Traitement des mots mal reconnus en compréhension de la parole

Caroline Bousquet-Vernhettes

IRIT
118, route de Narbonne – 31062 Toulouse, France
Tél.: ++33 (0)5 61 55 72 01 - Fax: ++33 (0)5 61 55 62 58
Mél: bousquet@irit.fr - http://www.irit.fr/ACTIVITES/DIAMANT

ABSTRACT

The aim of this paper is to propose an extension of the stochastic conceptual modeling to increase the robustness of the understanding process faced with misrecognitions and unknown words. Corpus analysis shows that some misrecognised words are more difficult to interpret than others, so we defined a *word ambiguity rate*. We performed trial series on train schedule inquiry application to evaluate the understanding rate when faced with misrecognised words and in particular, when these words are city names.

1. INTRODUCTION

L'objectif de cette communication est de proposer un modèle de compréhension robuste face aux erreurs produites par le module de reconnaissance de la parole utilisé en amont. Les deux principales sources d'erreurs de reconnaissance sont souvent dues à des conditions d'usage très différentes des conditions d'apprentissage des modèles acoustiques (environnement bruité, accent, …) et à l'emploi de mots inconnus du lexique de la reconnaissance (couverture lexicale incomplète). Dans le cas général, le mot prononcé est confondu avec un autre mot (ou un groupe de mots) phonétiquement proche. La présence d'un mot mal reconnu en entrée de la compréhension pose un problème d'ambiguïté sémantique : il faut détecter que le mot est mal reconnu pour ne pas l'interpréter tel quel et ensuite il faut trouver le sens de ce mot. De même, le problème posé par la présence d'un mot inconnu revient à identifier son sens alors qu'il est absent des connaissances linguistiques de la compréhension.

Des recherches ont été menées sur l'identification des mots inconnus au niveau du module de reconnaissance (voir entre autre [Gal96], [Baz00]) et sur l'interprétation de ces mots étiquetés comme inconnus (voir [Bor97], [Chu00], [Bou00]). Pour le traitement des mots mal reconnus, l'approche classique consiste à déterminer, en post-traitement de la reconnaissance, si un mot est correctement reconnu ou non en utilisant les taux de confiance donnés par la reconnaissance (voir entre autre [Haz00]).

Nous proposons dans ce papier une modélisation stochastique conceptuelle permettant d'interpréter à la fois les mots identifiés comme inconnus par la reconnaissance et les mots mal reconnus. Certains mots mal reconnus étant bien plus difficiles à interpréter que d'autres, nous avons défini un taux d'ambiguïté des mots. Les résultats sur la compréhension des sorties de reconnaissance d'une part, et sur le cas particulier des villes inconnues et mal reconnues d'autre part, sont présentés.

2. MODÉLISATION STOCHASTIQUE CONCEPTUELLE

2.1 Approche conceptuelle

L'objectif recherché est d'extraire le sens utile des énoncés de l'usager. Nous avons développé un module de compréhension fondé sur une approche stochastique et sur la notion de segments conceptuels [Pie95, Bou00]. Un segment conceptuel est une séquence de mots correspondant à l'unité de base du sens. Nous distinguons trois types de segments conceptuels : les référentiels qui permettent de représenter le domaine d'application, les illocutoires (par exemple *refus*, *demande*…) qui font référence à la théorie des actes de langage et le segment *poubelle*. Ce dernier regroupe tous les mots ou groupes de mots qui sont considérés comme inutiles pour la représentation sémantique de l'énoncé. Ce peut être par exemple, des simples hésitations, des phénomènes extra-linguistiques ou encore des digressions. Dans le cadre d'une tâche finalisée, l'ensemble des segments conceptuels est fini. Ces segments sont déterminés à partir de l'analyse manuelle des corpus de dialogue. Par exemple, les deux séquences de mots « vers quatre heures environ » et « à quatre heures du matin » sont deux instances du même segment conceptuel référentiel *horaire*. Selon ce principe, tout énoncé peut être annoté en une suite de segments conceptuels comme illustré ci-dessous :

J'aimerais un train pour Paris euh disons à neuf heures
Demande destination poubelle horaire

Le langage est modélisé par des chaînes de Markov cachées dont les états sont les segments conceptuels. Chaque segment est représenté par un modèle de Markov dont les états sont des classes de mots. Les classes de mots ont été introduites afin de réduire la complexité du modèle et le nombre de paramètres à estimer. A titre

d'illustration, le segment conceptuel de *ville de destination* est représenté de manière très simplifiée dans la figure 1. Les prépositions de destination sont regroupées dans la classe de mots C1 et tous les noms de villes dans C2. A chacune des classes de mots est associé un ensemble d'unités lexicales. L'union des lexiques de toutes les classes de mots composant le modèle de langage définit le lexique de la compréhension.

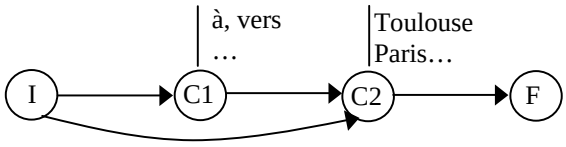

Figure 1 : Exemple de segment conceptuel.

La représentation sémantique d'un énoncé est construite à partir de son découpage en segments conceptuels. Cette représentation sémantique est donnée sous la forme de couples attributs / valeurs, appelés *slots*.

2.2 Objectifs

L'objectif est d'étendre l'approche conceptuelle décrite ci-dessus afin de pouvoir automatiquement détecter et interpréter les erreurs de reconnaissance du type mots mal reconnus au niveau de la compréhension. Cette modélisation doit permettre, en plus du traitement des mots mal reconnus, d'interpréter les mots identifiés comme inconnus.

Nous allons introduire la notion de mots hors-vocabulaire pour une classe de mots du modèle de langage de la compréhension : *un mot est considéré comme hors-vocabulaire pour une classe de mots s'il ne fait pas partie du lexique de cette classe.* Considérons l'énoncé « Je vais à Tours » reconnu « Je vais à jour ». Dans cet exemple, le mot *jour*, bien qu'il soit dans le lexique de la compréhension, est un mot hors-vocabulaire pour la classe de mots contenant les villes. Il serait intéressant de pouvoir identifier que l'usager parle d'une ville d'arrivée que le système n'a pas su reconnaître. Cette indication sur la nature de l'information attendue permettrait d'informer l'utilisateur ce qui n'a pas été compris et ainsi de mieux gérer la poursuite du dialogue. Si au contraire, le mot mal reconnu est en fait inutile pour la compréhension, sa présence ne doit pas pour autant gêner l'interprétation du reste de l'énoncé.

2.3 Modélisations

Le principe de modélisation consiste à ajouter une probabilité d'émission d'un mot hors-vocabulaire dans n'importe quelle classe de mots du modèle de langage. Ces probabilités peuvent être estimées de manière empirique ou bien être apprises automatiquement. Dans cette modélisation, la difficulté est de détecter qu'un mot est mal reconnu et de l'assigner à la bonne classe de mots alors qu'il fait partie du lexique d'une autre classe et qu'il peut être considéré comme mot hors-vocabulaire dans n'importe quelle classe. C'est le contexte apporté par les mots précédents ou suivants de l'énoncé qui permet de pouvoir interpréter correctement le mot mal reconnu. Afin de montrer l'influence de la modélisation proposée sur la compréhension des mots inconnus et mal reconnus, nous considérons trois modélisations :

Modélisation de base : Celle-ci utilise le modèle de base décrit en 2.1. Dans ce cas, le processus de compréhension est incapable d'interpréter un énoncé contenant un mot inconnu et un mot mal reconnu sera interprété tel quel.

Modélisation 'sans pondération' : Toutes les classes de mots ont ici la même probabilité d'émission de mots hors-vocabulaire.

Modélisation 'avec pondération' : Une analyse fine du corpus a permis de déterminer que la majorité des mots inconnus (et donc mal reconnus) sont des villes (environ 54% des cas) ou des mots considérés comme inutiles pour la compréhension (38% des cas). Ce constat nous à conduit à définir une modélisation dont la probabilité d'émission de mots hors-vocabulaire est plus élevée pour les classes de mots du segment conceptuel *poubelle* et pour la classe *ville*.

3. APPRENTISSAGE ET PRINCIPES D'ÉVALUATION

Le domaine d'application concerne la demande d'informations sur les horaires de train de la SNCF. Les corpus d'apprentissage et de tests sont issus de dialogues réels recueillis sur la plate-forme DEMON développé à l'IRIT dans le cadre du projet européen ARISE [Bag99]. Le modèle de langage comporte 29 segments conceptuels et 1220 mots répartis en une centaine de classes de mots. Le segment conceptuel *poubelle* est représenté par une seule classe de mots appelée aussi *poubelle*. L'apprentissage de la modélisation de base a été réalisé sur 4268 transcriptions orthographiques d'énoncés étiquetées manuellement en segment conceptuel. Pour les deux autres modélisations, les probabilités d'émission de mots hors-vocabulaire ont été définies de manière subjective après analyse des corpus. Le taux d'erreur de compréhension est calculé à l'aide de l'algorithme de Levenshtein en comparant les *slots* donnés par le module de compréhension avec ceux donnés par un expert. Le taux d'erreur est obtenu en ajoutant les insertions, les suppressions et les substitutions.

4. ÉVALUATION

L'objectif de cette évaluation est de montrer l'influence du type de modélisation. L'évaluation porte sur 2542 énoncés. L'entrée de la compréhension correspond aux N-meilleures solutions calculées à partir du graphe de mots de la reconnaissance. Nous avons réalisé ce test avec quatre valeurs de N (N = 1, 2, 3 ou 5). Pour ces quatre cas, la figure 2 montre les taux d'erreurs de compréhension obtenus pour les trois modélisations proposées. Le nombre d'erreurs de reconnaissance est important dans ce corpus. Si l'on considère seulement la

meilleure solution donnée par le module de reconnaissance, le taux d'erreur sur les mots est de 40.5% et il y a environ 23% de substitutions.

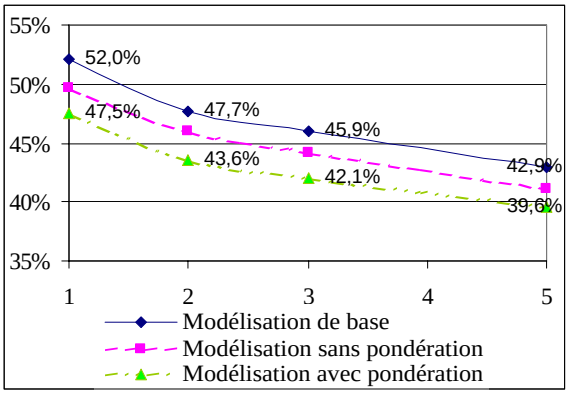

Figure 2 : Taux d'erreur selon le nombre de solutions de la reconnaissance et le type de modélisation.

Nous observons que les résultats les moins bons sont ceux obtenus avec la modélisation de base et les meilleurs sont ceux obtenus avec la modélisation avec pondération. Nous remarquons aussi que plus le nombre de solutions données par la reconnaissance est important, plus la différence entre les trois modélisations est faible (amélioration de 8.6% entre la modélisation de base et celle avec pondération pour N = 1 et 7.7% pour N = 5). Ces résultats sont logiques : plus on considère de solutions, plus il est probable que parmi celles-ci se trouve celle correspondant à l'énoncé. Or on ne peut juger de l'apport du traitement des mots mal reconnus que s'il y a des erreurs de reconnaissance.

5. TEST SUR LES VILLES MAL RECONNUES

Le principal objectif des tests présentés ici est d'évaluer l'apport de notre modèle pour le traitement des villes ou gares inconnues ou mal reconnues.

5.1 Taux d'ambiguïté des mots

Certains mots mal reconnus sont plus difficiles à détecter et à interpréter que d'autres. Cette difficulté est due au degré de confusion avec le mot effectivement reconnu, que nous appelons taux d'ambiguïté : c*e taux noté Θ(m) du mot m correspond à la difficulté d'interpréter correctement le mot m lorsque celui-ci fait référence à un autre mot en raison d'une erreur de reconnaissance de la parole.* Le taux d'ambiguïté d'un mot M est étroitement lié avec le modèle de langage. Soient :

- La probabilité d'émission $E(c_i, M)$ du mot M pour toutes les classes de mots c_i où il n'est pas considéré comme hors-vocabulaire ;
- Pour chaque segment conceptuel SC_j comportant une classe de mots c_i contenant le mot M, la probabilité maximum $T_{SCj}(I, c_i)$ du chemin pour aller de l'état initial I du segment conceptuel SC_j vers la classe c_i ;
- Pour chaque segment conceptuel SC_i qui comporte une classe de mot contenant M, les probabilités de transitions $T(SC_k, SC_i)$ de tous les segments conceptuels SC_k du modèle de langage vers celui-ci.

Le taux d'ambiguïté Θ(M) du mot M est une probabilité (notée sous la forme de pourcentage) calculée par la formule suivante :

$$\theta(M) = \sum_{C_i, SC_j} [E(c_i, M).T_{SC_j}(I, c_i).\max_{SC_k}(T(SC_k, SC_j)]$$

La section suivante montre l'influence du taux d'ambiguïté des mots sur les performances de compréhension.

5.2 Méthodologie de création du corpus de test

Ces tests ont été réalisés sur la transcription orthographique des énoncés. Nous avons retenu les énoncés qui sont correctement compris par notre module de compréhension et qui contiennent un ou plusieurs noms de villes. Ces villes sont alors remplacées par d'autres mots afin de simuler des erreurs de reconnaissance et ce selon leur taux d'ambiguïté. Les erreurs de compréhension seront donc uniquement dues à la présence de ces mots simulant les erreurs et donc considérés comme hors-vocabulaire pour la classe de mot *ville*. Le corpus de test comprend 1477 énoncés dont 1089 ne contiennent qu'une seule ville mal reconnue. Nous avons choisi de faire plusieurs tests en remplaçant à chaque fois le nom de la ville par cinq mots dont le taux d'ambiguïté diffère significativement :

Le mot 'truc' (Θ(truc)=0%) : ce mot n'appartient pas au modèle de langage de la compréhension, il est donc considéré comme hors-vocabulaire pour toutes les classes de mots du modèle. Ce mot est donc considéré comme l'étiquette hors-vocabulaire donnée par un module de reconnaissance [Gal96], [Baz00].

Le mot 'jour' (Θ(jour)=0.87%) : ce mot est utilisé uniquement pour l'expression de la date (ex : « mon jour de départ est le 3 janvier ») et il est peu fréquent.

Le mot 'sept' (Θ(sept)=1.31%) : nous avons choisi un chiffre car les chiffres sont assez ambigus étant donné qu'on les retrouve à la fois pour l'expression de la date (ex : « le 7-07-2001 ») et l'expression de l'horaire.

Le mot 'bon' (Θ(bon)=1.59%) : ce mot est très ambigu car on le retrouve couramment à la fois dans l'expression de réponses affirmatives et négatives (« Oui c'est bon », « non c'est bon ça suffit ») et il est aussi très fréquent comme hésitation (« bon ben disons… »).

Le mot 'alors' (Θ(alors)=5.96%) : ce mot est très fréquemment utilisé comme hésitations et c'est le plus ambigu de ces cinq mots.

5.3 Résultats et discussion

La figure 3 montre les résultats obtenus selon le taux d'ambiguïté et la modélisation choisie.

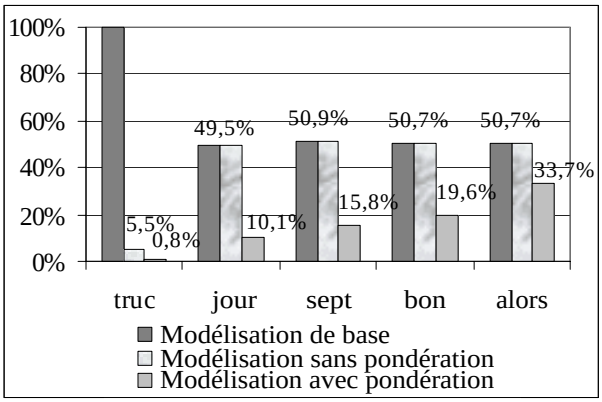

Figure 3 : Test sur les villes mal reconnues.

Les résultats montrent que l'influence de la modélisation sans pondération est nulle pour les mots mal reconnus mais est très élevée pour les villes inconnues de l'application (pour le mot 'truc'). Nous rappelons qu'avec la modélisation de base, il est impossible d'interpréter un énoncé contenant un mot inconnu ce qui explique le taux d'erreur de 100% pour le premier mot. De même, un mot mal reconnu ne sera pas interprété correctement. Cependant, dans le cas de la présence d'un mot mal reconnu, le reste de l'énoncé peut être correctement compris. Nous remarquons que quelque soit le taux d'ambiguïté du mot confondu, les taux d'erreurs de compréhension sont équivalents pour les deux premiers traitements (environ 50% de slots faux pour les quatre dernier mots). La modélisation dite 'sans pondération' n'améliore donc pas les performances pour les villes mal reconnues. En revanche, la troisième modélisation améliore sensiblement les résultats que la ville soit inconnue ou mal reconnue.

Pour la modélisation avec pondération, les performances dépendent fortement du taux d'ambiguïté du mot mal reconnu : le taux d'erreur de compréhension des énoncés passe de 0.8% pour le mot 'truc' (taux d'ambiguïté nul) à 33.7% pour le mot 'alors' (taux d'ambiguïté le plus élevé).

A partir d'une analyse détaillée du type d'erreurs, nous avons remarqué que dans l'ensemble, les erreurs d'interprétation ne se répercutent pas sur le reste de l'énoncé. La grande majorité de ces erreurs concernent les suppressions des slots concernant les villes. Certains slots sont parfois supprimés : ils sont « absorbés » par le segment conceptuel *poubelle*. Le troisième type d'erreur est la prise en compte du mot mal reconnu comme tel, c'est à dire que le processus de compréhension n'a pas su détecter que ce n'était pas le bon mot, ce qui engendre des confusions avec d'autres slots.

6. CONCLUSION ET PERSPECTIVES

Nous avons proposé une extension du modèle stochastique conceptuel pour le traitement des mots mal reconnus. Celui-ci permet aussi d'interpréter les mots étiquetés hors-vocabulaire. Nous avons vu que le traitement des mots mal reconnus est assez délicat : il faut pouvoir tout d'abord identifier qu'un mot est mal reconnu, et donc ne pas l'interpréter tel quel, et ensuite il faut interpréter ce mot correctement. Nous avons remarqué que certains mots sont bien plus faciles à identifier comme étant des mots mal reconnus que d'autres. Ceci nous a amené à définir un taux d'ambiguïté des mots. Les résultats obtenus dépendent directement du taux d'ambiguïté des mots mal reconnus : en effet, plus ce taux est élevé, plus il est difficile d'interpréter correctement le mot et le segment conceptuel qui le comporte. Ce modèle a cependant une limitation majeure : il est impossible de détecter qu'un mot est mal reconnu s'il appartient à la même classe de mot que le mot prononcé (par exemple lorsqu'une ville est confondue avec une autre ville). Une méthode pour lever cette difficulté serait de faire intervenir les scores de confiances données par le module de reconnaissance sur les mots afin de pouvoir détecter que le mot est mal reconnu.

BIBLIOGRAPHIE

[Bag99] Baggia P., Kellner A., Pérennou G., Popovici C., Sturm J. et Wessel F. (1999), "Language Modelling and Spoken Dialogue Systems – the ARISE Experience", EUROSPEECH, Vol. 4, pp. 1767-1770.

[Baz00] Bazzi I. et Glass J.R. (2000), "Modeling Out-of-vocabulary Words for Speech Recognition", ICSLP, Vol. 1, pp. 266-269.

[Bor97] Boros M., Aretoulaki M., Gallwitz F., Nöth E. et Niemann H. (1997), "Semantic Processing of Out-of-vocabulary Words in a Spoken Dialogue System", EUROSPEECH, pp.1887-1890.

[Bou00] Bousquet C., Vigouroux N. et Pérennou G. (2000), "Traitement des mots hors-vocabulaire en compréhension de la parole", JEP, pp. 309-312.

[Chu00] Chung G. (2000), "Automatically incorporating unknown words in JUPITER", ICSLP, Vol. 4, pp. 520-523.

[Gal96] Gallwitz F., Nöth E. et Niemann H. (1996), "A Category Based Approach for Recognition of Out-of-Vocabulary Words", ICSLP, pp. 228-231.

[Haz00] Hazen T.J., Burainek T., Polifroni J. et Seneff S. (2000), "Recognition Confidence Scoring for Use in Speech understanding Systems", Automatic Speech Recognition, pp. 213-220.

[Pie95] Pieraccini R. et Levin E. (1995), "A Spontaneous-Speech Understanding System for Database Query Applications ", ESCA Workshop on Spoken Dialogue Systems, pp. 85-88.