the useful meaning of the oral utterances. In this paper, we propose an original method for automatic Arabic speech understanding. The proposed method is strongly guided by the semantics. It has the advantage of being robust when coping with the disfluencies such as the hesitations, the repetitions and self-corrections. We also present, the SARF system, an interactive Arabic vocal server for Tunisian railway information. The speech-understanding module of SARF is based on the proposed method. The evaluation results of this module are very encouraging. Indeed, the obtained rate of F-Measure is 71.79%.

**Keywords**: Speech understanding, semantic analysis, disfluencies, Arabic speech processing, Human-machine dialogue, Arabic vocal server.

# Une méthode guidée par la sémantique pour la compréhension automatique des énoncés oraux arabes

La compréhension automatique de la parole vise à extraire le sens utile des énoncés oraux. Dans ce papier, nous proposons une méthode originale pour la compréhension automatique de la parole arabe. La méthode proposée est fortement guidée par la sémantique. Elle a l'avantage d'être robuste face aux disfluences comme les hésitations, les répétitions et les autocorrections. Aussi, nous présentons le système SARF, un serveur vocal arabe interactif pour des renseignements sur le transport ferroviaire tunisien. Le module de la compréhension de la parole de SARF est basé sur la méthode proposée. Les résultats de l'évaluation de ce module sont très encourageants. En effet, le taux obtenu de *F-Measure* est de 71.79%.

## 1 – INTRODUCTION

La compréhension automatique de la parole est une étape importante dans les Systèmes de Dialogue Oral Homme-Machine (SDOHM). Elle consiste à extraire le sens des énoncés oraux qui sont par nature incertains et ambigus. D'où, la nécessité d'un certain niveau de robustesse au niveau de l'analyse des énoncés pour surmonter les problèmes dus à la spontanéité de l'interaction orale. Plusieurs approches d'analyse ont été proposées. Parmi ces approches, nous citons l'approche guidée par la sémantique qui limite la compréhension à la recherche de séquences clés (*i.e.*, îlots ou segments conceptuels) dans l'énoncé. Le principal avantage de cette approche est qu'elle permet de faire ressortir facilement l'information, dite utile, d'un énoncé tout en ignorant de nombreux problèmes d'agrammaticalité et de spontanéité du langage parlé.

Nous proposons, dans ce papier, une méthode d'analyse des énoncés oraux transcrits (*i.e.*, séquences de mots reconnus par le système de la reconnaissance vocale), guidée par la sémantique et permettant de prendre en considération les problèmes de la spontanéité du langage parlé tels que l'autocorrection, l'hésitation et les répétitions.

Notre méthode se base sur le formalisme *frame grammar* (Bruce, 1975), inspiré de la grammaire de cas (Fillmore, 1968) pour la représentation sémantique des énoncés oraux et repose sur trois principales étapes à savoir, le pré-traitement de l'énoncé transcrit, l'analyse sémantique et la génération du ou des schéma(s) correspondant(s). Cette méthode permet de résoudre certains problèmes de disfluences, en particulier les hésitations, les répétitions et les autocorrections.

Comme application, nous avons choisi le cas d'un serveur vocal interactif pour le service de renseignement de la Société Nationale des Chemins de Fer Tunisiens (SNCFT). L'objectif de ce serveur vocal est de permettre à l'utilisateur de communiquer avec la machine, via la parole arabe, pour obtenir des renseignements sur le transport ferroviaire (*e.g.*, horaire d'un train, tarification, etc.).

Signalons, qu'à notre connaissance, il n'existe aucun serveur vocal capable de dialoguer en arabe, que ce soit dans le domaine du transport ferroviaire ou dans n'importe quel autre domaine. Cela est principalement dû au manque d'outils d'aide à l'élaboration de tels systèmes. En effet, les outils existants restent à l'échelle de prototypes de laboratoire. Nous citons à titre d'exemple l'outil de Dahmani et al. (2004) pour la reconnaissance automatique de mots enchaînés arabes, l'outil de Satori et al. (2007) pour la reconnaissance des chiffres arabes et l'outil de Ramsay et Mansour (2008) pour la synthèse de la parole arabe.

## 2 – BREF APERÇU DES APPROCHES DE COMPRÉHENSION AUTOMATIQUE DE LA PAROLE

La compréhension automatique de la parole cherche à attribuer une représentation sémantique aux énoncés prononcés. Elle doit faire face à plusieurs problèmes tels que les redondances d'informations, les répétitions, les autocorrections, les hésitations, les faux-départs, les contradictions, voire même les tendances à interrompre l'interlocuteur. Afin de surmonter ces problèmes, différentes

approches ont été proposées, dans la littérature, et testées dans divers contextes applicatifs. Nous distinguons principalement trois approches : l'approche guidée par la syntaxe, l'approche guidée par la sémantique et l'approche mixte (Bahou et *al.*, 2008a).

### 2.1 – Approche guidée par la syntaxe

L'approche guidée par la syntaxe est qualifiée d'approche par exploitation d'îlots syntaxiques (Goulian, 2002). Dans cette approche, une analyse syntaxique robuste et partielle est envisagée. En effet, en analyse automatique de la parole, il est apparu qu'une analyse syntaxique complète, cherchant à rendre compte des relations entre tous les mots de l'énoncé, pouvait faire décroître les performances en présence des disfluences, des constructions spécifiques à l'oral, des erreurs de reconnaissance et des phénomènes dus à la spontanéité de l'interaction (Bove, 2008).

Plusieurs systèmes ont été conçus en utilisant cette approche. Nous citons, à titre d'exemples, le système ROMUS (Goulian et *al.*, 2003) et le système LOGUS (Villaneau et *al.*, 2004). Il s'agit de deux systèmes de compréhension de la parole en français, appliqués au domaine de renseignement touristique et reposant sur des architectures générales comparables (*i.e.*, segmentation en *chunks* puis analyse de dépendances globales). Le système LOGUS utilise une stratégie d'analyse incrémentale reposant sur une analyse partielle par segments. Ce système utilise le *Lambda-calcul* pour la représentation sémantique de l'énoncé oral. Le système ROMUS repose sur une stratégie d'analyse incrémentale mixant des techniques de segmentation robuste et une analyse de dépendances sémantiques par grammaires de liens, pour la représentation sémantique de l'énoncé oral (Antoine, 2003).

### 2.2 – Approche guidée par la sémantique

L'approche guidée par la sémantique est qualifiée de sélective. Dans cette approche, la compréhension est limitée à la recherche du sens, dit utile, de l'énoncé (Luzzati, 1989). Elle repose sur l'identification de séquences clés, appelées segments conceptuels, à partir desquelles une structure sémantique prédéfinie sera instanciée. Donc, l'idée consiste à analyser uniquement les parties jugées pertinentes de l'énoncé. Nous présentons, dans ce qui suit, quelques systèmes qui reposent sur cette approche.

Le système MASK (Minker, 1999) traite des demandes d'information, en français, sur les trains ; notamment les horaires, les services à bord et les prix. Son analyseur sémantique est fondé sur le modèle de Markov caché. La représentation sémantique de l'énoncé est sous forme de schémas. L'analyse sémantique s'effectue en deux étapes : une étape d'apprentissage qui se base sur les séquences de mots (transcrites et prétraitées) et les séquences sémantiques correspondantes, et une étape de décodage/test qui fournit la séquence sémantique la plus probable. Préalablement à ces deux étapes, un pré-traitement des énoncés effectue la normalisation, le regroupement des catégories, et l'analyse lexicale (Minker et Bennacef, 2004).

Le système de Jamoussi (2004) offre aux touristes des informations touristiques diverses en langue française. L'analyseur sémantique utilise les réseaux de neurones qui sont connus par leur capacité de généralisation et par leur flexibilité. La représentation sémantique de l'énoncé est exprimée sous forme d'un ensemble de concepts. Le principe de la compréhension se résume en l'utilisation d'un perceptron multicouche (*i.e.*, un réseau de neurones artificiels) comme moyen d'association entre le langage d'entrée (*i.e.*, séquence de mots) et celui de sortie (*i.e.*, ensemble de concepts). Donc, l'idée est d'associer à chaque mot de l'énoncé un entier unique qui le symbolise et il en est de même pour les concepts. L'apprentissage se fait en présentant au réseau un ensemble de données constitué de paires de vecteurs d'entrées et de sorties désirées. Notons que l'analyse sémantique de l'énoncé transcrit en langage naturel (l'extraction des concepts) se fait de manière purement automatique en utilisant un réseau bayésien pour la classification non supervisée.

Pour la compréhension automatique de la parole arabe, nous citons le décodeur sémantique de la parole arabe spontanée de Zouaghi et *al.* (2008) appliqué au domaine de renseignement ferroviaire. Ce décodeur sémantique utilise une grammaire probabiliste qui permet de tenir compte de plusieurs informations contextuelles en même temps. Le décodage sémantique des énoncés est basé sur une analyse sémantique et ne considère que les éléments significatifs pour l'application. Chaque mot significatif pour l'application est

représenté à partir d'un ensemble de traits sémantiques noté  $Tse = \{domaine, classe\ sémantique, trait\ micro-sémantique\}$  et un ensemble de traits syntaxiques noté  $Tsy = \{genre, nombre, nature\}$ . L'analyse repose sur un modèle de langage probabiliste qui contribue à la sélection des Tse à affecter aux mots de l'énoncé, et sur un lexique sémantique qui décrit le sens de chaque mot via un ensemble de Tse et un ensemble de Tsy (Zouaghi et al., 2005).

Notons que l'approche sélective s'avère efficace pour des domaines très limités où l'ambiguïté sémantique est réduite. En plus, elle présente une certaine robustesse face aux difficultés causées par la spontanéité des énoncés oraux, l'agrammaticalité du langage parlé et la présence de mots non reconnus. Cette robustesse est obtenue par la nature sélective de cette approche. En effet, l'énoncé n'est pas analysé dans sa globalité mais concerne uniquement les parties porteuses d'informations relatives au domaine d'application.

Par ailleurs, les systèmes de compréhension basés sur l'approche sélective et utilisant le formalisme *frame grammar*, ont prouvé leur performance dans plusieurs domaines. La principale tâche pour l'élaboration de tels systèmes réside dans la définition des concepts et des mots de référence. Il s'agit d'une tâche importante, difficile mais réalisable pour des domaines restreints. De plus, il est possible en portant l'analyseur dans une autre application (mais d'un domaine assez proche), de conserver de nombreux concepts et mots de référence (Goulian, 2002).

### 2.3 – Approche mixte

Dans l'approche mixte, une analyse syntaxique complète est tout d'abord envisagée. En cas d'échec de cette analyse, des techniques sémantiques sont utilisées.

Le système TINA (Seneff, 1992) est basé sur l'approche mixte. Il offre un service de renseignement, en langue anglaise, sur le transport aérien aux États-Unis. L'analyseur sémantique de TINA se base sur une grammaire hors contexte contrainte. Le principe fondamental est que l'analyse, pour être efficace, doit associer syntaxe et sémantique. Il s'agit d'une analyse descendante basée sur des règles syntaxiques auxquelles s'ajoutent des contraintes

sémantiques. Les règles, qui sont celles d'une grammaire hors contexte, sont fabriquées directement à partir de l'analyse d'énoncés types. L'énoncé est analysé complètement à partir de sa forme syntaxique (Seneff, 1992). La grammaire est transformée en un automate probabiliste qui permet d'avantager les constructions les plus fréquentes. Devant l'insuffisance de cette analyse complète sur l'oral spontané, une stratégie d'analyse robuste par *chart* étendu a été intégrée. Cette analyse permet le recouvrement des groupes analysés correctement dans l'énoncé, en cas d'échec de l'analyse complète. Le résultat rendu est un arbre. Si plusieurs solutions sont possibles, la solution qui a consommé plus de mots est retenue. Les nœuds de l'arbre correspondent à des catégories qui peuvent être syntaxiques ou sémantiques.

## 3 – DIFFICULTÉS DE LA COMPRÉHENSION AUTOMATIQUE DE L'ARABE ORAL

L'objectif d'un système de compréhension de la parole est principalement la construction de la représentation sémantique de l'énoncé prononcé par l'utilisateur. Cette représentation est calculée en fonction de la séquence des mots reconnus, du *cotexte* (i.e., l'historique du dialogue) et du *contexte* (i.e., environnement, deixis, situations du dialogue, etc.) (Bousquet-Vernhettes, 2002).

Dans cette section, nous commençons, d'abord, par présenter les irrégularités langagières de l'oral, qui sont partagées par toutes les langues indépendamment de leurs familles ou origines. Ensuite, nous exposons les spécificités de l'arabe oral, notamment en ce qui concerne la diversité dialectale.

### 3.1 – Irrégularités langagières de l'oral

De nombreuses différences existent entre le langage écrit et le langage oral. L'auteur peut réfléchir à la formulation de sa phrase, la modifier jusqu'à complète satisfaction. De même, le lecteur peut relire une phrase en cas d'incompréhension ou de doute. En revanche, les erreurs de la parole sont corrigées en temps réel, d'où les phénomènes d'hésitation, de répétition, d'autocorrection, etc. (Minker et Bennacef, 2004).

Les caractéristiques intrinsèques au dialogue oral qui influent sur la compréhension automatique de la parole sont principalement :

- La spontanéité des énoncés qui peuvent contenir des redondances d'informations, des répétitions, des autocorrections, des hésitations, des faux-départs, des contradictions, voire même des tendances à interrompre l'interlocuteur.
- La structure agrammaticale des énoncés qui est liée non seulement à la spontanéité de l'interaction orale mais aussi au langage parlé lui-même.

Parmi les irrégularités langagières de l'oral (connues sous le nom de disfluences (Bove, 2008)), nous citons :

- Les répétitions : sont des suites de mots identiques et de même catégorie syntaxique. Une répétition peut être partielle (partie d'un syntagme) ou totale (syntagme complet). Les mots soulignés dans l'exemple (1) illustrent une répétition partielle.

(1) متى ينطلق القطار السريع آه إلى محطة تونس آه تونس  
[mtY ynTlq AlqTAr AlsrY E |h IY mHTp twns |h twns]

(Quand part le train rapide euh vers la gare de Tunis euh Tunis)

Notons que la répétition n'est pas toujours une redondance. Elle peut aussi avoir une fonction communicative. Par exemple, lorsqu'un locuteur n'est pas sûr que son message sera clairement perçu par son auditeur, il le répète. Par ailleurs, la répétition est un moyen pragmatique assez fréquent pour marquer une affirmation ou une insistance comme c'est le cas dans l'énoncé (1) où la répétition du mot تونس [twns] (Tunis) a une fonction d'affirmation.

- Les omissions : sont occasionnées par l'absence d'un ou de plusieurs constituants syntaxiques. Dans l'énoncé (2) les constituants omis sont mis entre accolades.

(2) ما هو ثمن {تذكرة القطار} لتونس  
[mA hw vmn {t\*krp AlqTAr} ltwns]  
(Quel est le prix {du billet de train} pour Tunis)

- Les autocorrections (dites aussi faux départs) : sont des corrections effectuées par le locuteur lui-même pour corriger son énoncé. Il s'agit, alors, soit d'une correction d'un simple mot par un autre mot (ou groupement de mots) comme illustré par l'énoncé (3) ; soit d'une correction de tout un segment par un autre segment comme c'est le cas de l'énoncé (4).

(3) أريد ثمن تذكرة ذهاب آه لا ذهاب وإياب من صفاقس إلى تونس

[Oryd vmn t\*krp \*hAb |h lA \*hAb wly~Ab mn Sfaqs Ily twns]

(Je voudrais connaître le prix d'un billet aller euh non aller-retour de Sfax vers Tunis)

(4) ما هو ثمن السفر لسوسة آه لا عفوا وقت السفر لسوسة  
[mA hw vmn Alsfr lswsp |h lA EfwA wqt Alsfr lswsp]

(Quel est le prix du voyage à Sousse euh non pardon l'horaire du voyage à Sousse)

À la lumière de ces deux exemples, on s'aperçoit que l'autocorrection peut s'étendre sur plusieurs mots et occuper une grande place dans les transcriptions.

- Les reprises : sont des interruptions suivies par un nouveau syntagme. Il y a alors abandon d'un segment et début d'un nouveau segment.

L'énoncé (5) comporte une reprise qui commence à la position indiquée par ↑.

(5) لا أعرف اسم آخر محطة للقطار الذي ينطلق على الساعة العاشرة ↑ باختصار هل القطار ذاهب إلى تونس

[lA OErf Asm |xr mHTp llqTAr Al\*y ynTlq Ely AlsAEp AlEA\$rp ↑ bAxtSAr hl AlqTAr \*Ahb Ily twns]

(Je ne connais pas le nom de la gare terminus du train qui part à dix heures ↑ bref est ce que ce train part à Tunis)

- Les amorces : sont des mots commencés et inachevés. Le résultat est alors des fragments de mots qui, selon le cas, peuvent être abandonnés ou repris et achevés par le locuteur comme c'est le cas de l'énoncé (6).

(6) ما هو ثمن تذكرة القطار العادي  
 [mʌ hw vmn t\*krp AlqTAr AIE |h  
 AIEAdi~]  
 (Quel est le prix du billet du train nor euh  
normal)

### 3.2 – Diversité dialectale de l’arabe parlé

La langue arabe, rappelons-le, appartient à la famille des langues sémitiques comme l’akkadien, l’hébreu, l’araméen et le sud arabe. Elle peut être considérée comme un terme générique regroupant principalement deux variétés (Bahou et al., 2008b) :

- L’arabe littéraire (standard) : langue employée dans les journaux et dans la plupart des écrits administratifs, médiatiques, scientifiques, techniques et littéraires. Elle est parlée, dans tous les pays arabes, ordinairement à la radio, dans la majorité des journaux télévisés, dans les cours et conférences universitaires et scientifiques, dans les discours officiels, etc. Il s’agit de la variété retenue comme langue officielle dans tous les pays arabes, et comme langue commune entre eux. Il devra être également admis qu’il s’agit de l’arabe standard moderne.
- Les arabes dialectaux : langues parlées au quotidien dans les pays arabes. Notons que ces dialectes varient d’un pays à un autre, voire même d’une région à une autre dans un même pays. Les dialectes arabes sont parfois tellement différents qu’il est difficile de les considérer comme une seule langue. Issus de l’arabe classique, leurs systèmes grammaticaux respectifs affichent de nettes divergences avec celui de l’arabe classique. La différenciation porte sur la prosodie, la phonologie, la morphologie et la syntaxe. Sur un plan très général, les dialectes se différencient de l’arabe classique par des traits qui leur sont communs. Par exemple, l’abandon des désinences casuelles et des flexions modales dans le verbe. Le système phonétique est en revanche plus riche. En ce qui concerne le vocabulaire, le fond lexical arabe est souvent enrichi

par des emprunts à des langues autochtones<sup>1</sup>.

Au delà de cette diversité dialectale, les sociétés arabes ont une conscience aiguë d’appartenir à une communauté linguistique homogène. Elles sont attachées à l’intégrité de leur langue, d’où l’importance de l’arabe standard qui constitue la langue commune pour cette large population. Notons que dans le cadre du présent travail, nous nous intéressons à l’arabe standard pour les raisons suivantes (Bahou et al., 2008c) :

- L’arabe standard est compréhensible et utilisé dans tous les pays arabes.
- Il est difficile pour un système de compréhension automatique de traiter différents dialectes à la fois.
- Si les outils d’analyse de l’arabe standard sont peu fréquents, ils sont quasi-absents pour les dialectes arabes.

Outre les variétés de la langue arabe parlée, signalons que certains phénomènes de l’arabe écrit sont aussi partagés par l’arabe parlé. Parmi ces phénomènes, nous citons l’agglutination des mots qui rend la segmentation de l’énoncé, en mots, difficile, et l’irrégularité de l’ordre des mots. Si ce deuxième phénomène pose des problèmes au niveau du traitement automatique de l’arabe écrit, il semble s’accroître, voire même se multiplier, pour le cas de l’arabe parlé où l’ordre des mots est totalement variable. Cela complique la tâche de construction du modèle de langage pour la compréhension, qui sert à analyser et interpréter les énoncés oraux de l’interlocuteur.

Toutes ces difficultés nous amènent à proposer une méthode de compréhension plus sémantique que syntaxique qui permet de décrire non pas la structure de l’énoncé mais son sens.

## 4 – NOTRE MÉTHODE DE COMPRÉHENSION

Afin d’appréhender les problèmes de compréhension des énoncés arabes oraux reconnus par un système de reconnaissance

<sup>1</sup> Langue qui existe depuis plusieurs générations, mais qui, sur un territoire donné, n’est plus parlée que par un groupe restreint de locuteurs, généralement âgés, et qui n’est souvent plus la langue maternelle de la nouvelle génération.

automatique de la parole, nous avons proposé une méthode de compréhension fortement guidée par la sémantique, qui entre dans le cadre de l'approche sélective. Le choix d'opter pour une approche guidée plutôt par la sémantique que la syntaxe a été dicté d'une part, par le fait que les approches non fondées sur une analyse syntaxique fine de l'énoncé sont généralement reconnues pour être plus robustes face aux difficultés de l'oral spontané et d'autre part, par les spécificités de la langue arabe, que nous avons décrites dans la section précédente (*e.g.*, l'ordre des mots est totalement variable). En effet, ces spécificités compliquent davantage la tâche de construction du modèle de langage pour la compréhension, si l'on se base sur une approche guidée par la syntaxe.

L'idée principale de notre méthode consiste à orienter l'analyse vers les noyaux syntaxiques locaux de l'énoncé. Ainsi, la syntaxe joue un rôle secondaire dans le processus de compréhension et ne nécessite pas des connaissances linguistiques très approfondies. Notre méthode se caractérise aussi par le traitement des disfluences, en particulier les hésitations, les répétitions et les autocorrections.

Par ailleurs, nous avons opté pour le choix du formalisme de *frame grammar* (Bruce, 1975) qui permet de construire, à partir d'un énoncé, une représentation sémantique sous forme d'un ou de plusieurs schémas. Ce formalisme est l'un des rares formalismes qui permettent le traitement de phrases agrammaticales. De plus, il offre un modèle de la structure profonde d'un énoncé, où la sémantique joue un rôle essentiel, sans exclure les contraintes syntaxiques. Ainsi, ce formalisme peut être vu comme un ensemble de règles syntaxico-sémantiques plus ou moins contraignantes (Minker et Bennacef, 2004).

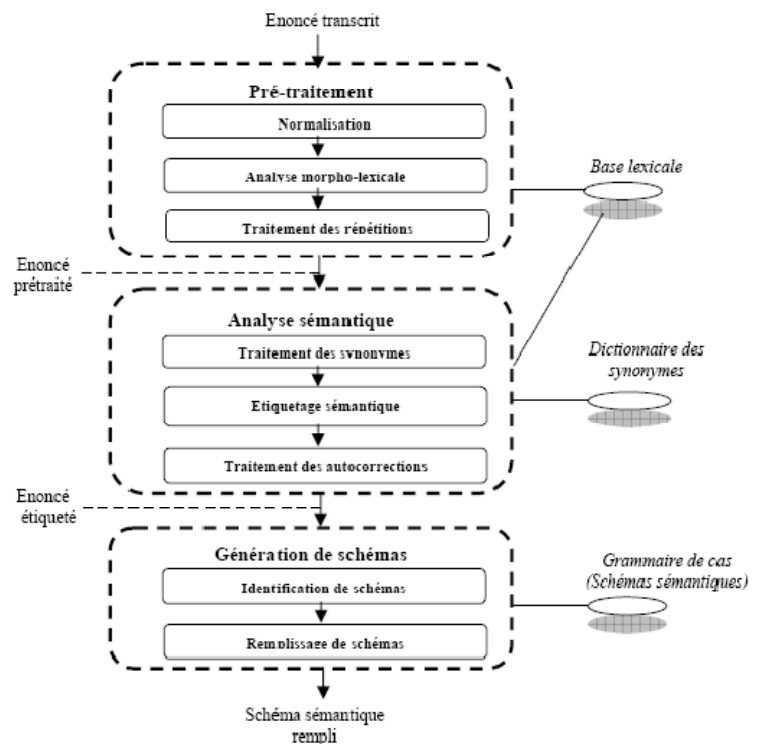
Rappelons que la grammaire de cas a été introduite à l'origine par Fillmore (1968) pour exprimer les relations entre un verbe et ses compléments. Cette théorie a été étendue par la suite par Bruce (1975) pour un système fondé sur les concepts. Ce formalisme construit une représentation sémantique d'un énoncé sous la forme d'un schéma. Le schéma correspond à un concept et comporte plusieurs attributs (voir figure 1). Un ou plusieurs mots de références permettent d'identifier le concept ainsi que le

schéma correspondant à ce concept. Les attributs du schéma sont instanciés avec les cas sémantiques identifiés, en se basant sur des marqueurs de cas.

**Schéma sémantique : *Tarif\_Voyage***  
Mots de référence : {ثمن، ثمن+سفر}  
Ville\_Départ : {من} [قابس، صفاقس، سوسة...]  
Ville\_Arrivée : {الى} [قابس، صفاقس، سوسة...]  
Type\_Billet : {ذهاب، ذهاب وإياب}  
Classe : {أولى، رفاهة}  
Type\_Train : {عادي، سريع}  
Jour\_Départ : {اثنين، ثلاثاء...1، 2...}  
Mois\_Départ : {جانفي، فيفري...}  
Heure\_Départ : {ساعة} [1، 2...]  
Minute\_Départ : [دقيقة} {1، 2...}

**Figure 1 :** Schéma sémantique *Tarif\_Voyage* simplifié

Notre méthode de compréhension repose sur trois principales étapes (Bahou et al., 2008c) : une étape de pré-traitement de l'énoncé transcrit, une étape d'analyse sémantique de l'énoncé prétraité et une étape de génération du schéma sémantique correspondant (voir figure 2).



**Figure 2 :** Étapes de la méthode de compréhension

#### 4.1 – Pré-traitement de l'énoncé

Un énoncé oral est par nature rigide et difficilement maîtrisable. Cela est dû, principalement, à la nature spontanée de

l'énoncé qui fait que ce dernier renferme divers types de disfluences (*i.e.*, répétitions, hésitations, autocorrections, reprises, etc.) ; phénomènes tout à fait normaux et habituels de la parole spontanée.

Ces phénomènes induisent des ambiguïtés qui peuvent produire des erreurs d'analyse. Une étape de pré-traitement s'avère donc nécessaire pour faciliter le traitement, de l'énoncé transcrit, par les étapes en aval. Cette étape consiste à supprimer les répétitions et les informations inutiles, à convertir les nombres écrits en toutes lettres, en chiffres, et à déterminer les formes canoniques<sup>2</sup> des mots. Pour se faire, l'énoncé reconnu subit une normalisation, une analyse morpho-lexicale et un traitement des répétitions (Bahou et *al.*, 2008b).

#### 4.1.1 – Normalisation de l'énoncé

L'objet de la normalisation de l'énoncé transcrit est de rendre ce dernier plus convenable pour les analyses ultérieures. Il s'agit de convertir les unités (mots ou suites de mots) représentant des dates, des horaires, des prix, des nombres, etc. sous forme de chiffres. La détection de ces unités se base essentiellement sur des marqueurs (*i.e.*, pré-marqueurs et post-marqueurs) de cas sémantiques relatifs aux dates de départ et d'arrivée, horaires de départ et d'arrivée, prix de billets, nombre de places à réserver, etc.

Ainsi par exemple, la détection d'un pré-marqueur d'horaire (*e.g.*, نحو+الساعة [*nHw*]+[*sAEp*] (vers+1'heure)) et d'un post-marqueur d'horaire (*e.g.*, دقيقة [*dqyqp*] (minute)) permet de localiser les unités qui représentent un horaire (heures+minutes) et qui sont délimitées par ces deux marqueurs. Ces unités seront, alors, converties en chiffres. Cette conversion est nécessaire, notamment pour permettre l'interrogation de la base de données en vue d'extraire les informations demandées par le locuteur (Bahou et *al.*, 2008a).

De plus, la normalisation permet de regrouper certaines locutions en un seul mot, ce qui simplifie la topologie du modèle de langage. Par exemple, les deux mots نحو [*nHw*] (vers) et

الساعة [*sAEp*] (l'heure) constituent une seule entité lexicale. Ce traitement a pour objectif de considérer ces deux mots comme une seule locution qui aura une seule étiquette sémantique (*i.e.*, *Marq\_Heure*).

#### 4.1.2 – Analyse morpho-lexicale

L'analyse morpho-lexicale de l'énoncé permet de déterminer pour chaque mot de l'énoncé sa forme canonique avant de vérifier son appartenance au lexique. En effet, les mots de référence, les marqueurs de cas des schémas sémantiques et les mots du lexique sont tous représentés sous leurs formes canoniques. Cela permet de réduire considérablement la taille des schémas ainsi que celle du lexique. Ainsi, à la place de stoker pour chaque entrée du lexique, le mot avec toutes ses formes dérivées possibles, seule la forme canonique du mot en question est représentée (Bahou et *al.*, 2008b).

De plus, à ce niveau d'analyse de l'énoncé, certains mots sont supprimés. Ces derniers correspondent à des hésitations et des extralinguistiques.

Les extralinguistiques sont généralement des mots non porteurs de sens pour l'application (*e.g.*, أريد [*Oryd*] (je voudrais), مرحبا [*mrHbA*] (bonjour/salut)). Ils sont recensés à partir de l'analyse des corpus. Ces mots sont supprimés car ils ne sont pas jugés utiles à l'interprétation sémantique de l'énoncé.

Nous soulignons cependant que toutes les hésitations ne sont pas forcément supprimées lors de cette étape. En effet, certains mots peuvent guider la résolution de certaines disfluences telles que les reprises et les autocorrections. Par conséquent, ce pré-traitement ne concerne que les mots dont il est certain qu'ils sont inutiles pour la compréhension.

Notons, par ailleurs, que les mots non reconnus au niveau de cette étape sont marqués par des étiquettes indiquant qu'il s'agit d'une disfluente. Ces mots représentent généralement des mots agrammaticaux résultant d'une amorce (*i.e.*, mots commencés par le locuteur mais inachevés) ou d'une mauvaise reconnaissance par le système de reconnaissance de la parole.

#### 4.1.3 – Traitement des répétitions

Par répétition, nous entendons la reprise d'un mot ou d'un ensemble de mots et pas celle

<sup>2</sup> La forme canonique représente les racines pour les verbes et généralement la forme du masculin non déterminé pour les mots et les adjectifs (Belguith Hadrich et Chaaben, 2006).



d'un segment différent avec le même sens (*i.e.*, la paraphrase). Les répétitions sont très fréquentes dans les énoncés oraux et peuvent être accompagnées par d'autres types de disfluences telles que les autocorrections et les reprises. On distingue les répétitions d'une unité simple (*i.e.*, un mot unique) et les répétitions d'unités complexes (*i.e.*, souvent des *chunks* entiers, des groupes de mots).

Ce traitement consiste à repérer les répétitions d'unités simples ou complexes et à les éliminer.

#### 4.1.4 – Exemple illustratif

Après pré-traitement, l'énoncé (7) est transformé en l'énoncé (8). Dans ce dernier, le mot أريد [Oryd] (je voudrais), qui n'appartient pas au lexique de mots utiles, est supprimé ; une occurrence du mot صفاقس [SfAqs] (Sfax) qui est répété deux fois, est aussi supprimée ; le groupement de mots représentant un horaire est transformé sous une forme numérique ; le mot الساعة [AlsAEp] est transformé sous sa forme canonique (*i.e.*, ساعة [sAEp] (heure)) ; les locutions ذهاب وإياب [\*hAb wLy~Ab] (aller retour) sont regroupées en une seule unité lexicale (*i.e.*, ذهاب-و-إياب (aller-retour) ; et de même, les locutions على ساعة [Ely sAEp] (à heure) et أه لا [h lA] (euh non) sont regroupées respectivement en على-ساعة (à-heure) et أه-لا (euh-non).

(7) أريد ثمن تذكرة ذهاب أه لا ذهاب وإياب من صفاقس أه صفاقس إلى تونس على الساعة الحادية عشر وثلاثة وثلاثون دقيقة

[Oryd vmn t\*krp \*hAb |h lA \*hAb wLy~Ab mn SfAqs |h SfAqs IY twns Ely AlsAEp AlHADyp E\$R wvLAvp wvLAvwn dgyqp]

(Je voudrais le prix d'un billet aller simple euh non aller retour de Sfax euh Sfax vers Tunis à onze heures trente trois minutes)

(8) ثمن تذكرة ذهاب أه-لا ذهاب-و-إياب من صفاقس إلى تونس على-ساعة 11 و33 دقيقة

[vmn t\*krp \*hAb |h lA \*hAb wLy~Ab mn SfAqs IY twns Ely AlsAEp 11 w 33 dgyqp]

(Prix billet aller simple euh-non aller-retour de Sfax vers Tunis à 11 heure 33 minutes)

## 4.2 – Analyse sémantique de l'énoncé

L'analyse sémantique consiste à attribuer des étiquettes sémantiques à tous les mots (ou groupements de mots) de l'énoncé prétraité. Ces mots représentent, en fait, des unités utiles, puisque les unités inutiles ont été supprimées ou marquées lors de l'étape de pré-traitement.

Le résultat de l'analyse sémantique de l'énoncé prétraité est alors une séquence de couples < étiquette sémantique, unité utile >. Afin de générer cette séquence, l'énoncé subit d'abord un traitement de la synonymie, ensuite un étiquetage sémantique et enfin un traitement des autocorrections (Bahou et al., 2008c).

### 4.2.1 – Traitement de la synonymie

Les mots considérés comme synonymes de mots appartenant au lexique, sont remplacés par les mots qui les représentent dans ce dernier. Par exemple, les mots سعر [sEr], كلفة [klfp], تعريف [tEryfp] et مبلغ [mblg] sont tous des synonymes du mot du lexique ثمن [vmn] (prix), qui les représente et qui se rapporte au schéma sémantique *Tarif Voyage*. Si l'un de ces mots figure dans l'énoncé, il sera alors remplacé par le mot ثمن [vmn] (prix).

Ce traitement permet de réduire le nombre de mots de référence des schémas sémantiques. Notons que lorsqu'il y a ambiguïtés (*i.e.*, cas d'un mot correspondant à deux mots de référence), toutes les solutions possibles sont représentées.

### 4.2.2 – Étiquetage sémantique

L'étiquetage sémantique permet d'attribuer une étiquette sémantique à chaque mot (ou groupement de mots), de l'énoncé, ayant un sens utile, ou représentant un marqueur de cas. Cet étiquetage se base sur un lexique des mots utiles étiquetés. À titre d'exemple, l'étiquette *Marq\_Ville\_Depart* est attribuée à la préposition من [mn] (de).

L'énoncé (9) est le résultat de l'étiquetage sémantique de l'énoncé (8).

(9) < Mot\_Ref\_Tarif, ثمن > < تذكرة, Marq\_Type\_Billet >  
 < ذهاب, Type\_Billet > < أه-لا, Marq\_Rectification >  
 < ذهاب-إياب, Type\_Billet > < من, Marq\_Ville\_Depart >  
 < صفاقس, Ville\_Depart > < إلى, Marq\_Ville\_Arrivée >  
 < تونس, Ville-Arrivée > < على-ساعة, Marq\_Heure >  
 < دقيقة, Marq\_Minute > < 33, Minute > < 11, Heure >

#### 4.2.2 – Traitement des autocorrections

Le traitement des autocorrections consiste à détecter les segments de l'énoncé qui expriment des autocorrections et à les corriger. Ainsi, si deux constituants successifs ont la même étiquette sémantique, le premier est considéré comme erroné et sera par conséquent éliminé, alors que le second sera retenu.

Par exemple, dans l'énoncé (10), les deux mots successifs سوسة [swsp] (Sousse) et صفاقس [SfAqs] (Sfax) ont la même étiquette sémantique *Ville\_Depart*. Il s'agit donc d'une autocorrection de la ville de départ. La première ville سوسة [swsp] (Sousse) est donc éliminée et la deuxième ville صفاقس [SfAqs] (Sfax) est retenue comme ville de départ.

(10) أريد معرفة سعر الرحلة من سوسة صفاقس نحو  
تونس في الخامس عشر ماي  
[Oryd mErfp sEr AlrHlp mn swsp SfAqs  
nHw twns fy AlxAms E\$R mAy]  
(Je voudrai connaître le prix du voyage  
de **Sousse Sfax** vers Tunis le quinze  
mai)

Par ailleurs, si dans un énoncé, il y a un marqueur de rectification (*i.e.*, marqueur d'autocorrection) alors le segment qui précède ce marqueur sera remplacé par celui qui le succède. C'est le cas par exemple de l'énoncé (8) qui présente une autocorrection du mot ذهاب [\*hAb] (aller simple) erroné, par le groupement de mots ذهاب-إياب [\*hAb-w-Iy~Ab] (aller-retour). Cette autocorrection est détectée par la présence du marqueur de rectification لا-أه [[h-lA] (euh-non). Après traitement de l'autocorrection, l'énoncé (8) sera transformé en l'énoncé (11).

(11) ثمن تذكرة ذهاب-إياب من صفاقس إلى تونس  
على-ساعة 11 و 33 دقيقة  
[vmn t\*krp \*hAb wIy~Ab mn SfAqs IY  
twns ELY sAEp 11 w 33 dgyqp]  
(Prix d'un billet aller-retour de Sfax  
vers Tunis à 11 heures 33 minutes)

Cependant, comme nous l'avons mentionné dans la section 3.1, les autocorrections peuvent s'étendre sur plusieurs mots et occuper une grande place dans les transcriptions. De plus, le locuteur peut corriger un mot (ou plusieurs mots) par un autre mot (ou groupe de mots) qui n'a pas forcément la même étiquette

sémantique que le mot erroné. C'est le cas de l'énoncé (4) présenté dans la section 3.1.

Par ailleurs, la présence des marqueurs d'autocorrection est certes importante pour la détection de l'autocorrection ; seulement ils ne sont pas toujours présents, auquel cas le recours à un dialogue avec l'utilisateur s'avère parfois nécessaire.

#### 4.3 – Génération des schémas sémantiques

L'étape de génération de schémas sémantiques consiste à identifier, dans un premier temps, le schéma sémantique qui correspond, au mieux, aux mots de référence et à remplir, dans un second temps, ce schéma en se basant sur les marqueurs de cas et les étiquettes sémantiques déterminées lors de l'analyse sémantique. En cas d'ambiguïtés, tous les schémas candidats sont générés.

##### 4.3.1 – Identification des schémas sémantiques

Rappelons qu'à chaque schéma sémantique correspond une liste exhaustive de mots de référence et qu'un énoncé peut contenir plusieurs mots de référence pouvant appartenir à des schémas différents.

L'identification des schémas se base principalement sur la liste des mots de référence. Il s'agit de calculer un score de similarité entre l'énoncé et chaque schéma sémantique. Ce score correspond au nombre de mots de référence communs entre l'énoncé et le schéma en question. Ainsi, le schéma ayant le score le plus élevé sera retenu. En cas de schémas ex-æquo, ils seront tous retenus. En cas d'absence de mots de référence (*i.e.*, cas d'un énoncé qui dépend du contexte, où les mots de référence ne figurent pas dans l'énoncé en question mais dans les énoncés antérieurs du dialogue), le schéma par défaut est retenu (Bahou et al., 2008a).

Pour l'énoncé (11), les deux schémas *Tarif\_Voyage* et *Réservation\_Billet* sont identifiés. En effet, le mot de référence ثمن [vmn] (prix) correspond au schéma *Tarif\_Voyage* alors que le mot تذكرة [t\*krp] (billet) correspond aux deux schémas *Tarif\_Voyage* et *Réservation\_Billet*. Seul le schéma *Tarif\_Voyage* est retenu vu qu'il possède le score le plus élevé (2 contre 1).

##### 4.3.2 – Remplissage des schémas sémantiques

Le remplissage du ou des schéma(s) identifié(s) se base principalement sur les

marqueurs et sur les étiquettes sémantiques des mots (ou groupements de mots) de l'énoncé. En effet, ce remplissage consiste à instancier la valeur de chaque paire « cas sémantique/valeur » du schéma, par des informations (unités porteuses de sens) contenues dans l'énoncé. Ces informations sont identifiées par leurs étiquettes sémantiques et les marqueurs de cas correspondants. La figure 3 représente le schéma *Tarif\_Voyage* instancié correspondant à l'énoncé (11).

**Schéma sémantique : Tarif\_Voyage**  
 Mot de référence : ثمن [vnm] (prix)  
 Ville\_Départ : صفاقس [SfAqs] (Sfax)  
 Ville\_Arrivée : تونس [twns] (Tunis)  
 Type\_Billet : ذهاب-و-إياب [\*hAb-w-Iy~Ab] (aller-retour)  
 Classe : \$  
 Jour\_Départ : \$  
 Mois-Départ : \$  
 Heure\_Départ : 11  
 Minute\_Départ : 33  
 Date\_Voyage : \$

**Figure 3** : Schéma sémantique *Tarif\_Voyage* simplifié correspondant à l'énoncé (11)

## 5 – PRÉSENTATION DU SYSTÈME SARF

Dans cette section, nous présentons le système SARF (Serveur vocal Arabe pour des Renseignements sur le transport Ferroviaire) basé sur la méthode de compréhension proposée. SARF est un serveur vocal interactif arabe offrant aux usagers un accès aux informations sur le transport ferroviaire tunisien (e.g., tarif, horaire, réservation, etc.) et ce en communiquant en arabe standard (Bahou et al., 2008c).

Nous présentons, dans ce qui suit, le corpus utilisé pour déterminer aussi bien les schémas sémantiques de la grammaire que le lexique relatif au domaine de la SNCFT.

### 5.1 – Corpus d'étude

Les corpus oraux représentent une proportion significative lors du développement d'un SDOHM. Ainsi, la compréhension automatique de la parole est étroitement liée à la disponibilité de ces corpus. Cependant, et comme nous l'avons signalé au début de ce papier, les ressources linguistiques arabes sont très rares, voire même indisponibles. C'est le cas des corpus oraux arabes. Ainsi, dans le cadre de ce travail, nous étions amenés à

construire notre propre corpus d'étude selon la technique du *Magicien d'Oz* (Dahlbäck et al., 1993).

Ainsi, nous avons utilisé des scénarios traitant des renseignements sur le transport ferroviaire tunisien. Toutes les requêtes du corpus (soit 1800 requêtes) ont été enregistrées, puis, transcrites manuellement selon des normes de transcription dans des fichiers XML, et étiquetées conformément aux standards proposés par la communauté ARPA (Minker et Bennacef, 2004).

La transcription manuelle a été faite d'une manière fidèle à ce qu'il a été enregistré. C'est-à-dire, les mots sont transcrits comme nous les entendons dans les enregistrements. D'où, la présence d'hésitations, d'autocorrections, de répétitions, d'amorces, de reprises, etc.

Nous distinguons trois types de requêtes à savoir, des requêtes indépendantes du contexte, des requêtes dépendantes du contexte et des requêtes aberrantes.

L'étude du corpus nous a permis d'identifier six concepts à savoir, le concept *Tarif\_Voyage*, le concept *Horaire\_Voyage*, le concept *Réservation\_Billet*, le concept *Durée\_Voyage*, le concept *Trajet\_Train* et le concept *Type\_Train* (Bahou et al., 2008b). À chaque concept, est associé un schéma contenant des mots de référence et des cas sémantiques relatifs à notre domaine applicatif. Ensuite, nous avons regroupé les mots, qui sont liés, sémantiquement dans des ensembles, puis nous avons attribué à chaque schéma les ensembles qui lui font référence. Ainsi, et pour réduire le nombre de mots de référence, nous n'avons gardé dans les schémas qu'un seul mot de référence de chaque ensemble caractérisant ce schéma. Aussi, à partir de ce corpus, nous avons pu construire le lexique relatif à notre domaine applicatif. Signalons que, seules les formes canoniques des mots sont représentées dans le lexique.

### 5.2 – Architecture du système SARF

Le système SARF se compose de cinq modules à savoir, le module de reconnaissance vocale, le module de compréhension, le module de gestion du dialogue, le module de génération de réponses et le module de synthèse vocale. La figure 4 illustre l'architecture générale du système SARF (Bahou et al., 2008a).

SARF intègre des techniques de reconnaissance, de compréhension du langage naturel, de gestion de dialogue et de synthèse de la parole. Ces techniques permettent d'extraire le contenu sémantique ou le sens d'un énoncé formulé par l'utilisateur afin de lui fournir des informations.

À partir du signal sonore émis par le locuteur, le module de reconnaissance de la parole délivre les listes de mots qui sont censés correspondre à l'énoncé source. Le module de compréhension fournit au gestionnaire de

dialogue une (ou plusieurs) représentation(s) sémantique(s) de l'énoncé transcrit. Ce module assure l'interface avec la base de données de l'application et propose des réponses ou éventuellement, des demandes d'informations complémentaires à l'utilisateur. Le module de génération des réponses assure la transformation des réponses, en phrases arabes compréhensibles par l'utilisateur. Le module de synthèse vocale permet la retransmission de ces phrases sous forme d'un signal sonore.

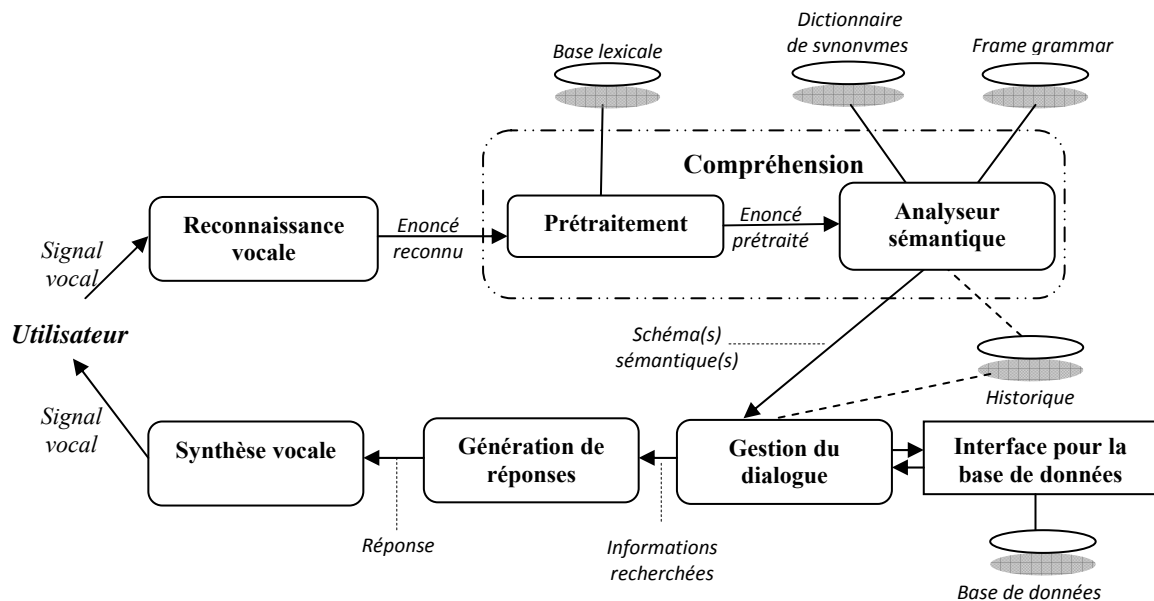


Figure 4 : Architecture générale du système SARF

Dans ce qui suit, nous nous intéressons au module de compréhension de la parole du système SARF. Notons que, pour le module de la reconnaissance vocale et le module de la synthèse vocale, nous comptons utiliser des systèmes commercialisés.

Le module de compréhension se compose de deux sous-modules : un module de pré-traitement de l'énoncé transcrit (*i.e.*, la normalisation, l'analyse morpho-lexicale et le traitement des répétitions) et un module d'analyse sémantique permettant l'analyse sémantique de l'énoncé prétraité (*i.e.*, le traitement de la synonymie, l'étiquetage sémantique et le traitement des autocorrections) et la génération des schémas sémantiques correspondant à cet énoncé. Les schémas sémantiques sont stockés dans un fichier XML avant d'être transmis au module de gestion de dialogue. Le module de compréhension est programmé avec le langage

de programmation JAVA sous l'environnement JBuilder 2007 (Bahou et al., 2008a).

### 5.3 – Évaluation du module de compréhension de SARF

Pour certaines langues comme le français, il existe actuellement des plateformes pour l'évaluation des modules de compréhension de systèmes de dialogue. Ces plateformes mettent notamment à disposition de la communauté un important corpus de dialogues réels annotés en sémantique. Ainsi, le très bon taux d'accord inter-annotateurs obtenu valide le choix du formalisme d'annotation. Certains consortiums ont également élaboré des outils d'évaluation spécifiques permettant de comparer les systèmes et de les évaluer dans le détail (Bonneau-Maynard et al., 2008).

La disponibilité de telles ressources (*i.e.*, corpus et outils d'évaluation) facilite les recherches dans le développement de systèmes

de dialogue robustes dans des situations réelles.

Néanmoins, ceci n'est pas le cas pour l'arabe où ces ressources sont quasi-absentes, à l'exception de quelques corpus diffusés par ELDA/ELRA<sup>3</sup>.

Ainsi, nous étions amenés à construire notre propre corpus d'évaluation selon la même technique du *Magicien d'Oz* utilisée pour la construction du corpus d'étude. Le corpus d'évaluation est constitué de 1003 requêtes (soit 12321 mots) de différents types, prononcées d'une façon spontanée et transcrites manuellement. Ces requêtes correspondent à des scénarios traitant des renseignements sur le transport ferroviaire tunisien.

L'évaluation du module de compréhension de SARF, sur le corpus d'évaluation, a montré que ce système génère 186 erreurs (soit en moyenne une erreur par 5 énoncés). Les mesures de rappel, de précision et de F-Mesure sont respectivement 73.00%, 70.62% et 71.79% et le temps moyen d'exécution d'un énoncé, de 12 mots, est de 0.279 secondes (Bahou *et al.*, 2008c).

Les cas d'échec s'expliquent principalement par la présence, dans l'énoncé, des disfluences qui résultent d'une amorce, d'autocorrection d'un segment par un autre segment sans que ce dernier ne soit introduit par un marqueur de rectification, ou de reprise sous forme d'abandon d'un segment et début d'un nouveau segment. Ces cas de disfluences ne sont pas encore pris en charge par le système SARF à son état actuel.

## 6 – LA GESTION DU DIALOGUE

En cas de présence de problèmes d'incohérence ou d'incomplétude au niveau du schéma sémantique généré ou aussi en cas de génération de plusieurs schémas, le recours à un dialogue avec l'utilisateur s'impose pour, d'une part, lever toutes sortes d'ambiguïtés inhérentes à l'énoncé lui-même et d'autre part, interpréter les schémas sémantiques et décider de la nature de l'intervention adéquate du système qui préserve l'aspect naturel du dialogue.

Les exemples présentés ci-dessous illustrent bien quelques uns des problèmes cités. L'énoncé (12) présente un cas d'incohérence entre la ville de départ et la ville d'arrivée. L'énoncé (13) illustre un cas d'omission de la ville de départ. Cette dernière n'a pas été précisée par l'utilisateur puisqu'elle représente, bien évidemment, la ville où il se trouve. L'énoncé (14) présente un cas d'ellipse. En effet, la ville de départ ne figure pas dans l'énoncé (14.c) vu qu'elle a déjà été donnée, par l'utilisateur, dans un énoncé précédent (*i.e.*, l'énoncé (14.a)) appartenant au même dialogue.

(12) أريد تذكرة ذهاب من سوسة إلى سوسة  
[Oryd t\*krp \*hAb mn swsp ILY swsp]  
(Je voudrai un billet aller simple de Sousse à Sousse)

(13) متى ينطلق القطار إلى تونس  
[mtY ynTlq AlqTAR ILY twns]  
(Quand partira le train vers Tunis)

(14.a) (14) المسافر: أريد تذكرة ذهاب إلى سوسة  
[AlmsAfr: Oryd t\*krp \*hAb ILY swsp]  
(L'utilisateur: je voudrai un billet aller vers Sousse)

(14.b) (14) الآلة: حسنا  
[Al/lp: HsnA]  
(La machine: d'accord)

(14.c) (14) المسافر: متى ينطلق القطار  
[AlmsAfr: mtY ynTlq AlqTAR]  
(L'utilisateur: quand partira le train)

## 7 – CONCLUSION

Dans ce papier, nous avons proposé une méthode de compréhension de la parole arabe spontanée fortement guidée par la sémantique. Cette méthode, basée sur l'approche sélective, se focalise sur l'analyse des segments porteurs de sens utiles et se base sur le formalisme *frame grammar* pour la représentation sémantique d'énoncés transcrits. Elle repose sur trois étapes principales à savoir, le pré-traitement de l'énoncé, l'analyse sémantique de l'énoncé pré-traité et la génération du ou des schéma(s) sémantique(s) correspondant(s).

L'étape de pré-traitement de l'énoncé transcrit permet de rendre ce dernier sous une forme

<sup>3</sup> <http://www.elda.fr> ; <http://www.elra.info>

plus convenable pour l'analyse sémantique. Cette étape consiste à supprimer les répétitions et les informations inutiles, à convertir les nombres écrits en toutes lettres, en chiffres, et à déterminer les formes canoniques des mots. Pour se faire, l'énoncé reconnu subit une normalisation, une analyse morpho-lexicale et un traitement des répétitions.

L'étape d'analyse sémantique a pour objectif d'attribuer des étiquettes sémantiques à tous les mots (ou groupements de mots) utiles de l'énoncé pré-traité. Elle consiste en un traitement de la synonymie, un étiquetage sémantique et un traitement des autocorrections.

L'étape de génération de schémas sémantiques consiste à identifier, dans un premier temps, le schéma sémantique qui correspond, au mieux, aux mots de référence et à remplir, dans un second temps, ce schéma en se basant sur les marqueurs de cas et les étiquettes sémantiques déterminées lors de l'analyse sémantique. En cas d'ambiguïtés, tous les schémas candidats sont générés.

La méthode de compréhension proposée a été testée à travers le système SARF, un serveur vocal interactif offrant aux usagers un accès aux informations sur le transport ferroviaire tunisien. L'évaluation du module de compréhension de SARF a montré que les taux de rappel, de précision et de F-Mesure sont respectivement 73.00%, 70.62% et 71.79%. Ces résultats sont encourageants même si SARF, dans sa version actuelle, ne traite pas certains problèmes de disfluences tels que l'autocorrection complexe qui ne fait pas intervenir des marqueurs de rectification, la reprise sous forme d'abandon d'un segment pour débiter un nouveau segment et les mots mal-reconnus.

Comme première perspective à court terme, nous envisageons d'étudier le problème des disfluences qui résultent d'une amorce (*i.e.*, mot commencé par le locuteur mais inachevé) en vue d'apporter des solutions, quant à leur résolution, par des techniques de calcul de distance entre le mot non reconnu (ou mal-reconnu) et les mots du lexique ayant la même étiquette sémantique que celle de l'étiquette potentielle du mot en question. En cas d'échec, nous envisageons de faire intervenir le gestionnaire du dialogue pour

poser une question à l'utilisateur afin de lever l'ambiguïté.

Comme deuxième perspective à court terme, nous projetons d'étudier les problèmes des autocorrections complexes (*i.e.*, correction d'un segment par un autre segment sans que ce dernier ne soit introduit par un marqueur de rectification) et des reprises (*i.e.*, abandon de segments et début de nouveaux segments). Certes le recours à un dialogue s'impose dans certains cas pour lever certaines de ces ambiguïtés.

## BIBLIOGRAPHIE

ANTOINE, J.-Y. (2003), "Pour une Ingénierie des Langues plus Linguistique", *HDR Computer Science*, University of South Bretagne, Vannes, France.

BAHOU, Y., BELGUITH HADRICH, L., BEN HAMADOU, A. (2008b), "Towards a Human-Machine Spoken Dialogue in Arabic", *6<sup>th</sup> Language Resources and Evaluation Conference (LREC'08), Workshop HLT Within the Arabic World: Arabic Language and Local Languages Processing Status Updates and Prospects*, Marrakech, Morocco, from 26 May to 1 June.

BAHOU, Y., BELGUITH HADRICH, L., BEN HAMADOU, A. (2008c), "Compréhension Automatique de la Parole Arabe Spontanée : Intégration dans un Serveur Vocal Interactif", *9<sup>th</sup> International Business Information Management Conference (IBIMA'08), Session Spéciale sur le Traitement de l'Information en Arabe*, Marrakech, Morocco, 4-6 January, pp. 1250-1259.

BAHOU, Y., SAFI, H., BELGUITH HADRICH, L. (2008a), "Analyse Sémantique des Énoncés Oraux Arabes dans un Contexte de Dialogue Homme-Machine", *XXVII<sup>èmes</sup> Journées d'Études sur la Parole (TALN-JEP'08)*, Avignon, France, 9-13 June.

BELGUITH HADRICH, L., CHAABEN, N. (2006), "Analyse et Désambiguïsation Morphologiques de Textes Arabes non Voyellés", *Actes de la 13<sup>ème</sup> Conférence sur le Traitement Automatique des Langues Naturelles (TALN'06)*, Leuven, Belgium, 10-13 April, pp. 493-501.

- BONNEAU-MAYNARD, H., DENIS, A., BECHET, F., DEVILLERS, L., LEFEVRE, F., QUIGNARD, M., ROSSET, S., VILLANEAU, J. (2008), *Media : Évaluation de la Compréhension dans les Systèmes de Dialogue*, L'Évaluation des Technologies de Traitement de la Langue, Les Campagnes Technolanguage (Traité IC2, Série Cognition et Traitement de l'Information), Stéphane Chaudiron et Khalid Choukri (Ed.), Hermès & Lavoisier, pp. 209-231.
- BOUSQUET-VERNHETTES, C. (2002), "Compréhension Robuste de la Parole Spontanée dans le Dialogue Oral Homme-Machine – Décodage Conceptuel Stochastique", *PHD Thesis*, University of Toulouse III–Paul SABATIER, France.
- BOVE, R. (2008), "A Tagged Corpus-Based Study for Repeats and Self-Repairs Detection in French Transcribed Speech", *Proceedings of the 11<sup>th</sup> International Conference on Text, Speech and Dialogue (TSD'08)*, Brno, Czech Republic, pp. 269-276.
- BRUCE, B. (1975), "Cases Systems for Natural Languages", *Artificial Intelligence*, Volume 6, pp. 327-360.
- DAHLBACK, N., JONSSON, A., AHRENBERG, L. (1993), "Wizard of Oz Studies - Why and How", *In Proceedings of the 1<sup>st</sup> International Conference on Intelligent User Interfaces (IUI'93)*, pp. 193-200.
- DAHMANI, H., BEDDA, M. (2004), "Conception d'un Système pour la Reconnaissance de Mots Enchaînés Arabes", *Traitement Automatique de la Langue Arabe Écrite et Parlée (TALN-JEP'04)*, Fes, Morocco, 19-22 April.
- FILLMORE, C. J. (1968), *The Case for Case*, In *Universals in Linguistic Theory*, E. Bach and R.T. Harms (Eds.), Holt Rinehart, New York.
- GOULIAN, J. (2002), "Stratégie d'Analyse Détaillée pour la Compréhension Automatique Robuste de la Parole", *PHD Thesis*, University of South Bretagne, Vannes, France.
- GOULIAN, J., ANTOINE, J.-Y., POIRIER, F. (2003), "How NLP Techniques Can Improve Speech Understanding: ROMUS – a Robust Chunk Based Message Understanding System Using Link Grammars", *EUROSPEECH'03*, Geneva, Switzerland, 1-4 September, pp. 2773-2776.
- JAMOUSSE, S. (2004), "Méthodes Statistiques pour la Compréhension Automatique de la Parole", *PHD Thesis*, University of Henri Poincaré, France.
- LUZZATI, D. (1989), "Recherches sur le Dialogue Homme-Machine : Modèles Linguistiques et Traitements Linguistiques", *PHD Thesis*, University of Paris III, France.
- MINKER, W. (1999), *Compréhension Automatique de la Parole Spontanée*, L'Harmattan.
- MINKER, W., BENNACEF, S. (2004), *Speech And Human-Machine Dialog*, Kluwer Academic Publishers Group, Pays-Bas.
- RAMSAY, A., MANSOUR, H. (2008), "Towards Including Prosody in a Text-To-Speech System for Modern Standard Arabic", *Computer Speech and Language*, Volume 22, Issue 1, pp. 84-103.
- SATORI, H., HARTI, M., CHENFOUR, N. (2007), "Arabic Speech Recognition System based on CMUSphinx", *In International Symposium on Computational Intelligence and Intelligent Informatics (ISCIII'07)*, Agadir, Morocco, 28-30 March, pp. 31-35.
- SENEFF, S. (1992), "TINA : a Natural Language System for Spoken Language Applications", *Computational Linguistics*, Volume 18, N°1, pp. 61-86.
- VILLANEAU, J., RIDOUX, O., ANTOINE, J.-Y. (2004), "LOGUS : Compréhension de l'Oral Spontané : Présentation et Evaluation des Bases Formelles de LOGUS", *Revue d'Intelligence Artificielle (RSTI-RIA'04)*, Volume 18, N°5-6, pp. 709-742.
- ZOUAGHI, A., ZRIGUI, M., ANTONIADIS, G. (2008), "Compréhension Automatique de la Parole Arabe Spontanée : Une Modélisation Numérique", *Traitement Automatique des Langues (TAL'08)*, Volume 49, N°1, pp. 141-166.
- ZOUAGHI, A., ZRIGUI, M., BEN AHMED, M. (2005), "Un Étiqueteur Sémantique des Énoncés en Langue Arabe", *Actes de la 12<sup>ème</sup> Conférence sur le Traitement Automatique des Langues Naturelles (TALN-RECITAL'05)*, Dourdan, France.