
Compréhension automatique de la parole arabe spontanée

Une modélisation numérique

Anis Zouaghi* — Mounir Zrigui** — Georges Antoniadis***

*Laboratoire RIADI, Équipe de Monastir
Faculté des sciences de Monastir
Anis.Zouaghi@riadi.rnu.tn

**Laboratoire UTIC, Équipe de Monastir
Faculté des sciences de Monastir
Mounir.Zrigui@fsm.rnu.tn

***Laboratoire LIDILEM
Université Stendhal, Grenoble 3
Georges.Antoniadis@u-grenoble3.fr

RÉSUMÉ. Notre travail s'intègre dans le cadre du projet intitulé « Oreillodule » : un système de reconnaissance, de traduction et de synthèse de la parole arabe. L'objectif de cet article est de présenter l'analyseur sémantique développé pour la compréhension automatique de la parole arabe spontanée. Cet analyseur est dédié, pour l'instant, à la compréhension de l'arabe standard. L'évaluation du système montre bien l'efficacité et la robustesse de ce dernier face aux caractéristiques principales de l'oral spontané, d'une manière générale, et aux spécificités de la langue arabe plus particulièrement.

ABSTRACT. This work is part of a large research project entitled "Oreillodule" aimed at developing tools for automatic speech recognition, translation, and synthesis for Arabic language. Our attention has mainly been focused on an attempt to present the semantic analyzer developed for the automatic comprehension of the standard spontaneous arabic speech. The findings on the effectiveness of the semantic decoder are quite satisfactory.

MOTS-CLÉS : compréhension automatique, parole arabe spontanée, analyse sémantique, modèle numérique, extraction automatique, contexte pertinent, information mutuelle moyenne, k-means.

KEYWORDS: automatic comprehension, spontaneous arabic speech, semantic analyse, probabilistic model, automatic extraction, pertinent context, mutual information, k-means.

1. Introduction

Nos travaux s'inscrivent dans le cadre de la compréhension automatique de la langue arabe, et dans le contexte des communications homme/machine finalisées. L'utilisation de modèles statistiques pour la reconnaissance et la compréhension de la parole ont l'avantage d'être portables vers d'autres domaines, ou vers des applications multilingues (Minker, 1999). Se fonder sur de tels modèles, pour l'interprétation sémantique des mots de l'énoncé, nécessite généralement la considération du contexte d'énonciation. Les informations contextuelles jouent un rôle primordial dans la détermination du sens adéquat d'un mot dans un énoncé donné. Ces informations permettent de lever les ambiguïtés d'interprétation et d'améliorer les performances du système de compréhension (Bousquet-Vernhettes, 2002). Dans l'approche standard, le décodage du sens d'un mot est généralement déduit en analysant le contexte qui le précède ou/et qui le succède immédiatement. Or, dans le cas de la compréhension de l'arabe spontané, ceci n'est pas toujours optimum. D'après l'étude que nous avons réalisée sur l'impact de la nature et du type du contexte sur le résultat de compréhension automatique de la langue arabe spontanée (voir figure 13, section 6), le taux d'une mauvaise compréhension est égal à 45 %, en considérant uniquement le sens du mot qui précède le mot à interpréter sémantiquement. Nous proposons alors de considérer plusieurs informations contextuelles en même temps pour l'analyse sémantique de l'oral arabe spontané : informations de type illocutoire, sémantique et linguistique. Ces informations contextuelles ont permis d'atteindre un taux de performance total de l'ordre de 29 % (voir figure 13). Ce taux passe à 9,9 %, dans le cas où le nombre des énoncés utilisés pour l'apprentissage est suffisant (voir figure 12).

2. La compréhension automatique de la parole arabe spontanée

L'efficacité et la performance d'un système de compréhension automatique de la parole arabe spontanée dépendent de sa robustesse et sa capacité à lever les difficultés de traitement du langage naturel. Certaines difficultés sont d'ordre linguistique. Celles-ci sont communes à la compréhension de l'écrit et de l'oral. Ces difficultés sont généralement dues à l'utilisation des références, des mots polysémiques, des prédicats vagues, de l'implicite, etc. D'autres sont dues aux caractéristiques de l'oral spontané d'une manière générale et aux spécificités de la langue arabe plus particulièrement.

2.1. *Les spécificités de la langue arabe*

La langue arabe, par ses propriétés et ses spécificités morphologiques et syntaxiques, est considérée comme une langue difficile à traiter automatiquement

(Larkey *et al.*, 2002). Parmi les caractéristiques de cette langue qui rendent son traitement ambigu, citons :

– un mot arabe peut signifier toute une expression en anglais ou en français (Chaabene *et al.*, 2003). Par exemple le mot «*أرأيت*» (*ara ayta*), permet d'exprimer l'équivalent en français de «*Est-ce que vous avez vu ?*». Ainsi la compréhension automatique de tels mots nécessite une segmentation préalable, tâche qui n'est pas évidente ;

– le double rôle de la lettre «*و*» (qui peut correspondre à la lettre «*w*», ou à la conjonction «*et*» en français), pouvant être employée comme une première lettre d'un mot (par exemple *رق* (*warak*) «*feuilles*»), ou comme une conjonction de coordination (voir énoncé E). Ce phénomène entraîne une ambiguïté au niveau de la distinction entre les deux rôles possibles de «*و*», ce qui rend difficile la tâche d'identification automatique des différentes propositions d'un énoncé en vue de leur analyse sémantique.

(E) حجز مكان وأريد معرفة توقيت القطار الذاهب إلى تونس

Traduction en français : *Je veux connaître l'horaire du train allant à Tunis et réserver une place ;*

– l'ordre des mots dans une phrase en langue arabe est relativement variable par rapport, par exemple, à la langue française, dont l'ordre d'agencement des mots dans une phrase respecte habituellement l'ordonnement SVO (sujet, verbe, objet) (Kurdi, 2003). En arabe, nous avons toujours le libre choix du terme que nous voulons mettre en valeur au début de la phrase (Blachère *et al.*, 1975). Ceci rend complexe la tâche de construction de la grammaire qui sera utilisée dans le processus d'interprétation sémantique. Il faut en prévoir toutes les règles de combinaisons possibles, décrivant l'inversion de l'ordre des mots dans une phrase ;

– certaines lettres de la langue arabe, comme par exemple : ف (f : prononciation Fa), ح (h : prononciation Hha'), خ (kh : prononciation Kha'), ض (d : prononciation Dad), ذ (d : prononciation Thal), ظ (z : prononciation Zah), sont prononcées au moyen d'une forte expiration ; ainsi la qualité du microphone peut influencer sur les résultats de reconnaissance de la parole (Saïdane *et al.*, 2005) ;

– la nature de prononciation de certaines lettres arabes, comme par exemple : غ (gh : prononciation Ghayn) (Saïdane *et al.*, 2005). Ce phonème n'a pas d'équivalent dans d'autres langues, comme, par exemple, le français ou l'anglais ;

– la possibilité d'existence de plusieurs réalisations graphémiques pour un même phonème (par exemple les graphèmes ض et ظ), ou de plusieurs réalisations phonétiques pour un même graphème (par exemple le graphème ل «*lam*» possède deux réalisations phonétiques différentes, dépendant des lettres qui le précèdent et le succèdent dans le mot, comme dans : بالله «*s'il te plaît*» et الله «*dieu*»). Certains graphèmes peuvent ne pas être considérés lors de la prononciation (Ghazali *et al.*, 1990). Ce phénomène rend la tâche de reconnaissance de la parole plus ardue, et, par conséquent, la tâche de compréhension aussi.

2.2. Les systèmes de compréhension de la langue arabe

La langue arabe est la sixième langue la plus parlée dans le monde avec environ 250 millions de locuteurs. Avec sa diffusion sur le Web, au cours de la dernière décennie, l'intérêt de l'étude de cette langue en vue de son traitement automatique est croissant. Nous observons de plus en plus de recherches et d'outils nécessaires au développement de son traitement automatique (Boualem, 1993 ; Jaccarini, 1997 ; Debili, 2001 ; Gaubert, 2001 ; Aloulou *et al.*, 2002 ; Dichy *et al.*, 2002 ; Baccour *et al.*, 2003 ; Alrahabi *et al.*, 2004 ; Tuerlinckx, 2004 ; Haddad *et al.*, 2005), etc. Des conférences et des workshops ont été organisés spécialement pour son traitement automatique, tels que *JEP-TALN'04* (traitement automatique de la langue arabe, à Fès au Maroc), *COLING'04* (Computational Approaches to Arabic Script-based Languages, à Genève Suisse), Arabic Language Resources and Tools Conference, au Caire Égypte (2004), et *CITALA'07* (Colloque International sur le Traitement Automatique de la Langue Arabe).

Malheureusement, et contrairement à certaines langues telles que l'anglais ou le français, le problème de la représentation et de l'analyse sémantique de la parole arabe spontanée reste encore très peu abordé au niveau de la recherche scientifique. La majorité des efforts déployés par les chercheurs du domaine sont concentrés sur la réalisation d'analyseurs morphologiques et syntaxiques, tels que (Al-Fedaghi, 1989 ; Beesely, 2001 ; Ditters, 2001 ; Ouersighni, 2001 ; Aloulou, 2003). Quelques rares travaux s'intéressent plus à l'interprétation sémantique de l'arabe écrit : c'est le cas de (El-Dessouki, 1988 ; Mankai, 1996 ; Al-Johar, 1997 ; Al-Muhtaseb, 1997 ; Meftouh *et al.*, 2001 ; Haddad, 2001 et 2005). Le système *Al Biruni* de (Mankai, 1996) repose sur une combinaison du formalisme de la grammaire de cas de Fillmore (Fillmore, 1968) avec la théorie sens texte de Mel'cuk, en vue de représentation, de manipulation et d'analyse de textes arabes. Le système de (Haddad *et al.*, 2005) utilise la grammaire d'unification HPSG qui permet d'intégrer des connaissances syntaxiques et sémantiques dans une même grammaire, afin d'aboutir à une analyse profonde. Dans les travaux de Laskri (Meftouh *et al.*, 2001), une approche neuronale est utilisée ; elle se base sur le formalisme de la grammaire de cas de Fillmore pour générer le sens d'une phrase en langue arabe. Voici, un exemple d'interprétation sémantique d'une phrase en employant le modèle des grammaires de cas (Laskri, 2002) :

– Phrase à analyser : *Khaled ira au stade demain pour regarder le match.*

– La représentation sémantique correspondant à cette phrase :

- agent : Khaled

- action : ira

- temps : demain

- destination : au stade

- but : pour regarder
- objet : le match

L'avantage majeur du formalisme des grammaires de cas est qu'ils font partie des rares formalismes autorisant la compréhension automatique des phrases ne respectant pas les règles de la grammaire de l'écrit. Dans la section suivante, nous présentons l'architecture générale et les différentes composantes du système développé pour la compréhension automatique de la parole arabe spontanée.

3. Présentation du système de compréhension

3.1. Architecture du système

Le système de compréhension développé permet de construire la représentation sémantique d'un énoncé, sous forme d'un ensemble d'associations attributs/valeurs (voir figure 1 et 4). Les systèmes utilisant une approche logique construisent, généralement, une formule logique, comme, par exemple, le système *LOGUS* (Villaneau *et al.*, 2005). Un ensemble de paires attributs/valeurs peut être vu comme un ensemble de couples [attribut = valeur], dont la *valeur* ne peut être qu'un entier, un réel, ou une chaîne. Ce formalisme est très employé en compréhension de la parole dans les systèmes de communication orale tels que (Pepelnjak *et al.*, 1995 ; Bousquet, 2002 ; Servan *et al.*, 2006).

– Énoncé présenté au système en vue de son analyse sémantique :

(1) أريد حجز مكان بالقطار الذهاب إلى تونس.

Je veux réserver une place dans le train allant à Tunis.

Représentation sémantique générée :

$$\left(\begin{array}{l} \text{نوعية (Type) = حجز (demande de réservation)} \\ \text{مدينة_وجهة (ville_destination) = تونس (Tunis)} \\ \text{عدد_مقاعد (nombre_places) = 1} \end{array} \right)$$

Figure 1. Exemple de représentation sémantique générée par le système

La figure 2 ci-dessous présente l'architecture générale du système de compréhension de la parole arabe spontanée.

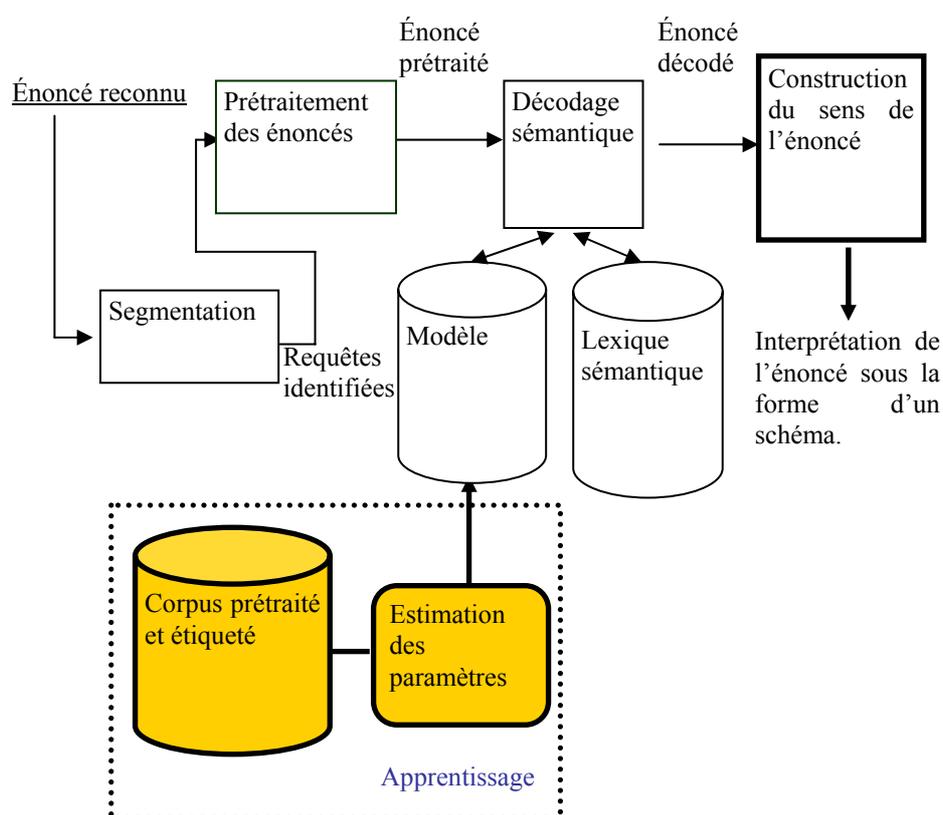


Figure 2. Architecture générale du système

3.2. Les composantes du système

Comme le montre la figure 2 ci-dessus, le processus de construction, par le système, du sens des énoncés comporte quatre étapes principales :

- segmentation des énoncés transcrits par le module de reconnaissance de la parole : ce traitement permet d'identifier les différentes requêtes du message du locuteur. Un même message peut être constitué d'une ou plusieurs requêtes à la fois. En ce sens, il est nécessaire que le système puisse identifier les différentes requêtes du message, afin d'interpréter la demande de l'utilisateur dans son intégralité.

- prétraitement des énoncés : il consiste principalement à éliminer les *mots vides* qui n'ont aucune influence sur la détermination du sens de l'énoncé (tel que le terme d'appui du discours *إِذًا* (*alors*)), et à regrouper certains mots en une seule entrée ;
- décodage sémantique : il permet de déterminer la signification de chaque mot de l'énoncé ;
- construction proprement dite du sens de l'énoncé : il permet de générer l'ensemble des paires attribut/valeur.

Le décodage sémantique des énoncés prétraités repose sur un modèle numérique qui encode les règles de la grammaire (voir section 4), et sur un lexique sémantique (voir figure 3 ci-dessous). Le lexique sémantique est un ensemble d'associations de la forme : mot/ensemble de traits sémantiques TSE décrivant le sens du mot (voir définition 1) + ensemble de traits syntaxiques TSY décrivant les caractéristiques du mot (nature, nombre et type). Par exemple, le sens du mot *الناهب* (*allant*) est décrit comme suit :

الناهب (*allant*) → TSE {(transport) *نقل*, (mouvement) *حركة*, (destination) *وجهة*} +
 TSY = {(masculin) *مذكر*, (singulier) *مفرد*, (nom) *اسم*}.

Définition 1 : ensemble de traits sémantiques (TSE)

Un *TSE* est un ensemble de trois traits sémantiques $\{D, C, TM\}$, permettant la représentation de la signification de chaque *mot non vide* :

- le premier trait est de type *domaine*, noté *D*. Il permet d'indiquer le domaine de l'application finalisée traitée comme, par exemple, des renseignements touristiques, ferroviaires, etc. ;
- le deuxième trait est de type *classe sémantique*, noté *C*. Il permet d'indiquer la classe sémantique à laquelle appartient un mot. Ce type de trait permet de regrouper les mots synonymes ou possédant un même rôle sémantique dans une application donnée ;
- le troisième est de type différentiel, noté *TM*. Il permet d'opposer les mots les uns aux autres et de préciser en quoi leur sens diffère.

La figure 3 suivante décrit la structure et le contenu du lexique sémantique employé.

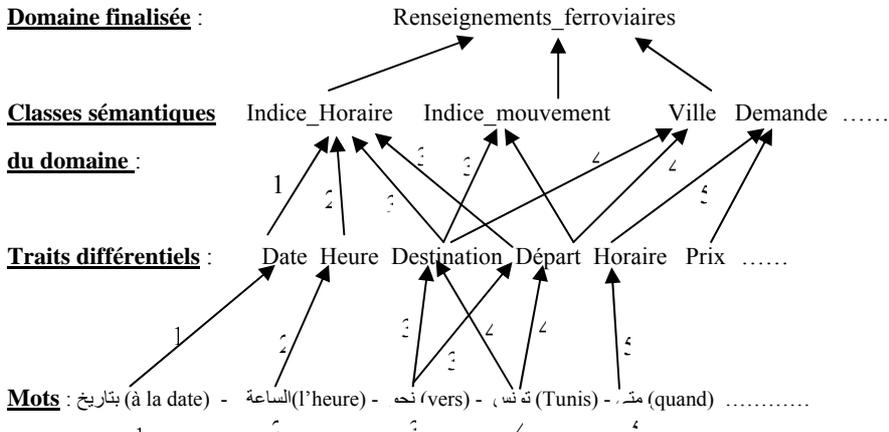


Figure 3. Description des mots d'un domaine à partir de TSE.

Les numéros employés dans la figure 3, indiquent les traits sémantiques qui peuvent être utilisés pour la représentation sémantique de chaque mot, à chaque niveau de description.

Par exemple :

$TSE[\text{نحو (vers)}] = TSE1\{\text{INDICE_HORAIRE, DESTINATION}\},$ ou $TSE2\{\text{INDICE_HORAIRE, DÉPART}\},$ ou $TSE3\{\text{INDICE_MOUVEMENT, DESTINATION}\}$

dans le domaine des renseignements ferroviaires. Voici, dans l'ordre, un exemple d'utilisation de chacun de ces ensembles de traits pour la représentation du sens du mot نحو (vers) :

اريدالقطار الذي يصل نحو الساعة الثامنة (2)

Je veux le train qui arrive vers huit heures.

اريد الانطلاق نحو الساعة الثامنة مساء (3)

Je veux partir vers huit heures du soir.

تونس (4) نحو اريد حجز تذكرة

Je veux réserver un billet vers Tunis.

4. Présentation du modèle numérique

4.1. Les principaux modèles existants

Il existe de nombreux systèmes de compréhension du langage oral. Pour la détermination du sens des énoncés, ces systèmes utilisent des formalismes de représentation des connaissances linguistiques de différentes natures. Parmi les principaux modèles employés citons :

- les grammaires hors contexte telles que (Seneff, 1992 ; Yan *et al.*, 2001) ;
- les grammaires de cas telles que (Bennacef *et al.*, 1994 ; Minker, 1999) ;
- les modèles de Markov cachés tels que (Bousquet, 2002) ;
- les réseaux de neurones tels que (Jamousi, 2004) ;
- les modèles de langage n-grammes tels que (Knight *et al.*, 2001) ;
- les logiques telles que (Sadek *et al.*, 1995 ; Villaneau *et al.*, 2004) ;
- les grammaires d'unification telles que (Lopez *et al.*, 1998 ; Kurdi, 2001).

4.2. Le modèle proposé

4.2.1. Principe

Les modèles numériques présentent plusieurs avantages par rapport aux modèles à syntaxe fixe (à base de règles) qui décrivent explicitement les interprétations possibles (Bennacef *et al.*, 1994). L'un des principaux avantages est qu'ils sont portables vers d'autres domaines ou vers des applications multilingues (Minker, 1999). Les modèles à syntaxe fixe sont adaptés d'une façon à obtenir des performances optimales pour une tâche spécifique, ce qui les rend plus rigides concernant l'analyse. De plus, les modèles numériques ont déjà montré leur efficacité pour le traitement automatique d'autres langues comme, par exemple, l'anglais.

Nous concernant, nous proposons un modèle de langage numérique hybride pour la compréhension automatique de la parole arabe spontanée. L'originalité de ce modèle réside dans le choix des ensembles de traits TSE candidats à la description du sens d'un énoncé, en tenant compte de plusieurs types de données à la fois. Les informations considérées par le modèle proposé sont les suivantes :

- le type de l'acte illocutoire accompli par l'énoncé : une demande, un refus ou une excuse ;
- le type de l'énoncé : il s'agit, par exemple, d'une demande de réservation, de

tarifs, etc. ;

– des informations linguistiques : les mots appartenant au contexte du mot courant à interpréter ;

– des informations sémantiques : les traits sémantiques qui ont été utilisés pour l’annotation sémantique des mots appartenant au contexte du mot courant à interpréter.

La considération de l’ensemble de ces informations en même temps, nous a permis d’atteindre un taux d’erreur égal à 9,9 % avec 2 726 énoncés appris. Ce taux est beaucoup plus élevé en n’utilisant que les mots appartenant au contexte droit, (compris entre 48,6 % et 57 %, voir figure 12 et 13, section 6).

Exemple : soit l’énoncé simple (5) suivant :

بكم سعر الذهاب إلى تونس ؟ (5)

Combien coûte le prix pour aller à Tunis ?

Pour construire la représentation sémantique de cet énoncé, notre système procède comme suit :

– chercher l’acte illocutoire : par défaut l’acte accompli est une demande ;

– chercher le type de l’énoncé : ceci consiste à détecter les mots-clés présents dans l’énoncé, afin d’activer l’un des schémas sémantiques (définis au préalable) susceptible de représenter le sens de cette requête. Ici, il s’agit du schéma *TarPRIX* (voir figure 4), identifié à travers la présence de بكم (combien) et سعر (prix). Ainsi, déterminer le sens de cet énoncé revient à remplir les différents attributs de ce schéma avec les valeurs correspondantes. Les valeurs des attributs sont directement déduites à partir des TSE (utilisés pour annoter les énoncés). Par exemple, si un mot est annoté avec le TSE : {VILLE, DESTINATION}, alors ce mot sera la valeur de l’attribut : ville_destination ;

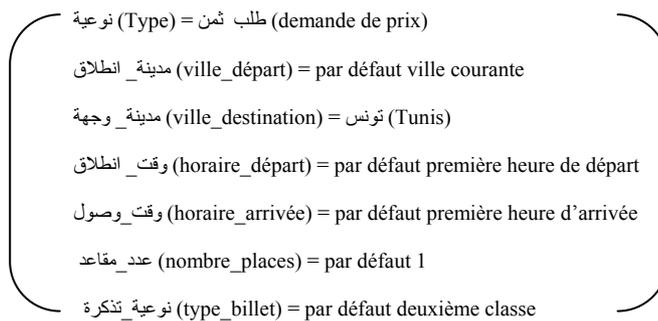


Figure 4. Schéma prix

– annoter sémantiquement les mots jugés porteurs de sens pour le processus de compréhension de cet énoncé (mots déterminés après une étape de prétraitement, voir section 3) : cette annotation est réalisée d'une manière incrémentale, les mots sont annotés dans l'ordre de leur apparition dans l'énoncé. À chaque incrémentation, on considère les mots présents dans le contexte, les traits sémantiques utilisés pour l'annotation du contexte, le type illocutoire et le type de l'énoncé. Supposons, par exemple, que l'on veut annoter sémantiquement le mot *Tunis* de l'énoncé considéré plus haut. Le décodeur commence par chercher le contexte pertinent du mot Tunis (voir la dernière partie de la section 4). Il utilise ensuite ce contexte annoté sémantiquement pour calculer le sens de *Tunis* (voir section 4.2.3). Il s'agit ici d'utiliser le contexte pertinent, الذهاب (aller) et إلى (à), qui représentent les informations linguistiques, et les ensemble de traits sémantiques, {Mouvement, Destination}, et {Indice_mouvement, Destination} qui représentent les informations sémantiques.

Le modèle est inspiré de l'approche classique fondée sur le critère du maximum d'entropie (MaxEnt), qui permet de mélanger diverses sources d'information dans le calcul des probabilités d'une suite de classes, étant donné une suite de mots.

Par rapport aux modèles de langage classiques de type Part Of Speech POS (tels que les modèles n-grammes ou les n-classes), ce modèle permet d'intégrer, en plus du contexte pertinent, des informations sur la nature illocutoire et le type de l'énoncé. Dans ce qui suit, nous présentons le modèle proposé pour le décodage sémantique de la parole arabe spontanée.

4.2.2. Description formelle du modèle

Les systèmes à base de modèles de langage numériques tentent de déterminer le score d'une séquence de mots $S = m_1, m_2, \dots, m_i$, dont la formule générale est décrite par l'équation [1] suivante :

$$P(S) = P(m_1) \times P(m_2 / m_1) \times \dots \times P(m_i / m_1, m_2, \dots, m_{i-1}) \quad [1]$$

Dans le cas de l'interprétation d'une séquence de mots significatifs M_1, M_2, \dots, M_n , par les ensembles de traits sémantiques $TSE_1, TSE_2, \dots, TSE_n$ (voir paragraphe 5 concernant la description de ces traits), le modèle tente de déterminer le score d'interprétation de chacun de ces mots, par chacun de ces ensembles traits.

Soit $I = TSE_1 \rightarrow M_1, TSE_2 \rightarrow M_2, \dots, TSE_n \rightarrow M_n$, le score de l'interprétation I (qui décrit le sens de M_1, M_2 et M_3 respectivement par TSE_1, TSE_2 , et TSE_3) est alors défini comme suit :

$$\begin{aligned} P(I) &= P(TSE_1 \dots TSE_n / M_1 \dots M_n) \\ &= P(TSE_1 / M_1) \times P(TSE_2 / TSE_1, M_1 M_2) \times \dots \times P(TSE_n / TSE_1 \dots TSE_{n-1}, M_1 \dots M_{n-1}) \\ &= P(TSE_1 / M_1) \times P(TSE_2 / TSE_1, M_2) \times \dots \times P(TSE_n / TSE_1 \dots TSE_{n-1}, M_n) \quad [2] \end{aligned}$$

Le passage de la deuxième à la troisième ligne de la formule [2] correspond à

une approximation du modèle, qui considère que la probabilité d'un TSE_i ne dépend, conditionnellement à la séquence complète des ensembles de traits, que du mot courant M_i. En fixant à l'avance le domaine l'application, chaque mot significatif M_i peut être interprété *via* un ensemble de traits sémantiques de la forme TSE_i = (C_i, TM_i) (voir définition 1, section 3.2.). L'équation [3] est alors transformée sous la forme suivante :

$$P(I) \approx P((C_1, TM_1) / M_1) \times P((C_2, TM_2) / (C_1, TM_1), M_2) \times \dots \times P((C_n, TM_n) / (C_1, TM_1) \times \dots \times (C_{n-1}, TM_{n-1}), M_n) \quad [3]$$

Afin de minimiser le nombre des ensembles de traits TSE candidats et améliorer les performances du décodeur sémantique, nous avons intégré dans l'équation [3] d'autres sources d'information (la nature illocutoire et le type de l'énoncé) pouvant participer à la sélection des TSE. Ainsi, la probabilité d'interprétation de chaque mot M_i par un ensemble de traits donné TSE_i = (C_i, TM_i), en tenant compte de la nature et du type de l'énoncé noté par NT_j, est donnée par :

$$P(I) = P(TSE_1, \dots, TSE_n / NT_j, M_1 \dots M_n) = P(NT_j / M_1 \dots M_n) \times P((C_1, TM_1) / NT_j, M_1) \times P((C_2, TM_2) / NT_j, (C_1, TM_1), M_2) \times \dots \times P((C_n, TM_n) / NT_j, (C_1, TM_1), \dots, (C_{n-1}, TM_{n-1}), M_n) \quad [4]$$

où P(NT_j / M₁, M₂, ..., M_{i-1}, M_i) est la probabilité pour que l'énoncé soit de type NT_j, sachant que l'énoncé est composé des mots significatifs M₁, ..., M_n. Signalons que le premier mot est traité à part, en l'annotant par défaut avec la classe C=demande, dans le cas de l'application finalisée considérée (demandes de renseignements ferroviaires). La détermination du trait TM est opérée en recherchant le premier mot dans la classe des mots demande (qui regroupe tous les mots exprimant une demande, tels que : *combien*, *quand*, etc.). Cette classe contient une dizaine de mots. Afin de réduire la complexité du modèle, et améliorer sa capacité de prédiction, nous avons appliqué certaines approximations.

4.2.3. Approximations du modèle

La première approximation appliquée à ce modèle consiste à ne considérer pour la détermination du type de l'énoncé que certains mots appelés mots-clés. Les mots-clés sont des mots dont l'occurrence dans un énoncé permet de déterminer le type de celui-ci (demande de réservation, demande de tarif, etc.). Les mots-clés sont déterminés à partir du corpus d'apprentissage, où l'on associe à chaque énoncé une étiquette identifiant le type de l'énoncé. Ces mots peuvent être des unigrammes, ou des bigrammes, expressions composées de deux mots (voir figure 11, section 5.3. concernant les mots-clés de l'application considérée). Considérons l'énoncé composé des n mots significatifs suivants : M₁ M₂ M₃ M₄ ... M_n; où, les mots notés par M_{r_j} avec j ∈ {1, ..., n} désignent les mots-clés de cet énoncé. En ne tenant compte que des mots de référence M_{r_j} pour le calcul de la probabilité qu'un énoncé soit de type NT_i, on obtient la substitution suivante :

$$P(NT_i / M_1, M_2, \dots, M_n) = P(NT_i / M_{r_1} \dots M_{r_k}) \quad [5]$$

$$\text{avec : } P(NT_i / Mr_1 \dots Mr_k) = N(Mr_1 \dots Mr_k) / N(NT_i(Mr_1 \dots Mr_k))$$

$N(Mr_1 \dots Mr_k)$ est le nombre d'énoncés contenant, en même temps, les mots Mr_1, \dots, Mr_k dans le corpus d'apprentissage, et $N(NT_i(Mr_1 \dots Mr_k))$ est le nombre d'énoncés qui contiennent les mots Mr_1, \dots, Mr_k à la fois et qui sont identifiés de type NT_j dans le corpus d'apprentissage. Cette substitution permet de réduire le nombre de paramètres à estimer, puisque le nombre de mots de référence est très inférieur à la taille du vocabulaire de l'application. Dans le cadre de l'application considérée, un énoncé contient au maximum trois mots de référence. La deuxième hypothèse de modélisation porte sur les relations d'indépendance conditionnelle dans le modèle et concerne la probabilité jointe :

$$P((C_i, TM_i) / M_i, NT_j, (C_1, TM_1), \dots, (C_{i-1}, TM_{i-1})).$$

Cette probabilité peut s'écrire comme suit :

$$\begin{aligned} P((C_i, TM_i) / M_i, NT_j, (C_1, TM_1), \dots, (C_{i-1}, TM_{i-1})) = \\ P(C_i / M_i, NT_j, (C_1, TM_1) \dots (C_{i-1}, TM_{i-1})) \times P(TM_i / M_i, C_i, NT_j, (C_1, TM_1) \dots (C_{i-1}, \\ TM_{i-1})) \end{aligned} \quad [6]$$

Afin de simplifier le modèle, nous avons considéré uniquement les ensembles des traits sémantiques pertinents TSEP(CP, TMP) (voir définition 2, ci-dessous) pour la prédiction de l'ensemble des traits sémantiques correspondant à un mot M_i noté par $TSE(M_i)$.

Définition 2 : les ensembles des traits sémantiques pertinents (TSEP)

Un ensemble TSE est considéré comme pertinent pour l'interprétation d'un mot M_i , si cet ensemble TSE a été utilisé pour la description du sens d'un autre mot M_j possédant une affinité sémantique forte avec le mot M_i .

La manière de déterminer le degré d'affinité sémantique entre les mots est détaillée dans le paragraphe suivant. La deuxième approximation considérée est que la classe sémantique C_i de l'ensemble $TSE(M_i)$, ne dépend que des deux classes pertinentes CP_{i-1} et CP_{i-2} employées dans la description du sens des mots appartenant au contexte droit de M_i et du type de l'énoncé. Nous avons ainsi :

$$P(C_i / NT_j, (C_1, TM_1) \dots (C_{i-1}, TM_{i-1}), M_i) \approx P(C_i / NT_j, CP_{i-1}, CP_{i-2}) \quad [7]$$

En tenant compte du fait que les mots appartenant à une même classe sémantique ne sont pas tous équiprobables, on peut intégrer dans l'équation [7] la probabilité de chacun des mots au sein de leur classe. On obtient dans ce cas la formule suivante :

$$P(C_i / NT_j, (C_1, TM_1) \dots (C_{i-1}, TM_{i-1}), M_i) = P(M_i / C_i) \times P(C_i / NT_j, CP_{i-1}, CP_{i-2})$$

$P(M_i / C_i)$ est la probabilité d'apparition du mot m dans la classe C_i . Cette probabilité est estimée comme suit : $P(M_i / C_i) = N(C_i(M_i)) / N(M_i)$

$N(M_i)$ est le nombre total d'occurrences de M_i dans le corpus d'apprentissage, et $N(C_i(M_i))$ le nombre d'utilisation de la classe C_i pour la description du sens de M_i .

La probabilité $P(C_i / NT_j, CP_{i-1}, CP_{i-2})$ est estimée comme suit :

$$P(C_i / NT_j, CP_{i-1}, CP_{i-2}) = N(NT_j(C_i, CP_{i-1}, CP_{i-2})) / N(NT_j(CP_{i-1}, CP_{i-2}))$$

$N(NT_j(C_i, CP_{i-1}, CP_{i-2}))$ est le nombre d'énoncés de type NT_j étiquetés simultanément par C_i, CP_{i-1} et CP_{i-2} . $N(NT_j(CP_{i-1}, CP_{i-2}))$ est le nombre d'énoncés de type NT_j étiquetés simultanément par CP_{i-1} et CP_{i-2} .

Une autre approximation considérée, est que le trait élémentaire TM_i du $TSE(M_i)$, à un instant t ne dépend que de la classe C_i affectée à M_i et de l'ensemble de traits pertinent précédant $TSEP_{i-1}$. Ainsi nous avons :

$$\begin{aligned} P(TM_i / M_i, C_i, NT_j, (C_1, TM_1), \dots, (C_{i-1}, TM_{i-1})) &\approx P(TM_i / C_i, TSEP_{i-1}) \\ &\approx P(TM_i / C_i, CP_{i-1}, TMP_{i-1}) \quad [8] \end{aligned}$$

$$\text{avec : } P(TM_i / C_i, TSEP_{i-1}) = N(TM_i, C_i, TSEP_{i-1}) / N(C_i, TSEP_{i-1})$$

$N(TM_i / C_i, TSEP_{i-1})$ est le nombre d'énoncés contenant des mots interprétés par $TSE_i = \{TM_i, C_i\}$ et $TSEP_{i-1}$. $N(C_i, TSEP_{i-1})$ est le nombre d'énoncés contenant des mots interprétés par C_i et $TSEP_{i-1}$.

À partir de ces deux approximations, [7] et [8], nous déduisons de l'équation [6] :

$$\begin{aligned} P((C_i, TM_i) / M_i, NT_j, (C_1, TM_1), \dots, (C_{i-1}, TM_{i-1})) &\approx \\ P(M_i / C_i) \times P(C_i / NT_j, CP_{i-1}, CP_{i-2}) \times P(TM_i / C_i, CP_{i-1}, TMP_{i-1}) & \quad [9] \end{aligned}$$

Enfin, à partir des équations [5] et [9], nous déduisons à partir de [4] la probabilité d'interprétation du mot M_i pour l'ensemble de traits sémantiques $TSE_i = \{C_i, TM_i\}$ sachant que l'énoncé est de type NT_j :

$$\begin{aligned} P((C_i, TM_i) / M_i, NT_j) &= P((C_i, TM_i) / M_i, NT_j) = \\ P(NT_j / M_i) \times P(M_i / C_i) \times P(C_i / NT_j, CP_{i-1}, CP_{i-2}) \times P(TM_i / C_i, CP_{i-1}, TMP_{i-1}) & \quad [10] \end{aligned}$$

Ainsi, à partir de l'équation [10], l'interprétation sémantique la plus probable IS d'un mot M_i est déterminée à partir de la formule [11] suivante :

$$IS(M_i) = \operatorname{argmax}_j P(NT_j / M_i) \times P(M_i / C_j) \times P(C_j / NT_j, CP_{j-1}, CP_{j-2}) \times P(TM_j / C_j, CP_{j-1}, TMP_{j-1}) \quad [11]$$

Ce modèle a été employé pour la sélection des ensembles TSE candidats à l'interprétation des mots des énoncés transcrits par le module de reconnaissance de la parole.

4.2.4. Amélioration du modèle

En remarquant que ce n'est pas obligatoirement les mots précédant immédiatement le mot à interpréter qui ont une influence sémantique sur ce dernier, nous avons décidé de n'employer, lors de la phase de décodage du sens d'un mot,

que les TSE des deux mots possédant la plus grande affinité sémantique avec celui-ci (la manière de calcul de la plus grande affinité sémantique est décrite par la formule [12] ci-dessous). Pour atteindre cet objectif, nous nous sommes basés sur la notion d'information mutuelle moyenne de (Rosenfeld, 1994), qui permet de calculer le degré de cooccurrence de deux mots donnés. Cette méthode nous a permis de ne plus utiliser systématiquement les TSE des deux mots qui précèdent immédiatement le mot à décoder. La formule de l'information mutuelle moyenne *IMM* est la suivante :

$$IMM(M_i, MED_j) = P(M_i, MED_j) \times \text{Log} [P(M_i / MED_j) / P(M_i).P(MED_j)] + P(\overline{M}_i, MED_j) \times \text{Log} [P(\overline{M}_i / MED_j) / P(\overline{M}_i).P(MED_j)] + P(M_i, \overline{MED}_j) \times \text{Log} [P(M_i / \overline{MED}_j) / P(M_i).P(\overline{MED}_j)] + P(\overline{M}_i, \overline{MED}_j) \times \text{Log} [P(\overline{M}_i / \overline{MED}_j) / P(\overline{M}_i).P(\overline{MED}_j)] ;$$

avec : $1 \leq j \leq i-1$ [12]

Les MED_j sont les mots qui appartiennent au contexte droit du mot M_i dans l'énoncé, c'est-à-dire les mots qui précèdent M_i . L'affinité sémantique la plus forte, ou maximale, *AffM* que possède le mot M_i avec son contexte droit est déterminée alors à partir de la formule [13] suivante :

$$AffM(M_i, MED) = \max_{1 \leq j \leq i-1} IMM(M_i, MED_j) \quad [13]$$

5. L'application finalisée considérée

5.1. Le domaine et la complexité de la tâche

Pour tester et estimer les paramètres du modèle, nous avons utilisé un corpus représentant le domaine des renseignements ferroviaires. La principale raison du choix de ce domaine d'application est la taille statistiquement représentative du corpus d'apprentissage dont nous disposons (voir figures 5 et 6 ci-dessous). Ce corpus a été collecté en demandant à mille personnes différentes de formuler des énoncés écrits relatifs aux renseignements ferroviaires, en utilisant la méthode du magicien d'Oz. Il s'agit, donc, d'un corpus simulé et non réel.

Domaine	Taille (Mo)	Nombre d'énoncés	Nombre de mots total	Nbre moyen de mots/énoncé	Nbre moyen de requêtes/énoncé	Nombre de locuteurs
Renseignements ferroviaires	3,4	10 000	430 000	43	entre 2 et 3	1 000

Figure 5. *Caractéristiques du corpus de point de vue volume*

Nature de la tâche	Renseignements sur les:				Réservations
	horaires	trajets	tarifs	durées	
Taux de sa représentation					10,41 %
	28,7 %	9,37 %	16,66 %	3,12 %	

Figure 6. *Caractéristiques du corpus de point de vue contenu*

Le tableau de la figure 7 suivante, donne des indications sur la complexité de la tâche considérée : nombre d'attributs, nombre de TSE, nombre de paires (attribut, valeur) différentes, nombre de formulaires décrivant le sens des énoncés, et nombre total de type de requêtes.

Nombre de TSE	37	Nombre de paires (attributs, valeurs)	15
Nombre de C	12	Nombre de schémas ou formulaires	6
Nombre de TM	20	Nombre de type de requêtes NT	10

Figure 7. *Quelques éléments indicateurs de la complexité de la tâche*

Le tableau de la figure 8 suivante, donne des indices de comparaison de la complexité de la tâche considérée, avec celle d'une tâche de même genre, ARISE (Minker, 1999).

Tâche	Indices de complexité	
	ARISE	Notre tâche
Types de requêtes	15	10
Attributs valeurs	76	37
Nombre de requêtes	15000	25000
Nombre moyen de mots/requête	6,8 8,0 (MASK) 9,1 (ATIS)	14,3

Figure 8. *Comparaison de la complexité de la tâche considérée avec celle d'autres tâches de même genre*

Nous pouvons remarquer dans le tableau de la figure 8, que la complexité de notre

tâche, par rapport à ARISE, est au niveau du nombre de mots utilisé par requête. Ce nombre élevé est dû principalement aux autocorrections et à la formulation de plusieurs demandes en même temps (généralement deux) dans une même requête.

5.2. Les ensembles de traits sémantiques

Le modèle défini sert à attribuer à chaque mot de l'énoncé transcrit par le module de reconnaissance de la parole un couple de traits sémantiques noté TSE. Chaque couple TSE est constitué de deux traits élémentaires : TSE = (classe sémantique C, trait micro-sémantique TM). Le premier trait sert à déterminer la classe sémantique à laquelle appartient le mot. Par exemple, toutes les villes du réseau ferroviaire sont représentées par la classe sémantique مدينة « medina » (ville). Pour l'application considérée (décrite ci-dessus), nous avons utilisé en tout 12 classes sémantiques C_i (voir figure 9).

Classes sémantiques C_i	Exemples d'instanciations
طلب (Demande)	متى (quand) ; بكم (combien) ; أحب (je veux) ; يوجد (existe) ; etc.
حركة (Mouvement)	الذاهب (qui va) ; اذهب (je vais) ; يصل (arrive) ; etc.
مؤشر_حركة (Indice_mouvement)	إلى (à) ; نحو (vers) ; عبر (à travers) ; من (de) ; etc.
	بتاريخ (à la date) ; الساعة (l'heure) ; etc.
رمز (Référence)	هذا (ce) ; هاته (cette) ; etc.
مدينة (Ville)	تونس (Tunis) ; سوسة (Sousse) ; etc.
ربط (Liaison)	و (et) ; etc.
عدد تذاكر (Nombre billets)	تذكرة (billet) ; مكان (place) ; تذكرتين (deux billets) ; etc.
حس (bruit)	أن (que) ; نهاركم (journée) ; etc.
نوع_التذكرة (Type_billet)	ذهاب (pour étudiants) ; للصحار (pour enfants) ; للطلبة (pour enfants) ; مسترسل (direct) ; وإياب (aller et retour) ; etc.
شرط (Condition)	لا تتجاوز أعمارهم (qui ne dépassent pas l'âge) ; etc.
عدد (Nombre)	2 ; 1 ; etc.

Figure 9. Les classes sémantiques de l'application considérée

Pour extraire les classes sémantiques de l'application, nous avons appliqué l'algorithme des K-means proposé par (McQueen, 1967). Nous avons utilisé

l'information mutuelle moyenne IMm de (Rosenfeld, 1994) au lieu de la distance euclidienne pour mesurer la distance sémantique entre les différents mots du vocabulaire de l'application finalisée. Ceci, nous a amené à remplacer dans l'algorithme le critère d'évaluation $\arg \min_{j=1, \dots, k} d_2(m_i, cg_j)$ par $\arg \max_{j=1, \dots, k} d(m_i, cg_j)$ (voir figure 10). Mis à part le fait que cet algorithme facilite la tâche d'identification des classes sémantiques, il a l'avantage d'être :

- rapide face à des données de taille importante, puisqu'il converge à une vitesse linéaire de l'ordre de $O(n.k.t)$, où n, k et t désignent respectivement le nombre des mots à classer, le nombre des classes sémantiques et le nombre d'itérations maximales ;
- simple à implémenter.

Présentation de l'algorithme des k-means

Choisir d'une manière arbitraire les centres de gravité ($cg_1, cg_2, cg_3, \dots, cg_k$) des k classes sémantiques ($cs_1, cs_2, cs_3, \dots, cs_k$).

Début ; Étiquette :

Pour tout mot m_i de m_1 à m_n faire

Chercher la classe cs_k du mot m_i en question :

$cs_k = \arg \max_{j=1, \dots, k} d(m_i, cg_j)$;

où, $d(m_i, cg_j) = IMm(m_i, cg_j) = P(m_i, cg_j) \times \text{Log} [P(m_i / cg_j) / P(m_i).P(cg_j)] + P(m_i, cg_j) \times \text{Log} [P(m_i / cg_j) / P(m_i).P(cg_j)] + P(m_i, cg_j) \times \text{Log} [P(m_i / cg_j) / P(m_i).P(cg_j)]$

Recalculer le centre de gravité de la classe cs_k :

$cg_k = 1/N_k \sum_{m_i \in cs_k} m_i$; où N_k désigne dans cet algorithme le nombre de mots dans la classe cs_k .

Fin Pour.

- Arrêt du traitement si les centres de gravité sont inchangés.

- Retourner à *Etiquette* sinon.

Fin

Figure 10. L'algorithme des k-means en utilisant l'IMm comme métrique.

Le problème principal de cette méthode est la dépendance du résultat du classement final des informations données en entrée (les k centres de gravité des k classes sémantiques à déterminer sont choisis d'une manière totalement arbitraire). Cette limite ne pose pas de problèmes pour nous, puisque nous avons utilisé cette méthode rien que pour aider et donner une idée à l'utilisateur (surtout si cet utilisateur n'est pas un expert du domaine) sur la classification possible des mots de l'application d'un point de vue sémantique. Cependant les cartes auto-organisatrices de (Kohonen, 1989) offrent une alternative efficace lorsque l'on cherche de meilleurs résultats de partitionnement (Jamoussi, 2004). Nous avons utilisé le logiciel libre TANAGRA (Rakotomalala, 2005) pour l'extraction des classes. En ce qui concerne le deuxième trait de l'ensemble TSE, nous avons utilisé en tout vingt traits microsémantiques TSM_i , soit presque le double des C_i . Ces traits sont les suivants : عام_طلب (demande_générale) ; ثمن_طلب (demande_prix) ; توقيت_طلب

(demande_horaire); وجهة (destination); انطلاق (départ); عبور (correspondance); لحظة (moment); ساعة (heure); تاريخ (date); يوم (jour); درجة (classe); etc.

5.3 Les mots-clés de l'application

Le tableau de la figure 11 ci-dessous, illustre les mots-clés des différentes classes d'énoncés, identifiés dans le cas d'une application de demandes de renseignements ferroviaires. Par exemple, le bigramme القطار (le train) أريد (je veux) constitue un mot-clé, permettant d'identifier les énoncés de type réservation. En effet, ce bigramme n'est employé par les locuteurs que lorsqu'il s'agit d'une demande de réservation. Cette constatation est déduite après étude du corpus d'apprentissage. Nous avons lemmatisé les mots qui sont dérivés à partir d'une même racine trilitère. Par exemple, l'ensemble des mots suivant: الحجز (la réservation); تحجز (le verbe réserver conjugué à la deuxième personne du singulier); حجز (réservation); أحجز (réserver à la première personne du singulier) sont présentés par la racine حجز («hajaza» qui est l'équivalent en français du verbe réserver à l'infinitif).

Classe d'énoncés Sous-classes	Demande d'horaires (HORAIRE)	
Demande de l'horaire de départ (HorDEPART)	موعدا + سفر (date + verbe voyager lemmatisé); متى+اطلق (quand + partir lemmatisé); موعدا + خرج (date + sortir lemmatisé); موعدا + طلق (date + partir lemmatisé); طلق + ساعة (heure + partir lemmatisé); وقت + خرج (horaire lemmatisé + sortir lemmatisé); موعدا + القطار (horaire + train); موعدا + طلق (horaire lemmatisé + partir lemmatisé)	Mots-clés
Demande de l'horaire d'arrivée (HorARRIVEE)	متى + وصل (quand + arriver lemmatisé); متى + وصل (quand + arriver lemmatisé); ساعة + وصل (heure + arriver lemmatisé); وقت + يأتي (horaire lemmatisé + venir lemmatisé); وقت + وصل (horaire lemmatisé + arriver lemmatisé); موعدا + وصل - زمن + وصل (date + arriver lemmatisé)	
Demande des horaires de départ (HorDEPART(S))	أوقات + ذهب (horaires lemmatisé + verbe aller lemmatisé); أوقات + حجز (horaires lemmatisé + verbe réserver lemmatisé); ما هي + أوقات (quelles + horaires lemmatisé); أوقات + القطار (horaires lemmatisé + train)	
Demande de durée du trajet (HorDUREE)	مدة (durée lemmatisé); متى (met); كم+وقت (combien + horaire lemmatisé); تستغرق (passer lemmatisé); مكث (rester lemmatisé)	
Classe d'énoncés	Demande de réservation (RESERVATION)	

	حجز (réserver lemmatisé) ; القطار + أريد (vouloir lemmatisé + train) ; اقتناء (saisie lemmatisé) ; + أريد شراء (vouloir lemmatisé + achat lemmatisé).	Mots-clés
Classe d'énoncés Sous-classes	Demande de tarifs (TARIF)	
Demande de prix (TarPRIX)	سعر (coût lemmatisé) ; تساوي (coûte lemmatisé) ; ثمن (prix lemmatisé) ; بكم (combien) ; تعريفة ; تبلغ ; قيمة (valeur)	Mots-clés
Demande d'informations sur les réductions (TarREDUCTION)	تخفيض (réduction lemmatisé)	

Figure 11. Les mot-clés des classes et sous-classes d'énoncés

Nous pouvons remarquer dans le tableau de la figure 11, qu'il y a en tout trois classes d'énoncés différentes : la classe *HORAIRE* qui regroupe toutes les requêtes de type demande d'horaires ; la classe *RESERVATION* qui regroupe toutes les requêtes de type demande de réservation ; la classe *TARIF* qui regroupe toutes les requêtes de type demande de tarifs.

6. Application du modèle et résultats

Pour tester la performance du modèle numérique défini, 95 % du corpus collecté a été consacré à l'apprentissage et 5 % aux tests (soit 500 énoncés). Le corpus a été préalablement étiqueté et prétraité. Ce prétraitement consiste, comme pour le prétraitement des énoncés, à éliminer les mots vides et à regrouper certains mots en une seule entrée. Le corpus d'apprentissage a été, en plus, annoté manuellement. Cette annotation consiste à étiqueter les énoncés avec les TSE, et à attribuer à chaque demande une étiquette indiquant son type et sa nature illocutoire. Nous avons utilisé les mesures de performances suivantes :

– le taux d'erreur du décodage sémantique : $Taux_{erreur} = N_f / N$; où N est le nombre total de traits attribués à l'énoncé à interpréter, et N_f le nombre de TSE correctement attribués ;

– le taux de précision : $Taux_{precision} = N_v / N$; où N_v est le nombre de TSE correctement attribués.

La figure 12 suivante, présente les taux d'erreur et de précision, trouvés en utilisant le modèle décrit précédemment. Ces taux sont répartis selon le type de renseignement demandé par l'utilisateur : demande de réservation (DR), ou de renseignements sur le trajet (DT), l'horaire (DH), le prix (DP), ou la durée du voyage (DD). Nous pouvons toujours aussi relever le taux d'erreur des énoncés incorrectement décodés sémantiquement, en considérant le rapport entre les énoncés

mal interprétés et le nombre total d'énoncés considérés dans le test (ici 500).

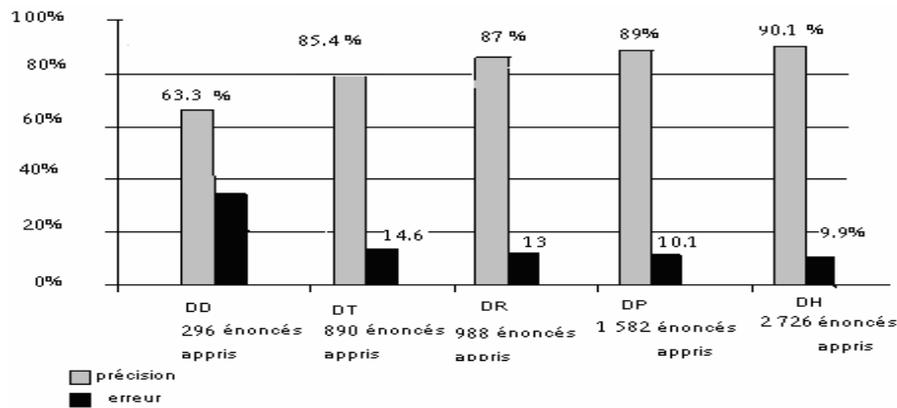


Figure 12. Taux d'erreur et de précision selon le type de la demande de l'utilisateur et le nombre d'énoncés appris

Le taux d'erreur réellement trouvé lors de la mesure de la performance du modèle est de l'ordre de 30,7 %. En analysant davantage les résultats, nous avons conclu qu'un mauvais décodage est obtenu chaque fois qu'il y a un manque de données d'apprentissage. La figure 12 ci-dessus illustre bien ceci. En effet, les résultats de décodage sont bons dans presque tous les types de renseignements demandés par l'utilisateur (DT, DR, DP et DH). Le plus mauvais décodage correspond aux énoncés de type DD. Ceci est dû au fait, que le nombre des énoncés DD considérés lors de la phase d'apprentissage du modèle (3,12 % du corpus) est insuffisant. En effet, nous avons constaté qu'à partir de 1 600 énoncés appris, le modèle numérique devient performant. À partir de ce seuil, le taux d'erreur est inférieur à 24 %. Au-dessous de la barre de 500 énoncés, les résultats deviennent inacceptables. Le taux d'erreur atteint 50 % pour 296 énoncés appris, alors qu'il est de 23,9 % pour 2 726 énoncés appris (voir figure 12 ci-dessus). Une mauvaise interprétation est donc due essentiellement à un manque de données d'apprentissage, et non au type ou à la topologie du modèle numérique défini. En éliminant les énoncés de type DD, la performance de ce dernier passe de 30,7 à 25,8 %. Ces taux sont déjà très satisfaisants, si nous les comparons aux résultats trouvés auparavant : 45 % et 41,2 % (voir figure 13 suivante). Ces résultats sont obtenus lorsqu'on considère une approche standard qui ne tient compte que du contexte (mots) précédent ou suivant pour déterminer le sens d'un énoncé. Dans ce genre d'approches, le contexte pertinent, le type et l'acte illocutoire accompli par l'énoncé ne sont pas généralement considérés.

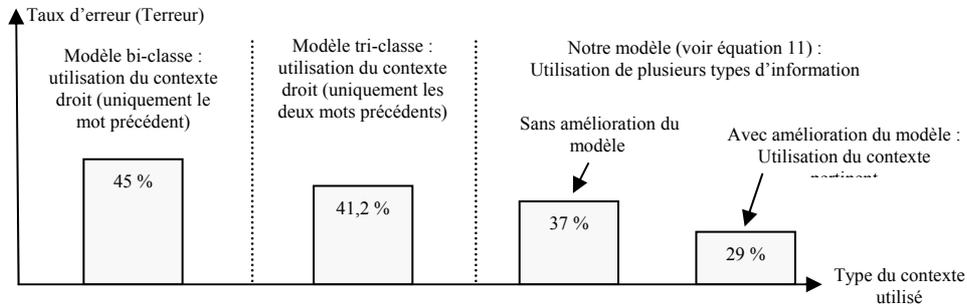


Figure 13. Impact de la considération de plusieurs types d'informations.

Les résultats obtenus, de la variation du taux d'erreur de compréhension en fonction du nombre d'énoncés utilisés pour l'apprentissage, se rapprochent de ceux de (Maynard *et al.*, 2002) lors d'une étude réalisée pour le français dans le même domaine. Selon cette étude, le taux d'erreur se stabilise à partir de 2 000 requêtes (voir figure 14 ci-dessous). Les résultats ont été obtenus avec les transcriptions automatiques des énoncés de test et dans deux conditions d'apprentissage : avec (trait plein) et sans (trait pointillé) classe lexicale pour représenter les villes. Sans la classe lexicale, le double de requêtes (4 000) est nécessaire pour atteindre un niveau de performance équivalent. Au-delà de 11 000 requêtes, l'influence de la classe devient négligeable. En utilisant notre modèle le taux d'erreur se stabilise à partir d'environ 3 955 requêtes = 1 582 énoncés \times 2,5 (d'après la table de la figure 5, un énoncé contient en moyenne entre 2 et 3 requêtes).

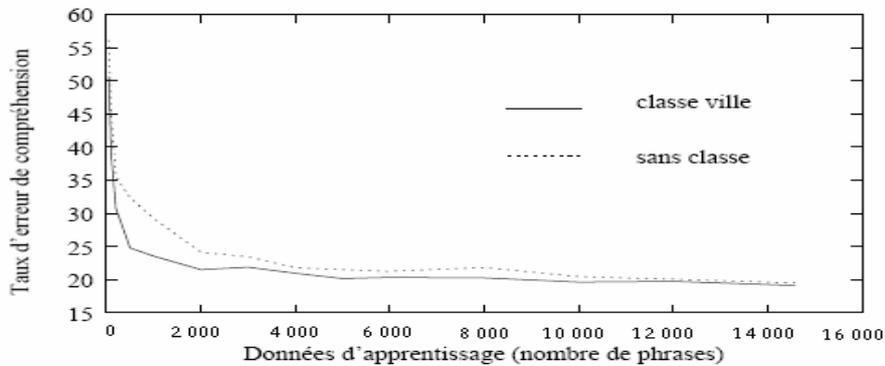


Figure 14. Impact du nombre d'énoncés utilisés pour l'apprentissage sur le taux d'erreur de compréhension (Maynard *et al.*, 2002)

La figure 15 suivante, permet de comparer nos résultats avec des résultats obtenus sur le même genre de tâches par les analyseurs ATIS (Bennacef *et al.*, 1994), MASK (Gauvain *et al.*, 1995) et ARISE (Minker, 1999). Elle indique les erreurs de décodage sémantique fournies au niveau du schéma sémantique par ces analyseurs par méthode probabiliste et par règles.

Systeme / Méthode	ATIS	MASK	ARISE	Notre système
Règles	14,4	13,8	-	-
Probabiliste	13,7	7,2	9,1	9,9 (voir figure 12)

Figure 15. Les taux d'erreurs de décodage sémantique obtenus par d'autres systèmes de compréhension automatique de la parole spontanée

En analysant notre corpus de test, nous avons remarqué que les taux d'erreurs obtenus en employant le modèle numérique présenté précédemment, sont dus principalement :

- aux énoncés ayant une structure syntaxique complexe (énoncés avec plusieurs autocorrections) ;
- aux demandes ne faisant pas partie du domaine de l'application ;
- à l'utilisation de mots qui ne figurent pas dans le vocabulaire de l'application.

Afin de résoudre ce problème, certains systèmes combinent une analyse syntaxique profonde avec une analyse sélective ; c'est le cas pour le système TINA de (Seneff, 1992). D'autres systèmes utilisent les stratégies d'analyses du TAL robuste (Antoine *et al.*, 2003) et (Van Noord *et al.*, 1999). Ces systèmes sont performants dans des applications ouvertes.

7. Conclusion

Nous avons présenté dans cet article un décodeur sémantique basé sur un modèle de langage hybride, qui permet d'intégrer des données contextuelles lexicales, sémantiques ainsi qu'illocutoires en même temps. Il permet, de plus, de ne tenir compte que des ensembles de traits sémantiques TSE pertinents dans l'historique du mot à interpréter. Pour cela, nous avons développé une méthode permettant d'extraire automatiquement ces TSE pertinents qui décrivent les sens des mots ayant une influence sémantique sur le mot à interpréter. Ceci est atteint, en se basant sur la notion d'information mutuelle moyenne de Rosenfeld (Rosenfeld, 1994). Les résultats trouvés sont satisfaisants. Nous comptons par la suite évaluer notre modèle par rapport aux modèles dits distants ou aux modèles obtenus par combinaison linéaire de modèles de langage bien connus comme le maximum d'entropie.

8. Bibliographie

- Aloulou C., Belguith Hadrich L., Ben Hamadou A., « Utilisation des grammaires HPSG pour l'analyse de l'Arabe », *JEI'2002*, Hammamet, Tunisie, 2002.
- Alrahabi M., Mourad G., Djioua B., « Filtrage sémantique de textes en arabe en vue d'un prototype de résumé automatique », *JEP-TALN'04*, Fès, Maroc, 2004.
- Baccour L., Mourad G., Belguith Hadrich L., « Segmentation de textes arabes en phrases basée sur les signes de ponctuation et les mots connecteurs », *troisième journées scientifiques des jeunes chercheurs en génie électrique et informatique*, Mahdia, Tunisie, 2003.
- Bennacef S., Bonneau-Maynard H., Gauvain J-L., Lamel L., Minker W., « A spoken language system for information retrieval », *Actes de International Conference on Spoken Language Processing (ICSLP'94)*, 1994, p. 1271-1274.
- Blachère R., Gaudefroy-Demombynes M., *Grammaire de l'arabe classique*, Maisonneuve & Larose, Paris, 1975.
- Boualem M., Système de conversion de formalismes de langages techniques dans un environnement à syntaxe contrôlée et à contexte limité, Thèse de doctorat, Université de Sophia Antipolis, 1993.
- Bousquet-Vernhettes C., Compréhension robuste de la parole spontanée dans le dialogue oral homme-machine – Décodage conceptuel stochastique, Thèse de l'université de Toulouse III, 2002.
- Chaabene N., Belguith L., « L'étiquetage morpho-syntaxique : comment lever l'ambiguïté dans les textes arabes non voyellés ? », *Proceedings des 3^{ème} journées scientifique des jeunes chercheurs en génie électrique et informatique*, Mahdia, Tunisie, 2003.
- Debili F., Traitement automatique de l'arabe voyellé ou non, Correspondances n°46, Correspondances de l'IRMC, Tunis, 2001.
- Dichy J., Braham A., Ghazali S., Hassoun M., « La base de connaissances linguistiques DIINAR.1, Dictionnaire INformatisé de l'Arabe », *Actes du colloque International Sympotium*, université de Manouba, Tunisie, 2002.
- Gaubert C., Stratégies et règles minimales pour un traitement automatique de l'arabe, Thèse de doctorat, Université Aix-Marseille I, 2001.
- Gauvain J.L., Bennacef S., Devillers L., Lamel L., Rosset S., « Spoken language component of the MASK kiosk », *Human confort & security systems*, édité par Varghese K., Pflieger S., Springer Verlag, 1997, p. 93-103.
- Ghazali S., Habaili H., Zrigui M., « Correspondance graphème phonème pour la synthèse de la parole arabe à partir du texte », IRSIT, *Actes du Congrès dialogue homme machine*, Tunis, 1990.
- Haddad B., Yaseen M., « Towards understanding arabic : A logical approach for semantic representation », *Arabic NLP workshop, ACL'01*, 2001
- Haddad B., Yaseen M., « A Compositional Approach Towards Semantic Representation and Construction of ARABIC », *Proceedings of LACL*, 2005.

- Jaccarini A., Grammaire modulaires de l'arabe. Modélisation, mise en œuvre informatique et stratégies, Thèse de doctorat, Université de Paris IV – Sorbonne, 1997.
- Knight S., Gorell G., Rayner M., Milward D., Koeling R., Lewin I., « Comparing grammar-based and robust approaches to speech understanding : a case study », *Proceedings of European conference on speech communication and technology*, 2001.
- Kurdi M-Z., Contribution à l'analyse du langage oral spontané, Thèse de l'université Joseph-Fourier, Grenoble, 2003.
- Larkey L.S., Ballesteros L., Connell M., « Improving Stemming for Arabic Information Retrieval : Light Stemming and Cooccurrence Analysis », *In Proceedings of the 25th Annual International Conference on Research and Development in Information Retrieval (SIGIR 2002)*, Tampere, Finland, 2002, p. 275-282.
- Lopez P., Roussel D., « Which rules for the robust parsing of spoken utterances with lexicalized tree adjoining grammars ? », *Fourth international workshop on tree adjoining grammars (TAG+4)*, Philadelphie, Etats-Unis, 1998, p. 96-99.
- Mankai Naanaa C., Compréhension automatique de la langue arabe. Application : Le système Al Biruni, Thèse de doctorat de l'université de Tunis II, 1996.
- Maynard H., Lefèvre F., « Apprentissage d'un modèle stochastique de compréhension de la parole », *Actes de JEP'2002*, 2002.
- Meftouh K., Laskri M.T., « Generation of the Sense of a Sentence in Arabic Language with a Connectionist Approach », *Actes de AICCSA*, 2001.
- Minker W., Compréhension automatique de la parole spontanée, L'Harmattan, Paris, 1999.
- Ouersighni R., « A major offshoot of the Dinar-MBC project : AraParse, a morphosyntactic analyzer for unvowelled Arabic texts », *Proceedings of ACL/EACL*, 2001.
- Pepelnjak K., Gros J., Mihelic F., Pavešic N., « Linguistic Analysis in a Slovenian information retrieval system for flight services », *ESCA Workshop on Spoken Dialogue Systems*, Vigso, Danemark, 1995, p. 65-68.
- Rakotomalala R., « TANAGRA : un logiciel gratuit pour l'enseignement et la recherche », *Actes de EGC'2005*, vol. 2, n° 3, 2005, p.697-702.
- Rosenfeld R., Adaptive statistical language modelling : A maximum entropy approach, Thèse de doctorat de l'université de Carnegie Mellon, 1994.
- Sadek D., Bretier P., Cadoret V., Cozannet A., Dupont P., Ferrieux A., Panaget F., « A Cooperative Spoken Dialogue System Based on a Rational Agent Model : A First Implementation on the AGS Application », *ESCA Workshop on Spoken Dialogue Systems*, Vigso, Danemark, 1995, p. 145-148.
- Saïdane T., Zrigui M., Ben Ahmed M., « Arabic speech synthesis using a concatenation of polyphones : the results », *Proceedings of Canadian Conference on AI*, Montréal, Canada, 2005.
- Seneff S., « Robust parsing for spoken language systems », *Actes de ICASSP, 1992*, p.189-192.
- Servan C., Bechet F., « Décodage conceptuel et apprentissage automatique : application au

- corpus de dialogue Homme-Machine MEDIA », *TALN 2006*, Leuven, Belgium, 2006.
- Tuerlinckx L., « La lemmatisation de l'arabe non classique », *JADT'04*, Paris, 2004.
- Van Noord G., Bouma G., Koeling R., Nederhof M.J., Robust grammatical analysis for spoken dialogue systems, Natural Language Engineering, 1999.
- Villaneau J., Ridoux O., Antoine J-Y., « LOGUS : compréhension de l'oral spontané
Présentation et évaluation des bases formelles de LOGUS », *RSTI – RIA*, 2004, p. 709-742.
- Yan P., Zheng F., Xu M., « Robust Parsing in Spoken Dialogue Systems », *European conference on speech communication and technology (EUROSPEECH'01)*, Aalborg, Denmark, Vol. 3, 2001, p. 2149-2152.