

COMPRÉHENSION ET ÉVALUATION DANS LE DOMAINE ATIS

Wolfgang MINKER, Samir BENNACEF

LIMSI-CNRS - BP 133, 91403 Orsay cedex
email: {minker,bennacef}@limsi.fr

ABSTRACT

A spoken language system for vocal access in French to the Air Travel Information Services (ATIS) task has been developed at LIMSI. The semantic analysis in this system is based on a case grammar. We have ported the natural language understanding component to American English using the ARPA ATIS corpus. We discuss the advantages and the performance limitations of the approach when applied to an increased domain coverage and to other languages.

INTRODUCTION

Un système d'interaction verbale dans le domaine ATIS (Air Travel Information Services) permet à son utilisateur d'acquiescer dans des conditions de simulation des informations issues d'un guide officiel des compagnies aériennes américaines et canadiennes [1]. Le domaine ATIS est une des applications choisies par différents laboratoires américains fédérés dans un programme lancé par l'agence ARPA (Advanced Research Project Agency) pour le développement de systèmes du traitement du langage écrit et parlé, pour la collecte de données et la mise à disposition d'outils pour l'évaluation des systèmes.

Un système d'interaction verbale pour le français, L'ATIS, a été développé au LIMSI [2]. Son architecture décrite sur la figure 1 intègre différentes composantes: reconnaissance vocale, compréhension, dialogue ainsi que génération de la requête SQL (System Query Language) et de la réponse du système. Le module de reconnaissance vocale transforme la parole en entrée en une suite de mots qui est transmise à l'analyseur sémantique. Celui-ci détermine la signification de la phrase et établit une représentation sémantique appropriée. La gestion du dialogue complète cette représentation en utilisant l'historique ou en posant des questions à l'utilisateur. Le résultat est utilisé par le module de génération de la requête pour construire une séquence de commandes SQL et extraire les informations de la base de données. La génération de la réponse s'appuie sur les résultats fournis par la base de données et sur la représentation sémantique complétée.

L'analyse sémantique dans L'ATIS est fondée sur la grammaire des cas [3]. Après une présentation de l'originalité de l'approche, nous décrivons les résultats obtenus récemment dans le domaine de la

compréhension, et présentons une étude de flexibilité de la méthode pour l'augmentation de la couverture du domaine ainsi qu'une analyse de portabilité vers d'autres langues. En utilisant le corpus ARPA ATIS [4], l'analyseur sémantique de L'ATIS a été adapté à l'anglais-américain et évalué avec des données de test officielles.

REPRÉSENTATION SÉMANTIQUE

Dans le domaine de la demande d'informations, l'interaction homme-machine est spontanée, ce qui peut se manifester par des répétitions, des hésitations ou des requêtes disloquées (ruptures de construction) qui ne respectent pas la grammaire de l'écrit. L'extraction sémantique ne peut donc pas s'appuyer totalement sur l'analyse syntaxique nécessairement incomplète, mais doit se limiter aux éléments porteurs de sens de la requête tout en ignorant les parties redondantes ou non-essentiels pour l'application. La compréhension dans L'ATIS utilise cette approche fondée sur l'utilisation d'une grammaire des cas permettant de détecter dans chaque requête le ou les concepts liés à l'application et de les valider par un jeu de contraintes.

Tableau 1: Concepts utilisés par l'analyseur sémantique du système français L'ATIS. Les mots-clés soulignés dans les exemples facilitent l'identification du concept.

Concept	Exemple
vol	<i>Je voudrais aller de Oakland à Denver</i>
tarif	<i>Je voudrais les tarifs des vols de Denver à Atlanta</i>
escale	<i>Quel est le lieu de l'escale</i>
type	<i>Quel est le type d'avion pour la compagnie American</i>
réserver	<i>Je choisis le vol numéro 317</i>

Par exemple, dans la phrase *J'aimerais aller de Washington à Boston le cinq juin*, le concept est **vol** et les contraintes sont **ville-départ**, **ville-arrivée** et **date**. D'un point de vue de la grammaire des cas, le concept correspond à la structure casuelle alors que les contraintes correspondent aux cas.

Pour déterminer les concepts et leurs contraintes, un corpus français qui couvre un sous-domaine de l'application a été utilisé [5]. Comme le montre le tableau 1, cinq concepts ont été identifiés à partir de l'analyse manuelle de 655 phrases issues de ce corpus et associés à un ensemble de mots clés tels

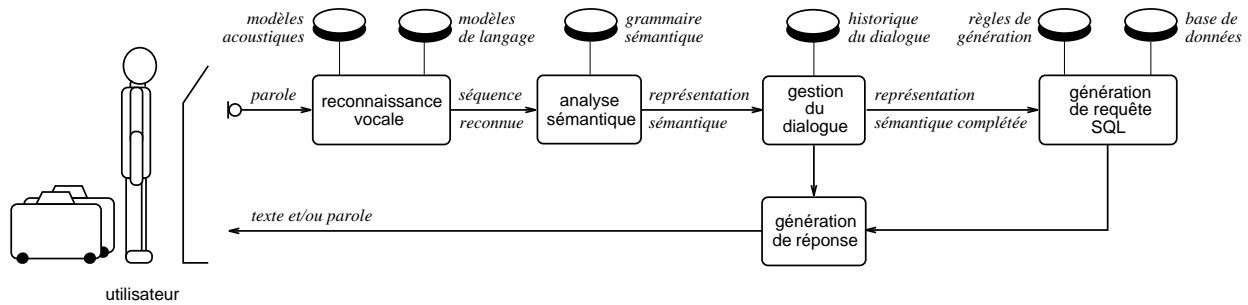


Figure 1: Architecture du système d'interaction verbale L'ATIS.

que ceux soulignés dans les exemples. Un ensemble de 38 contraintes sémantiques représente les catégories d'information, très souvent communes à tous les concepts, par exemple itinéraire, désignation-vol, durée-vol, tarif et ville. Des marqueurs de cas constituent les contraintes syntaxiques nécessaires pour l'extraction d'une représentation sémantique. Dans la phrase *de Washington à Boston* la préposition *de* désigne *Washington* comme une ville de départ et *à* désigne *Boston* comme une ville d'arrivée.

Dans L'ATIS, la grammaire est décrite dans un langage défini dans un fichier déclaratif. Elle contient l'ensemble des structures casuelles et des cas liés à l'application. L'analyse d'une phrase consiste en l'identification d'une structure casuelle pour construire une représentation sémantique sous forme de schémas. Les attributs des schémas sont instanciés à partir de certaines parties de la phrase en utilisant les marqueurs de cas.

La figure 2 montre les structures établies à différents niveaux d'analyse, la représentation sémantique (*RS*), la séquence des commandes SQL (*S*) et la réponse du système (*R*). Pour la requête *RQ*, l'analyseur choisit le concept *vol*, reconnu par le mot clé *aller*. La représentation sémantique *RS* se construit en instanciant les attributs *de*, *à* et *escale-à* avec respectivement *Philadelphie*, *San Francisco* et *Dallas*. La séquence *S* des commandes SQL permet d'accéder à la base de données et d'en extraire l'information. Dans l'exemple, la commande *SELECT from-airport, to-airport, departure-time, arrival-time FROM flight* se construit à partir du concept instancié *vol* de *RS*. Si les attributs *de*, *à* et *escale-à* contiennent des valeurs, les parties *@de*, *@à* et *@escale-à* dans *S* sont remplacées respectivement par les valeurs des attributs *de*, *à* et *escale-à*. Avant d'être présentée à l'utilisateur, l'information est mise en forme et accompagnée d'une réponse *RP* en langage naturel éventuellement synthétisée.

DONNÉES ET ÉVALUATION - UN EFFORT MULTI-SITE

En 1991, le groupe MADCOW (Multi-site ATIS Data Collection Working Group) de coordination pour la collecte multi-site d'un nombre important de données destinées au développement des sys-

tèmes en traitement du langage parlé [4] s'est créé au sein de la communauté ARPA. Le groupe a également élaboré un paradigme d'évaluation standardisée en parole et langage naturel dans le domaine ATIS [6, 1].

Dans la plupart des cas les enregistrements du corpus ont été réalisés de manière classique en utilisant les techniques du Wizard of Oz (Magicien d'Oz) où l'on fait croire aux sujets qu'ils parlent à un système complètement automatique. Des scénarios ont été proposés aux candidats sur la base desquels ils devaient formuler leurs demandes.

Pour la mise en forme, les requêtes ont été orthographiquement transcrites en utilisant des conventions de référence. Lors du processus dit d'annotation, les phrases du corpus ont été étiquetées conformément aux standards proposés par la communauté ARPA [4]. On distingue entre les requêtes de type A (indépendantes du contexte), de type D (dépendantes du contexte) et de type X (qui regroupe les phrases aberrantes). Pour chaque requête de types A et D, des réponses de référence de la base de données ont été établies. Ces références sont utilisées dans une évaluation standardisée [6]. Le corpus pour ATIS consiste en 12 000 requêtes de parole avec des transcriptions orthographiques exactes, des fichiers contenant leurs catégories, les séquences d'entrée du Magicien à NLParse¹, les séquences de commandes SQL pour les réponses de référence ainsi que les références correspondantes. Le corpus a été divisé en deux parties, développement et test. Des tests officiels ont eu lieu entre 1990 et 1994 [7].

Un objectif supplémentaire de MADCOW est la création d'une infrastructure pour l'évaluation automatique des systèmes afin de simplifier le développement et d'offrir l'opportunité de comparer des résultats entre les sites. MADCOW a proposé d'effectuer l'évaluation à partir de la réponse du système plutôt qu'à partir de la représentation sémantique. La méthode consiste à comparer une paire de réponses de référence minimale et maximale construites manuellement au préalable à une hypothèse générée par le système [6, 1].

¹NLParse est un produit de Texas Instruments pour l'accès à une base de données

<i>RQ</i>	<i>Je veux aller de Philadelphie à San Francisco avec escale à Dallas</i>																								
<i>RS</i>	<vol> de: philadelphia à: san-francisco escale-à: dallas																								
<i>S</i>	SELECT airline_code, flight.flight_id, flight.departure_time, flight.arrival_time, stops, stop_airport FROM flight, flight_stop WHERE from-airport=@de AND to-airport=@à AND stop-airport=@escale-à																								
<i>RP</i>	Voici les vols de Philadelphie à San Francisco faisant escale à Dallas																								
	<table border="1"> <thead> <tr> <th><u>COMPAGNIE</u></th> <th><u>NUM_VOL</u></th> <th><u>DÉPART</u></th> <th><u>ARRIVÉE</u></th> <th><u>ESCALES</u></th> <th><u>VILLE_ESCALE</u></th> </tr> </thead> <tbody> <tr> <td>DELTA</td> <td>217/149</td> <td>08h30</td> <td>13h25</td> <td>1</td> <td>DALLAS/FORT-WORTH</td> </tr> <tr> <td>AMERICAN</td> <td>459</td> <td>15h00</td> <td>20h23</td> <td>1</td> <td>DALLAS/FORT-WORTH</td> </tr> <tr> <td>DELTA</td> <td>589/395</td> <td>19h15</td> <td>23h50</td> <td>1</td> <td>DALLAS/FORT-WORTH</td> </tr> </tbody> </table>	<u>COMPAGNIE</u>	<u>NUM_VOL</u>	<u>DÉPART</u>	<u>ARRIVÉE</u>	<u>ESCALES</u>	<u>VILLE_ESCALE</u>	DELTA	217/149	08h30	13h25	1	DALLAS/FORT-WORTH	AMERICAN	459	15h00	20h23	1	DALLAS/FORT-WORTH	DELTA	589/395	19h15	23h50	1	DALLAS/FORT-WORTH
<u>COMPAGNIE</u>	<u>NUM_VOL</u>	<u>DÉPART</u>	<u>ARRIVÉE</u>	<u>ESCALES</u>	<u>VILLE_ESCALE</u>																				
DELTA	217/149	08h30	13h25	1	DALLAS/FORT-WORTH																				
AMERICAN	459	15h00	20h23	1	DALLAS/FORT-WORTH																				
DELTA	589/395	19h15	23h50	1	DALLAS/FORT-WORTH																				

Figure 2: Structures établies dans le système d'interaction verbale L'ATIS. Requête (RQ), représentation sémantique (RS), séquence de commandes SQL (S) et réponse formatée (RP).

L'ATIS EN ANGLAIS-AMÉRICAIN

Adapter la partie compréhension et génération de la requête du système français L'ATIS à l'anglais-américain revient à traduire les fichiers déclaratifs contenant la grammaire des cas et les règles de la génération de requête. La version anglaise a été développée en utilisant les 3 275 requêtes du type A du corpus ATIS [4]. Lors du développement, les règles étaient modifiées, puis l'analyseur sémantique évalué de façon itérative [8]. Pour le français, la collecte de données décrite dans [5] a permis de définir un sous-domaine de l'application en fonction de l'importance accordée par la population visée à un concept sémantique particulier.

Tableau 2: Les concepts utilisés dans le domaine ATIS et quelques mots clés utilisés par l'analyseur sémantique du système anglais de L'ATIS.

<i>Concept</i>	<i>Mots clés</i>
abbreviation	abbreviation, define, explain
quantity	capacity, count, number of
meal	eat, food, meal
ground-service	transportation, ground
airport	airport, airports
airfare	airfare, cost, price, rate, ticket
aircraft	description, kind, type
flight-class	class
restriction	restriction
airline	airline
city	city, where
time-zone	time zone
flight	flight, operate, run, travel, trip

Le concept *time-zone*, par exemple, n'est pas inclus dans le système français car il n'est pas utilisé. Par contre dans le domaine ATIS, tel que défini par ARPA, les données et la couverture du domaine sont imposées. C'est pourquoi des concepts supplémentaires, *meal*, *ground-service*, *time-zone* ont été ajoutés. Le tableau 2 montre les 13 concepts sémantiques accompagnés de quelques mots clés déterminés pour l'anglais. Un total de 69 contraintes représente presque le double par rapport à la version française. Pour la désignation du vol, par exemple, le système français ne contient que les

contraintes numéro du vol et compagnie. Dans la version anglaise, des contraintes supplémentaires existent: *type d'avion* et *capacité*. A partir de la transcription, des marqueurs et des mots-clés sont ajoutés, enlevés ou leur succession est changée dans la grammaire. Par exemple, en français, *midi* et *minuit* ne se comportent pas exactement comme des nombres. Contrairement à l'anglais, ils peuvent être directement suivis de minutes comme dans la phrase *demain midi trente*. C'est ainsi qu'ils servent seulement dans le système français comme des prémarqueurs pour l'attribut *départ-minute*.

PERFORMANCES ET LIMITATIONS

En utilisant une métrique standardisée introduite par la communauté ARPA [6], l'analyseur sémantique anglais de L'ATIS a été évalué sur les données de test officielles en Février 1992. En s'appuyant sur la réponse du système, la méthode consiste à comparer l'hypothèse à une paire de réponses de référence minimale et maximale. On obtient un résultat de 81,8% de réponses correctes.

Le tableau 3 montre des exemples de requêtes qui échouent systématiquement. La suite des symboles entre parenthèses permet de répertorier la requête dans le corpus. La complexité du système, par exemple, devrait augmenter considérablement pour tenir compte des concepts marginaux et pouvoir répondre à la requête *RQ1*. Le système est également incapable d'analyser correctement des requêtes contenant des contraintes multiples ou plus d'une seule catégorie sémantique (*RQ2*, *RQ3*). Le traitement des auto-annulations qui est important dans un système conçu pour des applications réalistes, reste le problème le plus difficile à résoudre. Bien souvent il n'est pas évident de savoir quelle partie sémantique devrait être annulée (*RQ4*). Le problème des répétitions, par contre, est maîtrisable. Dans la plupart des cas, il peut être suffisant de supprimer un des mots identiques (*RQ5*).

Dans le domaine du trafic aérien, les nombres jouent un rôle important. Ils peuvent être des dates, heures, numéros de vols, nombre d'escales ou nom-

Tableau 3: Quelques requêtes problématiques pour l'analyseur sémantique du système anglais de L'ATIS. La suite des symboles entre parenthèses permet de répertorier la requête dans le corpus.

Requête	Type d'erreur
RQ1 <i>What's the next smallest plane after a turboprop</i> (fb0092sx) (Quel est l'avion de taille inférieure à une turboprop)	concept marginal
RQ2 <i>List all the flights from Denver to Pittsburgh and list the fares</i> (b200a1sx) (Montrez tous les vols de Denver à Pittsburgh et montrez les tarifs)	deux concepts
RQ3 <i>Which airline serves Denver Pittsburgh and Atlanta</i> (4a00b1sx) (Quelle compagnie dessert Denver Pittsburgh et Atlanta)	contraintes identiques multiples
RQ4 <i>Show me flights from San Francisco from Pittsburgh to San Francisco on Monday</i> (j50021sx) (Montrez-moi des vols de San Francisco de Pittsburgh à San Francisco lundi)	auto-annulation
RQ5 <i>I need to fly from Dallas to San Francisco and be in San Francisco by four p m</i> (i100a5sx) (Je devrais aller de Dallas à San Francisco et être à San Francisco aux alentours de seize heures)	répétition
RQ6 <i>Show me flights from Baltimore to Philadelphia arriving after twenty one hundred</i> (rc0073sx) (Montrez-moi des vols de Baltimore à Philadelphia arrivant après vingt et une heure)	contraintes faibles pour l'heure

bre de personnes. En anglais, c'est le contexte qui peut être décisif pour attribuer un nombre à une heure (RQ6). Ceci entraîne un risque de confusion, car si l'on instancie un cas, des ambiguïtés surgissent si les marqueurs et/ou les mots clés manquent.

RÉSUMÉ ET PERSPECTIVES

Cet article décrit les résultats de l'adaptation à l'anglais-américain de l'analyseur sémantique développé initialement pour le système d'interaction verbale français L'ATIS. L'analyseur utilise une grammaire des cas afin d'extraire le contenu sémantique d'une requête formulée par l'utilisateur. Cette technique est jugée plus adaptée pour des applications dans le domaine d'interaction homme-machine spontanée, que les méthodes fondées sur une analyse purement syntaxique. L'extraction sémantique s'appuie cependant sur des éléments syntaxiques (marqueurs) ce qui rend la méthode robuste face aux effets du langage parlé ou aux requêtes dont la syntaxe n'est pas connue du système. Les connaissances sémantiques sont acquises au préalable lors de l'analyse manuelle d'un corpus de développement.

L'approche proposée est suffisamment flexible ce qui permet d'augmenter la couverture du domaine en ajoutant des concepts supplémentaires. La grammaire des cas est également portable vers d'autres langues car il suffit de traduire le jeu de règles en tenant compte de quelques spécificités de la langue. Introduite par la communauté ARPA, l'évaluation standardisée a permis de tester de façon itérative les adaptations nécessaires du système et permet donc d'en simplifier le processus. L'analyse de quelques requêtes problématiques donne une idée des limitations de l'approche. L'augmentation de la couverture du domaine améliore les performances du système, mais également sa complexité. Et dans ce cas, si l'on essaie de formaliser la grammaire, celle-ci ne peut jamais tenir compte de tous les concepts sémantiques tels que ceux imaginés par l'utilisateur. L'approche se montre également limitée face aux phénomènes qui ne sont pas accompagnés de contraintes syntaxiques parmi lesquels notamment les annulations.

Une alternative à un système fondé sur des règles peut consister en un apprentissage automatique de concepts à partir d'un grand nombre d'interactions. Un tel système fondé sur un modèle stochastique issu de l'analyse automatique d'un corpus d'apprentissage, peut intégrer les concepts et les contraintes en fonction de leur occurrence. Ce système est en cours de développement.

BIBLIOGRAPHIE

- [1] P. Price. Evaluation of Spoken Language Systems: The ATIS Domain. In *Proceedings of ARPA Workshop on Human Language Technology*, pages 91–95, June 1990.
- [2] S. K. Bennacef, H. Bonneau-Maynard, J. L. Gauvain, L. F. Lamel, and W. Minker. A Spoken Language System For Information Retrieval. In *Proceedings of ICSLP*, pages 1271–1274, September 1994.
- [3] Ch. J. Fillmore. The case for case. *Universals in Linguistic Theory*, pages 1–90, 1968.
- [4] MADCOW. Multi-Site Data Collection for a Spoken Language Corpus. In *Proceedings of DARPA Speech and Natural Language Workshop*, pages 7–14, February 1992.
- [5] H. Bonneau-Maynard, J. L. Gauvain, L. F. Lamel, J. Polifroni, and S. Seneff. A French version of the MIT-ATIS System. In *Proceedings of the European Conference on Speech Technology, EUROSPEECH*, pages 2059–2062, September 1993.
- [6] M. Bates, S. Boisen, and J. Makhoul. Developing an Evaluation Methodology for Spoken Language Systems. In *Proceedings of DARPA Speech and Natural Language Workshop*, pages 102–108, February 1992.
- [7] D. S. Pallett, J. G. Fiscus, W. M. Fisher, J. Garofolo, B. A. Lund, A. Martin, and M. A. Przybocki. 1994 Benchmark Tests for the Spoken Language Program. In *Proceedings of ARPA Workshop on Human Language Technology*, pages 5–36, November 1994.
- [8] W. Minker. An English Version of the LIMSI L'ATIS System. Technical Report 9512, LIMSI-CNRS, April 1995. Notes et Documents LIMSI.