Pour une interlangue utile en traduction automatique de la parole dans des domaines limités

Bouillon, Pierrette; Rayner, Emmanuel; Estrella, Paula; Gerlach, Johanna; Georgescul, Maria

Pour une interlangue utile en traduction automatique de la parole dans des domaines limités

Pierrette Bouillon[*] — **Manny Rayner**[*] — **Paula Estella**[**] — **Johanna Gerlach**[*] — **Maria Georgescul**[*]

** Université de Genève, ETI/TIM – 40, bd du Pont-d'Arve, CH-1211 Genève 4, Suisse*
{Pierrette.Bouillon,Emmanuel.Rayner,Johanna.Gerlach,Maria.Georgescul}@unige.ch

*** FaMAF, U. Nacional de Córdoba – 5000, Córdoba, Argentine*
pestrella@famaf.unc.edu.ar

RÉSUMÉ. *Dans cet article, nous présentons une méthodologie pour construire des interlangues lisibles et vérifiables, qui combinent à la fois expressivité et simplicité. Comme celles-ci sont représentées dans le même formalisme que celui utilisé pour l'analyse syntaxique et la génération, leur bonne formation peut être décrite par des grammaires d'unification, qui permettent naturellement (1) de générer une glose lisible par l'être humain d'une représentation structurée, (2) pour déterminer si cette dernière est bien formée sémantiquement. Bien plus que de simples pivots pour la traduction, les interlangues participent ainsi directement à l'amélioration de la reconnaissance vocale (en filtrant les phrases asémantiques) et du cycle du développement (qui peut se faire de manière entièrement monolingue, via la glose).*

ABSTRACT. *We present a methodology for constructing automatically verifiable and human-readable interlinguas, which simultaneously combine expressivity and simplicity. By defining the interlingua using the same formalism as the one used to define the analysis and generation grammars, it is feasible, in a natural way, both (1) to generate a* **human-readable** *gloss for an interlingua representation, (2) to determine whether said representation is* **well-formed**. *As well as serving as a pivot language for carrying out translation, we show how the interlingua can be used for two other important purposes: improving the quality of recognition, by filtering out hypotheses which are not semantically well-formed, and simplifying the development cycle, which thanks to the gloss form can be transformed into a disjoint set of monolingual tasks.*

MOTS-CLÉS : *traduction automatique de la parole, reconnaissance de la parole fondée sur les grammaires, interlangue, sous-langages.*

KEYWORDS: *speech translation, grammar-based speech recognition, interlingua, sublanguages.*

TAL. Volume 52 – n° 1/2011, pages 133 à 160

1. Introduction

Aujourd'hui, les interlangues sont décriées pour différentes raisons, notamment elles seraient difficiles à définir, maintenir et évaluer objectivement (Blanchon, 2004). Il serait aussi malaisé de garantir à la fois l'expressivité et la simplicité nécessaires : une interlangue trop complexe ne pourrait pas être utilisée de manière systématique ; une interlangue trop simple ne serait pas assez précise pour représenter tous les faits linguistiques et risquerait ainsi d'être peu portable vers de nouvelles applications (Levin *et al.*, 2002 ; Levin *et al.*, 2000). Nous voulons au contraire plaider en faveur d'une approche par interlangue quand d'autres approches ne sont pas possibles. Nous montrons que, pour des systèmes multilingues de traduction automatique (TA) de la parole dans des domaines limités et où la précision est plus importante que le rappel, celle-ci présente des avantages réels, pour autant que deux conditions principales soient remplies :

1) l'interlangue, quelle que soit sa forme structurée, doit être associée à une **glose** facilement lisible par les êtres humains, de manière à pouvoir assurer à la fois expressivité et simplicité ;

2) la bonne formation sémantique de l'interlangue doit pouvoir être vérifiée automatiquement, ce qui permettra de distinguer les phrases sémantiquement bien formées (avec une interlangue correcte) de celles qui ne le sont pas (avec une interlangue incorrecte).

Ces deux propriétés font de l'interlangue un mécanisme essentiel pour un système de traduction de la parole multilingue dans un domaine limité. D'une part, la glose simplifie le cycle de développement du système de TA, qui peut se faire de manière rationnelle, indépendamment de la langue cible traitée, facilitant ainsi l'évaluation des traductions et le développement des règles de mise en correspondance avec l'interlangue. D'autre part, l'information sur la sémanticité des phrases est une source de connaissances très utile, notamment pour aider le reconnaisseur vocal à ne produire que des phrases sémantiquement correctes. Comme le montre (Huet *et al.*, 2006) dans son état de l'art, l'exploitation d'informations de haut niveau sur le langage pour la reconnaissance reste encore très marginale aujourd'hui. Ce travail est ainsi l'une des rares expériences à aller dans ce sens et à en proposer une intégration réussie.

Dans la suite, nous décrivons notre approche de l'interlangue plus en détail, en montrant son apport concret à un système de traduction automatique de la parole dans le domaine médical, MedSLT, notamment au niveau du développement des règles de traduction et de la reconnaissance de la parole. Ce système a été largement décrit dans la littérature ((Bouillon *et al.*, 2008) pour un descriptif récent) ; nous nous contentons donc d'abord d'en rappeler les éléments qui motivent l'approche fondée sur les règles et conduisent à la conception de l'interlangue décrite ici (section 2) ; ensuite, nous focalisons directement sur l'interlangue : nous décrivons le formalisme de représentation (section 3) et la méthodologie adoptée pour la rendre lisible par un être humain et vérifiable automatiquement (section 4). Finalement, nous montrons comment nous exploitons concrètement l'interlangue ainsi développée pour améliorer le

cycle de développement du système (section 5) et la performance, sur le plan de la reconnaissance vocale (section 6) et de la traduction automatique (section 7). L'interlangue nous permet en effet aussi d'envisager un système de traduction hybride qui combine les méthodes linguistique et statistique et augmente ainsi la robustesse, sans perte de précision (section 7). Notre but ici n'est pas de défendre l'interlangue contre d'autres approches (par exemple, statistiques), mais de montrer comment l'exploiter au mieux dans des contextes où elle est motivée (systèmes multilingues, dans des domaines limités où la précision est fondamentale, comme la médecine) et où l'approche statistique semble donner de moins bons résultats (Rayner *et al.*, 2009).

2. MedSLT

MedSLT[1] est un système de traduction automatique de la parole pour le diagnostic d'urgence de patients étrangers. Il permet à un médecin de poser des questions orales à un patient dans des domaines médicaux particuliers, par exemple les maux de gorge ou de tête, et ceci dans différentes langues (pour l'instant l'anglais, le français, le japonais, le catalan, l'espagnol et l'arabe, dans toutes les combinaisons possibles). Suivant la version du système (unidirectionnelle ou bidirectionnelle), le patient répond soit de manière non verbale, soit de manière verbale, par des phrases elliptiques ou complètes, comme dans l'extrait du dialogue en (1), entre un médecin anglais et un patient espagnol[2]. Toute l'architecture du système est donc conçue pour permettre l'adaptation la plus aisée possible à beaucoup de domaines et de langues différentes, tout en répondant à **deux contraintes** au coeur du domaine médical : (a) comment obtenir la précision nécessaire pour le diagnostic, (b) tout en ne donnant pas l'impression à l'utilisateur d'être limité de manière non rationnelle dans son pouvoir expressif ?

1. Ce projet est développé par l'Université de Genève et est financé par le Fonds national de la recherche suisse, SNF.
2. Le site http ://www.issco.unige.ch/en/research/projects/medslt/ donne accès à différentes vidéos du système.

(1) Médecin : Where is your pain ? (“Où est votre douleur ?”)
(Trad. MedSLT : ¿Dónde le duele ?)
Patient : Me duele la garganta (“Me fait-mal la gorge”)
(Trad. MedSLT : I have a sore throat)
Médecin : How long have you had sore throat ?
(“Depuis combien de temps vous avez mal à la gorge ?”)
(Trad. MedSLT : ¿Desde cuándo le duele la garganta ?)
Patient : Cuatro días (“Quatre jours”)
(Trad. MedSLT : I have had a sore throat for four days)
Médecin : Are you allergic to any medicine ? (“Etes vous allergique à certains médicaments ?”)
(Trad. MedSLT : ¿Tiene alergia a alguna medicina ?)
Patient : La penicilina me da alergias (“La péniciline me donne des allergies”)
(Trad. MedSLT : Yes I am allergic to penicillin)
Médecin : Is the pain getting better ? (“Est la douleur devenant mieux ?)”
(Trad. MedSLT : ¿el dolor disminuye ?)
Patient : No sé (“Je ne sais pas”)
(Trad. MedSLT : I don’t know)

Pour répondre à la **première contrainte**, MedSLT est avant tout un système **fondé sur les règles**. La reconnaissance de la parole est linguistique. Elle est contrainte par une grammaire indépendante de contexte, dérivée pour le reconnaisseur *Nuance* ; la traduction suit la méthode interlingue.

Par rapport aux autres systèmes linguistiques récents dans le domaine médical, Nespole ! (Blanchon, 2004) et S-MINDS (Ehsani *et al.*, 2006), MedSLT innove cependant puisque toutes les grammaires du système (pour la reconnaissance, l’analyse et la génération) sont compilées par la plate-forme générique *Regulus* (Rayner *et al.*, 2006), à partir de grammaires générales d’unification, motivées linguistiquement et partagées pour les langues proches (Bouillon *et al.*, 2006b ; Bouillon *et al.*, 2006c). Ceci se fait en deux phases : les grammaires générales sont d’abord spécialisées automatiquement pour des domaines spécifiques par une méthode d’apprentissage fondée sur des corpus, appelée EBL (*Explanation-Based Learning*, (van Harmelen et Bundy, 1988 ; Rayner, 1988)). Cette première étape permet d’obtenir les grammaires les moins ambiguës possibles, mais qui généralisent le mieux un corpus d’apprentissage particulier, comme expliqué en détail dans le chapitre 10 de (Rayner *et al.*, 2006). Ensuite, dans une seconde étape, les grammaires d’unification, ainsi spécialisées par domaine, sont compilées pour les différentes tâches du système : la reconnaissance et l’analyse avec *Nuance* et la génération. Cette approche linguistique, fondée sur les corpus, présente différents avantages : grâce à la spécialisation, le meilleur compromis entre généralité et précision peut être trouvé pour chaque sous-domaine (ici, maux de tête, douleurs abodminales, examen médical, etc.) et locuteur (médecin ou patient). Les différentes grammaires requises par le système deviennent aussi plus faciles à maintenir et restent cohérentes entre elles, notamment toute phrase reconnue sera nécessairement bien analysée, ce qui n’est pas nécessairement le cas avec un reconnaisseur statistique. Il est également possible de produire automatiquement des grammaires qui répondent directement aux spécificités des différentes tâches du système. Par exemple, la gram-

maire de reconnaissance doit évidemment être plus générale que celle de génération pour accepter le plus de variations possibles, alors que la grammaire de génération ne devrait produire que la meilleure variante.

Comme ce type d'architecture fondée sur les règles ne donne des résultats compétitifs que pour les phrases couvertes par le système (Rayner *et al.*, 2005), il convient d'aborder **la seconde contrainte**. Pour aider l'utilisateur à apprendre la couverture du système et à formuler des questions/réponses qui pourront être traitées, nous utilisons un système d'aide (Starlander *et al.*, 2005 ; Chatzichrisafis *et al.*, 2006). Au médecin, celui-ci propose, après chaque question, d'autres phrases similaires, couvertes par le système, dont il pourra s'inspirer ; au patient, il donne aussi des exemples de réponses possibles après chaque question. Pour dériver l'aide, le système se fonde sur un corpus de phrases préenregistrées avec des questions et les réponses liées. Pour les questions, le système fait en parallèle une reconnaissance de type statistique (toujours avec *Nuance*) et compare ensuite le résultat de cette reconnaissance avec les phrases préenregistrées pour extraire les plus similaires (en termes de n-grammes). C'est ainsi que nous introduisons de la robustesse dans un système contrôlé ; pour les réponses, le système d'aide cherche dans le corpus la question la plus similaire à celle du médecin et propose les réponses correspondantes. Les évaluations de Nespole (Blanchon, 2004) concluent que plus de 25 % des interactions étaient hors domaine. Dans MedSLT, il a été montré dans (Chatzichrisafis *et al.*, 2006) que le système d'aide joue un rôle essentiel pour apprendre la couverture aux utilisateurs et augmenter ainsi leur performance avec le système.

L'utilisation du système se fait par une interface de type "cliquer" et "parler", comme dans la plupart des systèmes de ce type (Bouillon *et al.*, 2006a). Pour répondre au besoin de précision nécessaire, la phrase reconnue est toujours rétrotraduite en langue source (LS). Cette rétrotraduction doit ensuite être validée par le médecin pour que la phrase soit traduite pour le patient : ceci permet de vérifier que la phrase ait été bien comprise par le système et d'assurer qu'il n'y aura pas de problème de sens en langue cible (LC). Différentes évaluations de MedSLT ont montré l'intérêt de cette architecture fondée sur les règles pour ce type de domaine où la précision est essentielle, notamment par rapport à une reconnaissance statistique (Rayner *et al.*, 2004) et une traduction statistique (Rayner *et al.*, 2009). Dans la suite, nous tentons de motiver une approche par interlangue. Nous montrons que celle-ci, si elle est bien conçue, ne complexifie pas l'architecture du système, en la rendant peu portable et extensible. Au contraire, elle permet non seulement de simplifier considérablement le développement du système de traduction automatique de la parole, mais aussi d'en améliorer les performances, pour un coût finalement limité dans un système fondé sur les règles.

3. La traduction par interlangue dans MedSLT

La traduction dans MedSLT est assez similaire à ce qui se fait dans le seul autre système récent de TA de la parole par interlangue, à savoir Nespole !³. Elle présente cependant une caractéristique essentielle, illustrée dans la figure 1 : toutes les représentations (source, interlangue et cible) utilisent le même type de formalisme, ce qui aura deux conséquences importantes sur le système en général. Sur le plan informatique, il sera potentiellement possible d'utiliser les mêmes outils à tous les niveaux, que ce soit pour extraire ces représentations, les valider ou générer des phrases/gloses qui en expriment le sens. C'est un point essentiel sur lequel nous reviendrons. Sur le plan linguistique, l'interlangue correspond *mutatis mutandis* à l'une des représentations de l'anglais, ce qui en facilite la définition : il suffit de choisir en anglais la paraphrase dont la représentation servira de forme canonique, en général la plus explicite possible. Par exemple, la question *la douleur est-elle aggravée par le thé ?* est représentée dans un sens qui peut être paraphrasé par "does the pain become worse when you drink tea", qui rend explicite l'action "boire" liée au nom "thé", indispensable par exemple pour la traduction en japonais (figure 2). Le choix de la langue pour l'interlangue n'est plus important ici, puisque celle-ci sera liée à une glose, qui pourra être générée dans n'importe quelle langue (section 4.2).

```
SOURCE: he consultado un médico hace una semana
CIBLE: i visited the doctor one week ago

REPR. SOURCE :  [hace=[number,1], hace=[timeunit,semana],
                 null=[action,consultar], null=[tense,past],
                 null=[utterance_type,dcl], null=[voice,active],
                 obj=[person,médico], subj=[pronoun,yo]]
INTERLANGUE :  [null=[action,visit],
                 arg2=[agent,doctor], arg1=[pronoun,i],
                 ago=[spec,1], ago=[timeunit,week],
                 null=[tense,past], null=[voice,active]]
                 null=[utterance_type,dcl],
REPR. CIBLE :  [null=[action,visit], object=[agent,doctor],
                 agent=[pronoun,i], null=[tense,past],
                 ago=[spec,1], ago=[timeunit,week],
                 null=[utterance_type,dcl], null=[voice,active]]
```

Figure 1. *Étapes de traduction dans MedSLT*

Pour l'analyse, la plate-forme Regulus permet d'obtenir différents types de représentations, plus ou moins complexes, mais dans le cadre de cette application, nous avons choisi le formalisme le plus simple possible, de manière à en facili-

3. http ://nespole.itc.it

```
SOURCE : La douleur est-elle aggravée par le thé
INTERLANGUE : [(when =
                  [clause,
                     [null=[action,drink], arg2=[cause,tea],
                      arg1=[pronoun,you], null=[tense,present],
                      null=[utterance_type,dcl],
                      null=[voice,active]]]),
                null=[event,become_worse], arg1=[symptom,pain],
                null=[tense,present], null=[utterance_type,ynq],
                null=[voice,active]]
```

Figure 2. *Interlangue pour "la douleur est-elle aggravée par le thé ?"*

ter la manipulation/transformation et à rester compatible avec le reconnaisseur statistique. Il s'agit à la base d'une liste de paires attributs-valeurs, avec un niveau d'enchâssement possible pour les phrases subordonnées ou relatives par exemple. Les paires attributs-valeurs représentent les mots au niveau des langues source ou cible (par exemple `[body_part, cou]`) ou le concept au niveau de l'interlangue (`[body_part, neck]`). Il est important de noter pour la suite que ces représentations ne mentionnent pas le type d'article (défini ou indéfini) : ceux-ci sont en effet tellement mal reconnus et difficiles à traduire par des règles générales que nous avons préféré les enlever de la représentation ; c'est la grammaire spécialisée pour la génération qui se chargera ici du choix du meilleur déterminant en fonction du nom (Bouillon *et al.*, 2006b). Pour augmenter l'expressivité des représentations, les paires attributs-valeurs sont aujourd'hui étiquetées au niveau syntaxique avec des fonctions syntaxiques (comme `hace`, `subj`, `obj`) et, au niveau de l'interlangue, avec des rôles sémantiques (`arg2`, `arg1`, etc.). Ce type de représentation, appelé AFF (*Almost Flat Functional Representation* ; (Rayner *et al.*, 2008)), semble le meilleur compromis linguistique et informatique entre, d'une part, une structure entièrement plate dont la surgénération doit être traitée avec des règles de préférences ou des sélections de restrictions *ad hoc* et, d'autre part, des formes logiques complexes comme les QLF (Alshawi, 1992), difficiles à manipuler au niveau de la traduction (Rayner *et al.*, 2000)).

Avec un formalisme de représentation de ce type, la traduction vers l'interlangue et à partir de celle-ci peut s'accomplir de manière efficace : elle consiste simplement à transformer des listes de paires attributs-valeur (étiquetées) dans d'autres listes (étiquetées) similaires. Elle se fait en trois étapes principales, illustrées dans la figure 1 pour la phrase *he consultado un médico hace una semana* ("J'ai consulté un médecin il y a une semaine"). D'abord, la phrase source est analysée avec la grammaire spécialisée pour la reconnaissance/analyse dans la représentation source (`REPR. SOURCE`) ; s'il s'agit d'une phrase elliptique (ce qui n'est pas le cas ici), l'ellipse est résolue de

manière à permettre une traduction en contexte, essentielle dans ce domaine (Bouillon *et al.*, 2007). La représentation source résolue est ensuite convertie dans l'interlangue (`INTERLANGUE`) avec des règles de traduction, puis de l'interlangue vers la représentation de la langue cible (`REPR. CIBLE`). De là sera finalement produite la phrase en langue cible avec la grammaire spécialisée pour la génération.

Puisque les règles de traduction transforment essentiellement des listes, elles devraient *a priori* permettre n'importe quelle transformation de listes en listes, requise par la traduction. Pour ce faire, elles sont de trois types. Les premières permettent de transférer les étiquettes fonctionnelles, par exemple la fonction `obj` dans le rôle sémantique `arg2` en (2) ; les deuxièmes, lexicales, transforment des paires atomiques d'attributs-valeurs, par exemple la représentation du mot espagnol *mover* dans le concept `[action, move]` (3) ; les troisièmes, structurales, manipulent des listes d'attributs-valeurs, par exemple, en (4), elles traduisent la représentation source complexe de *tener tos* (`[[action, tener], obj=[cause, tos]]`) dans la représentation interlingue atomique : `[[action, cough]]`.

(2)
```
bidirectional_role_transfer_rule(arg2, obj).
```

(3)
```
bidirectional_transfer_lexicon([action, move], [action, mover]).
```

(4)
```
bidirectional_transfer_rule([[action, cough]],
  [[action, tener], obj=[cause, tos]]).
```

Toutes les règles peuvent être unidirectionnelles ou bidirectionnelles. Il est aussi possible de spécifier les contextes dans lesquels les règles doivent ou non s'appliquer, ainsi que d'utiliser des macros à l'intérieur des règles ou des variables traductionnelles. Par exemple, la règle suivante (5) s'applique à tous les noms de parties du corps définis dans la macro `ef_body_part`.

(5)
```
bidirectional_transfer_rule([[adj,lower], A],
  [[part,partie_inférieure], de=@ef_body_part(A)]).
```

Même si l'interlangue de MedSLT et les règles de traduction ont été conçues de manière à représenter le meilleur compromis possible entre expressivité et simplicité, elle pose cependant différents problèmes au niveau du développement du système. Il est facile de comprendre pourquoi les interlangues ont été décriées. Quand les phrases sont longues, comme en (6), les représentations deviennent facilement complexes, ce qui ralentit le développement.

(6) SOURCE : ha estado en contacto con alguien que tiene una infección de la garganta por estreptococo

INTERLANGUE :
```
[null=[aspect,perfect],
 arg2 =
 [clause,
   [arg1=[pronoun,rel],
    arg2=[symptom,strep_throat]
    null=[state,have_symptom],
    null=[tense,present],
    null=[voice,active]]]),
 arg2=[head,anyone], arg1=[pronoun,you],
 null=[tense,present],
 null=[utterance_type,ynq],
 null=[verb,be_in_contact_with],
 null=[voice,active]]
```

Il est en effet difficile de voir rapidement si la représentation interlangue est bien formée ou pas ou de trouver pourquoi une traduction échoue, surtout dans le contexte multilingue qui est le nôtre (Dorr *et al.*, 2004).

Dans la suite, nous montrons comment les propriétés de l'interlangue décrites au début de cette section nous ont permis de passer facilement à une interlangue lisible et vérifiable automatiquement. Ces deux caractéristiques vont conduire à une approche beaucoup plus rationnelle de la traduction par interlangue, dont nous pourrons tirer parti à la fois pour le développement des règles (section 5) et la reconnaissance de la parole (section 6).

4. Lisibilité et vérification automatique

Comme l'interlangue est représentée dans le même formalisme que les analyses source et cible, nous avons à notre disposition un moyen très simple, naturel et précis pour vérifier automatiquement la bonne formation des interlangues, sans compliquer toute l'architecture du système : nous pouvons utiliser le même type de grammaire que pour l'analyse et la génération. La seule différence est que cette grammaire ne fait pas le lien entre une représentation source ou cible et la langue naturelle correspondante, comme c'est le cas habituellement : elle prend en entrée une représentation interlangue et produit une glose, comme illustré en (7).

(7) INTERLANGUE :
```
[in_loc=[part,side], arg1=[pronoun,you],
 arg2=[symptom,pain], null=[tense,present],
 null=[state,have_symptom],
 null=[utterance_type,ynq],
 null=[voice,active]]
```

GLOSE : YN-QUESTION you have pain PRESENT ACTIVE in-loc side-part

Nous pouvons ainsi considérer qu'une représentation interlingue est bien formée, s'il est possible d'en générer la glose avec la grammaire d'interlangue. En termes plus techniques, la grammaire d'interlangue est donc une grammaire d'unification, qui, comme toutes les grammaires de ce type, peut être compilée dans un générateur qui prend comme entrée une représentation et produit en sortie une chaîne si la représentation est bien formée, ou échoue si cette dernière est mal formée (Shieber *et al.*, 1990). La grammaire d'interlangue fonctionne donc à la fois comme comme un accepteur (elle détermine quelles sont les représentations interlingues bien formées) et un transducteur (elle transforme une interlangue bien formée en glose).

L'avantage de cette approche est que cette glose, si elle est bien conçue, devient une forme beaucoup plus lisible que l'interlangue elle-même. La grammaire d'interlangue combine donc ici deux fonctions essentielles pour l'application dont nous pourrons tirer parti au niveau du développement et de la reconnaissance : elle permet (1) de générer une glose **lisible** d'une représentation structurée (et *vice versa*) et (2) de déterminer ainsi si cette dernière est **bien formée**. Si la représentation est mal formée, la glose ne pourra donc pas être générée. Ce qui est donc nouveau et intéressant dans cette approche, ce n'est pas tellement la glose elle-même, c'est le fait qu'elle indique la bonne formation de la représentation et qu'il est possible, étant donné une glose, de générer les phrases en langue naturelle qui y correspondent. Dans la suite, nous décrivons d'abord la grammaire d'interlangue plus en détails, et l'évaluons dans ces deux fonctions, avant d'en voir les retombées sur le système.

4.1. *La grammaire d'interlangue*

Le but de la grammaire est ici de décrire la bonne formation de l'interlangue, tout comme la grammaire syntaxique décrivait la bonne formation au niveau syntaxique. Il s'agit en fait d'une grammaire sémantique qui prend en compte des contraintes sémantiques, difficiles à modéliser dans la grammaire syntaxique, sans complexifier les règles avec des contraintes spécifiques au domaine. Cette grammaire a l'avantage d'être beaucoup plus simple que celles utilisées pour l'analyse syntaxique : les mots de la grammaire sont en effet des gloses de concept sans propriétés grammaticales. Les règles lexicales de ce type de grammaire ne font donc qu'associer un concept à une glose, par exemple la règle (8) associe la glose "constipation" au concept de l'interlangue `[symptom, constipation]` :

(8) `basic_symptom:[sem=[[symptom, constipation]]] -->`
`constipation.`

La syntaxe de la glose est toujours la même : les règles de la grammaire vérifient uniquement si la glose est bien formée sémantiquement, c'est-à-dire si elle contient bien tous les éléments nécessaires pour une relation sémantique donnée, par exemple en (9) un concept qui dénote un symptôme peut être un symptôme seul (comme "constipation" plus haut) ou un symptôme avec un qualificatif de symptôme (`symptom_quality`), etc.

(9)
```
symptom:[...] -->
  basic_symptom:[...].

symptom:[...] -->
  symptom_quality:[...],
  basic_symptom:[...].
```

À titre indicatif, l'interlangue de MedSLT couvre actuellement quatre sous-domaines du diagnostic médical, les maux de tête et de gorge ainsi que les douleurs thoraciques et abdominales. La grammaire d'interlangue correspondante contient 494 règles lexicales (ou concepts, par exemple `[symptom, constipation]`). Ces concepts sont regroupés en vingt-neuf catégories de base (par exemple `basic_symptom`), qui forment à leur tour quarante et un nouveaux concepts ou relations (par exemple, `symptom` en (9)). Comme tous les sous-domaines du diagnostic traitent finalement des mêmes concepts (durée d'un symptôme, fréquence, début et fin, etc.), la prise en compte d'un nouveau sous-domaine ne demande aucune modification majeure. Par exemple, l'ajout du domaine des maux de gorge aux trois autres domaines n'a impliqué que l'addition de quelques concepts et relations, notamment la relation `be_in_contact`. Il faut noter que cette grammaire d'interlangue échappe aux critiques traditionnelles des grammaires sémantiques (Ritchie, 1983), puisqu'elle ne décrit ici que la bonne formation des concepts de l'interlangue dans un domaine particulier ; elle offre en revanche le moyen le plus simple pour générer/valider des chaînes de caractères, comme nous allons le voir maintenant.

4.2. *La glose*

Le premier avantage de la grammaire d'interlangue est qu'elle permet d'associer aux phrases une glose, qui est aisément modifiable au niveau de la syntaxe et du vocabulaire grâce à la grammaire d'interlangue et qui devrait être plus facilement lisible que la représentation elle-même. L'idée poursuivie a été de partir d'une glose par défaut, la plus simple possible au début, et d'évaluer son impact sur le développement et la performance du système, avant de l'affiner davantage. Pour les expériences décrites dans la suite, nous avons ainsi choisi un ordre SVO pour tous les types de phrases ; le vocabulaire de la glose est en anglais. Nous indiquons toujours le type de phrase au début de la glose (STATEMENT, WH-QUESTION, YN-QUESTION) et les informations grammaticales sur le temps (PRESENT, PAST) et la voix (ACTIVE, PASSIVE) après le verbe, comme dans les exemples de (10).

(10) Où la douleur irradie-t-elle ?
WH-QUESTION pain radiate PRESENT ACTIVE where
Avez-vous vomi ?
YN-QUESTION you vomit PAST ACTIVE
Avez-vous eu mal cette semaine ?
YN-QUESTION you have PAST ACTIVE pain this-week
Souffrez-vous le matin ?
YN-QUESTION you have PRESENT ACTIVE pain in-time morning
Les mouvements de tête brusques augmentent vos maux de tête ?
YN-QUESTION headache become-worse PRESENT ACTIVE
sc-when [you move-suddenly PRESENT ACTIVE head]
J'ai mal à la gorge.
STATEMENT i have PRESENT ACTIVE pain in-loc throat

Même si ces gloses peuvent encore être clairement améliorées, il est facile de voir qu'elles permettent de parer aux principaux facteurs connus pour compliquer la lecture des interlangues structurées (Dorr *et al.*, 2004). D'une part, l'ordre suit une véritable syntaxe, déterminée par la grammaire d'interlangue, ce qui n'est pas le cas avec la représentation interlingue structurée, dont les éléments apparaissent alphabétiquement. D'autre part, certains éléments peuvent être réunis, comme le démonstratif et le nom (`this-week`), indiquant ainsi un groupe syntaxique. La glose est aussi beaucoup plus compacte et fait abstraction de traits linguistiques parfois complexes comme l'aspect.

Lors de la définition de la glose par défaut, la question était surtout de savoir si celle-ci ne conduit pas à une perte d'informations. L'évaluation effectuée dans (Gerlach, 2008) sur une version légèrement antérieure de la glose confirme que ce n'est pas le cas et que ce type de glose améliore la lisibilité. Dans cette expérience pilote, nous avons demandé à des êtres humains qui n'avaient pas été impliqués dans MedSLT de générer manuellement des phrases à partir des deux formes de l'interlangue (comme proposé par exemple dans la tâche G du projet IAMTC, *Interlingual Annotation of Multiligual text Corpora*, (Belvin, 2004 ; Dorr *et al.*, 2004)). Cette expérience montre que le nombre d'erreurs double quand on part de l'interlangue structurée, plutôt que de la glose. Le temps requis est aussi beaucoup plus court avec la glose (moitié de temps nécessaire), comme on s'y attendait. Nous sommes à présent en train d'affiner cette glose par défaut sur la base des résultats de l'évaluation (par exemple, il semblerait judicieux de mettre le mot interrogatif après l'élément WH-QUESTION et de faire un lien vers le concept UMLS (Lindberg *et al.*, 1993) pour éviter certaines ambiguïtés au niveau du vocabulaire). Ce qui nous importe cependant le plus, c'est de disposer d'une méthodologie générale et efficace pour associer aux phrases qui ont une interlangue correcte une glose, dont la syntaxe précise puisse être modifiée aisément en fonction des besoins et qui puisse être utilisée à la place des représentations structurées pour simplifier le développement du système et améliorer ses performances. La glose est en effet plus lisible qu'une représentation structurée, mais plus facile à produire qu'une traduction puisqu'il n'est pas nécessaire de produire une représentation cible.

4.3. *Vérification de la bonne formation de l'interlangue*

Le deuxième avantage de la grammaire d'interlangue est qu'elle nous permet de vérifier si (1) une phrase a ou pas une bonne interlangue et (2), dans une certaine mesure, si la phrase est correcte sémantiquement. Chacune de ces deux vérifications a évidemment un objectif spécifique et sera utilisée différemment dans la suite. Dans le premier cas, il s'agit principalement d'aider le développeur à l'écriture des règles de traduction : celui-ci doit pouvoir considérer que, lors du développement, une phrase avec une interlangue correcte ne doit pas être modifiée et qu'une phrase sans glose a bien une mauvaise interlangue, qui potentiellement doit être corrigée ; dans le second cas, il s'agit d'identifier les phrases sémantiquement correctes, pour aider, par exemple, la reconnaissance de la parole. La question est donc de savoir si la grammaire d'interlangue remplit bien ces deux tâches, pour anticiper son intérêt pour la reconnaissance et la traduction.

La tâche (1) est moins ambitieuse que la (2) et clairement atteinte – nous ne nous y attarderons d'ailleurs pas trop ici. Pour donner une idée des problèmes possibles à ce niveau, nous avons sélectionné un corpus de 2 017 questions en français, collectées auprès d'étudiants en médecine dans le sous-domaine des maux de tête avec le système MedSLT, lors de simulations (Chatzichrisafis *et al.*, 2006). Le médecin devait diagnostiquer le type de maux de tête dont souffrait le patient standardisé. Celui-ci avait reçu un document avec le type de maladie dont il souffrait et les principaux symptômes. Pour avoir une idée du nombre de faux positifs, nous avons vérifié manuellement si les phrases avec une glose avaient effectivement une interlangue correcte. Sur les 1 595 phrases recevant une glose, 15 ont été considérées comme incorrectes par l'évaluateur humain (0,94 % de faux positifs, 15/1 595). Il s'agit typiquement de phrases ambiguës, qui peuvent potentiellement avoir plusieurs interlangues. Par exemple, la glose de (11) serait correcte pour la phrase "avez-vous mal sur le côté de la tête ?" *versus* "avez-vous mal sur un côté de la tête ?" : elle correspond à la représentation où "un" est considéré comme un déterminant (absent de la représentation, *cf.* section 3) et non un chiffre.

(11) Source : Est-ce que la douleur est d'un côté de la tête ?
Interlangue : YN-QUESTION pain be PRESENT ACTIVE
in-loc head side-part

Il peut aussi s'agir de problèmes de résolution d'ellipses (le système fait une addition à la place d'une substitution, conduisant ainsi à une phrase sémantiquement correcte, qui n'est donc pas rejetée par la grammaire) ou de bogues au niveau du lexique (un mot est associé à un mauvais concept, mais sans conduire à une phrase sémantiquement incorrecte). Il est clair que l'évaluateur humain a pu laisser passer des erreurs lors de la vérification manuelle, mais l'expérience montre, en tout cas, que les résultats de la vérification manuelle sont approximativement les mêmes que ceux de la vérification automatique. Pour vérifier les faux négatifs, nous avons compté dans le même corpus combien de phrases, parmi celles qui ont reçu une analyse syntaxique, mais pas de glose, ont une interlangue bien formée. Sur les 422 phrases sans glose, 409 n'ont pas reçu d'analyse syntaxique et 10 ont été considérées comme correctes

(bien formées) par l'évaluateur humain et devraient être couvertes par la grammaire d'interlangue, ce qui donne un pourcentage de faux négatifs de 10/422 (2,4 %). Le nombre de phrases non analysées peut sembler élevé. Il est pourtant normal dans ce type d'architecture où l'utilisateur est supposé s'adapter à la couverture du système, notamment grâce au système d'aide (section 2). Nous verrons dans la section 7 comment l'interlangue nous permet d'améliorer la robustesse grâce à un système de traduction automatique hybride. Ce qu'il est important de retenir de cette évaluation, c'est qu'il semble donc possible d'arriver à une grammaire d'interlangue avec une couverture suffisante pour identifier les interlangues correctes et incorrectes parmi les phrases analysées par la grammaire syntaxique.

La tâche (2) est plus difficile à évaluer *a priori*. Elle revient à se poser la question suivante : les phrases avec une glose sont-elles toujours sémantiquement correctes ? Et de manière liée : la grammaire d'interlangue permet-elle de distinguer les phrases sémantiquement correctes de celles qui ne le sont pas d'une manière utile pour le reconnaissance ? La réponse est *oui*, dans une certaine mesure. La grammaire d'interlangue reste un compromis entre efficacité et précision. Elle n'est pas capable à elle seule de déterminer si une phrase est pragmatiquement correcte, mais elle identifie dans beaucoup de cas les phrases incorrectes.

Pour vérifier dans quelle mesure l'interlangue permet de trier les phrases correctes et incorrectes, nous avons généré aléatoirement avec la grammaire de reconnaissance dans le domaine du diagnostic des maux de tête deux ensembles de 500 phrases. Le premier contient toutes les phrases produites par la grammaire de reconnaissance/analyse, avec ou sans glose ; le second ne contient que des phrases filtrées, qui ont reçu une glose.

Questions de diagnostic	Jugement
Des crises vous irradient-t-elles sous le côté gauche ?	incorrect
La fréquence des douleurs brûlantes débute-t-elle ?	incorrect
Avez-vous mal quand vous avez des massages depuis plus de septante trois mois ?	incorrect
La douleur est-elle plus fréquente à la maison ?	correct
Quand le cholestérol vous atteint-t-il souvent ?	incorrect
Quelle est la durée quand vous fumez le gros repas ?	incorrect
Avez-vous moins connaissance plus de vingt et un heures ?	incorrect
Vos maux de tête sont-ils précédés par votre état de somnolence ?	incorrect
Des maux de tête irradient-ils la tête ?	incorrect
Vos maux de tête sont-ils lancinants ?	correct
Vos maux de tête serrent-ils quand vous avez mal quand vous marchez ?	incorrect
Quand vous mangez des noix ou des noisettes avez-vous déjà mal lorsque vous avez mal habituellement chaque jour ?	incorrect

Tableau 1. *Phrases générées automatiquement et non filtrées*

Questions de diagnostic	Jugement
Avez-vous mal après de l'alcool ?	correct
Mangez-vous pendant plus de six mois ?	incorrect
Vos maux de tête sont-ils habituellement précédés par un traumatisme ?	correct
Les frissons peuvent-ils vous améliorer habituellement ?	incorrect
Avez-vous mal à la maison ?	correct
La douleur irradie la colonne vertébrale ?	correct
Avez-vous des maux de tête ?	correct
Les mouvements brusques soulagent-ils les maux de tête ?	correct
La douleur est-elle causée par une lumière forte ?	correct
Les maux de tête sont-ils déclenchés par une activité physique ?	correct
Vos maux de tête sont-ils accompagnés par des troubles de vision ?	correct
Un état dépressif soulage-t-il vos maux de tête ?	correct

Tableau 2. *Phrases générées automatiquement et filtrées*

Nous avons ensuite vérifié combien de phrases sont sémantiquement correctes dans les deux cas. Pour ce faire, nous avons demandé à deux personnes de marquer les phrases interrogatives asémantiques (auxquelles il leur aurait été impossible de répondre en tant que patient) et de se mettre d'accord sur les éventuelles différences de jugement. Les tableaux 1 et 2 reprennent les 12 premiers exemples de chacun de ces deux types de phrases, avec les jugements. Les résultats résumés dans le tableau 3 montrent l'intérêt potentiel de la grammaire d'interlangue pour filtrer les résultats de la reconnaisance vocale (section 6).

	Phrases asémantiques
Ensemble filtré	22/500
Ensemble non filtré	204/500

Tableau 3. *Nombre de phrases asémantiques*

Dans la suite, nous voyons comment nous pouvons tirer parti d'une interlangue lisible et vérifiable dans MedSLT lors du développement des règles de traduction, ainsi que pour améliorer la reconnaissance de la parole et la traduction automatique.

5. Développement des règles de traduction

Même si l'approche par interlangue est conçue théoriquement pour faciliter le développement des règles bilingues dans des systèmes multilingues, celui-ci pose très vite différents problèmes pratiques. Bien sûr, comme souvent mentionné dans les ouvrages d'introduction à la TA (Arnold *et al.*, 1994), l'interlangue permet à chaque développeur d'écrire les règles depuis et vers l'interlangue, sans se préoccuper de la

langue cible, mais la vérification de la traduction ne peut se faire sans passer par une paire de langues données. On vérifie donc la traduction dans une paire de langues de référence (souvent la langue source vers la langue source ou la langue source vers une langue commune, comme l'anglais) et on suppose ensuite que si la traduction fonctionne pour cette paire, ceci devrait être vrai aussi dans les autres paires de langues puisque, par définition, l'interlangue est la même. Malheureusement, la situation n'est pas aussi simple que cela dans la réalité. D'abord, comme la traduction de la langue source vers la langue source ou vers une langue de référence implique deux ensembles de règles (vers et de l'interlangue), il est difficile pour le développeur de voir rapidement d'où vient le problème (analyse vers l'interlangue ou génération de l'interlangue ?). Il est aussi impossible de garantir que les règles soient complètes pour un domaine donné et qu'elles aient été testées systématiquement. Chaque langue source est en effet l'objet de collectes de données spécifiques, qui peuvent donner lieu à un ensemble d'interlangues différentes en fonction des scénarios de collecte de données et de la culture. Par exemple, il semble commun en Catalogne de demander si les maux de tête sont causés par des charcuteries, ce qui ne se fera certainement pas dans un pays arabophone. Une possibilité serait évidemment de se résoudre à tester toutes les paires de langues du système, mais il est alors nécessaire d'avoir un spécialiste pour chaque paire de langues, ce qu'un système par interlangue devrait justement permettre d'éviter.

Notre solution est plutôt **de tirer parti de l'interlangue lisible et vérifiable**. Puisque chaque représentation interlangue est associée à une glose lisible et que la glose indique normalement une interlangue correcte, il devient en effet possible, comme dans la figure 3, de créer automatiquement un corpus combiné avec toutes les gloses produites par toutes les langues du système, accompagnées de la traduction dans toutes les langues (`to(L)`) et des phrases qui produisent la glose (`from(L)`). Pour l'instant, nous générons un corpus de ce type par sous-domaine, par exemple le corpus combiné pour le domaine des maux de tête contient 1 195 gloses.

Tout le développement des règles s'articule dès lors autour du corpus combiné de gloses, qui évolue en fonction des différentes collectes de données. Chaque développeur teste les règles de traduction pour les gloses/phrases cibles de ce corpus combiné. Comme l'interlangue est lisible par l'être humain, le développement peut maintenant s'accomplir de manière entièrement **monolingue**, de la langue source à la glose et de la glose à langue cible, sans passer par une langue cible de référence. Il est ainsi possible de tester séparément les règles de et vers l'interlangue, ce qui représente un gain de temps considérable. Tous les développeurs travaillent aussi avec le même corpus de test, ce qui garantit la cohérence du système : on est certain que toute interlangue du corpus, quelle que soit la langue qui l'a produite, aura été testée pour la traduction. Finalement, le corpus combiné nous permet de vérifier aisément : (1) si chaque glose est bien générée par toutes langues sources du système, (2) si chaque glose peut être traduite et (3) si les éventuelles variantes proposées pour une langue sont correctes, ce qui est le cas dans la figure 3. Une glose qui ne serait produite que par une seule langue est erronée ou bien indique un problème de couverture. La glose joue donc un rôle fondamental lors du développement du système. Nous allons voir maintenant comment

```
interlingua_item(
  'YN-QUESTION headache become-worse sc-when
[ you drink red-wine PRESENT ACTIVE ] PRESENT ACTIVE',
  [from(ara)-'hal yachtaddou al soudaa indama tachroub al khamr',
   from(eng)-'does red wine make them worse',
   from(eng)-'is the headache worse when you drink red wine',
   from(fre)-'le vin rouge aggrave t il votre mal de tête',
   from(fre)-'vos maux de tête empirent ils quand vous buvez
                du vin rouge',
   from(jap)-'aka wain de zutsu ga hidoku nari masu ka',
   from(jap)-'aka wain wo nomu to zutsu wa hidoku nari masu ka'
   to(ara)-'hal yachtaddou al soudaa indama tachroub al khamr',
   to(eng)-'are the headaches worse when you drink red wine',
   to(fre)-'vos maux de tête empirent-ils quand vous buvez
                du vin rouge',
   to(jap)-'aka wain wo nomu to zutsu wa hidoku nari masu ka',
  ]).
```

Figure 3. *Extrait du corpus combiné de gloses*

elle permet d'améliorer la performance du système, au niveau de la reconnaissance et de la traduction.

6. Reconnaissance de la parole

Un modèle linguistique pour la reconnaissance permet d'obtenir des résultats plus précis qu'un modèle statistique, mais il est loin d'être parfait (Rayner *et al.*, 2006). La grammaire garantit en effet que le reconnaisseur ne prenne en compte que des phrases syntaxiquement correctes et conformes à certaines restrictions de sélection, mais elle ne peut exclure toutes les phrases asémantiques, comme le montrent les exemples de mauvaises reconnaissances en (12).

(12) Phrases prononcée : avez-vous cette douleur tous les jours ?
Phrases reconnue : avez-vous ce repos ?
Phrases prononcée : avez-vous mal au niveau de yeux ?
Phrases reconnue : perdez-vous mal au niveau des yeux ?
Phrases prononcée : la douleur est-elle sur le côté de la tête ?
Phrases reconnue : vous êtes sur le côté ?

Il arrive aussi souvent que le reconnaisseur fournisse comme première hypothèse une phrase elliptique à la place d'une phrase complète, même si celle-ci est impossible dans le contexte, comme en (13) ou que, faute de contexte, une phrase elliptique soit mal reconnue (Bouillon *et al.*, 2007). Dans ces cas, il est difficile de voir quel trait dans la phrase pourrait être utilisé pour améliorer la reconnaissance : les mots ne sont

probablement pas significatifs ; les bigrammes non plus puisque, pour ce domaine, la reconnaissance statistique donne des résultats plus mauvais que la reconnaissance linguistique, au moins pour les données couvertes par la grammaire (Rayner *et al.*, 2004 ; Knight *et al.*, 2001).

(13) Contexte : qu'est-ce qui soulage la douleur ?
Phrase prononcée : est-ce que vous avez mal sous la mâchoire ?
Phrases reconnue : sous la mâchoire ?

Pour éviter ce genre de mauvaise reconnaissance qui conduit à des phrases asémantiques, **nous proposons d'utiliser à nouveau l'interlangue lisible et vérifiable**. Comme celle-ci permet de distinguer les phrases sémantiques de celles qui ne le sont pas, il devient possible de l'utiliser pour améliorer le modèle du langage. Plutôt que de ne prendre en considération que la première hypothèse (*1-best*) fournie par le reconnaisseur, nous allons garder les six meilleures (*6-best*) (ce qui ressort comme le meilleur compromis au niveau de l'efficacité et de la performance) (Rayner *et al.*, 1994). Nous allons ensuite sélectionner la première hypothèse qui conduit à une interlangue correcte. (14) et (15) illustrent cet algorithme. Dans ces deux exemples, le premier résultat (*1-best*) de la reconnaissance ne peut pas être glosé – l'interlangue est incorrecte conformément à la grammaire d'interlangue. C'est donc le deuxième résultat, qui donne lieu à une interlangue correcte, qui sera préféré ici, conduisant dans ces deux cas à une bonne reconnaissance. De la même manière en (16), le premier résultat (*1-best*) est rejeté car l'ellipse, une fois résolue, aboutit à une interlangue incorrecte (qui correspondrait ici à "qu'est-ce qui soulage la douleur sous la mâchoire ?"). C'est donc la troisième hypothèse, correcte ici, qui est sélectionnée. Il est important de noter déjà ici que la méthode proposée ne peut ainsi qu'améliorer les résultats puisque les phrases reconnues sans interlangue correcte ne sont pas traduites par le système.

(14) Phrase prononcée : avez vous mal au niveau des yeux ?
1 : "a elle mal au niveau des yeux"
2 : "avez vous mal au niveau des yeux" (SÉLECTIONNÉ)
3 : "elle a mal au niveau des yeux"
4 : "a elle mal au dessus des yeux"
4 : "a elle mal au dessus des yeux"
5 : "une fois au niveau des yeux"
6 : "a elle mal au fatigue"

(15) Phrase prononcée : la douleur commence-t-elle brusquement ?
1 : "commencent elles brusquement"
2 : "la douleur commence elle brusquement" (SÉLECTIONNÉ)
3 : "commence elle brusquement"
4 : "augmentent elles brusquement"
5 : "la douleur augmente elle brusquement"
6 : "la durée commence t elle brusquement"

(16) Contexte : qu'est-ce qui soulage la douleur ?
Phrase prononcée : est-ce que vous avez mal à la mâchoire ?
1 : "sous la mâchoire"
2 : "vers la mâchoire"
3 : "est-ce que vous avez mal à la mâchoire" (SÉLECTIONNÉ)
4 : "sous la même mâchoire"
5 : "êtes vous vers la mâchoire"
6 : "aux mâchoires"

Pour quantifier l'intérêt de cette approche, nous avons d'abord évalué les résultats de la reconnaissance dans le système MedSLT, avec et sans la prise en compte de l'interlangue. Pour ce faire, nous avons sélectionné des ensembles de données dans trois langues différentes, en anglais, français et japonais. Il s'agit de questions sur les maux de tête, collectées auprès d'étudiants en médecine ou, pour le japonais, auprès de personnes qui ont joué le rôle du médecin, comme décrit en 4.3 pour le français. Nous avons extrait de ces données les phrases couvertes par la grammaire de reconnaissance puisque ce sont celles qui, par définition, sont susceptibles d'être améliorées par une sélection fondée sur l'interlangue. Pour ces phrases, nous avons ensuite calculé, comme on le fait traditionnellement, le taux d'erreurs au niveau des mots (*WER*) et le taux d'erreurs sémantiques, au niveau de l'attribution de l'interlangue (*SemER*) pour trois versions du système : *1-best* (sans la sélection fondée sur l'interlangue), *6-best* (avec la sélection fondée sur l'interlangue) et *oracle* (le système idéal qui prendrait toujours la bonne hypothèse si elle se trouvait dans la liste des six hypothèses fournies par le reconnaisseur). Il faut noter que le *SemER* mesure bien ici la performance réelle du système : il indique le nombre d'énoncés qui n'ont pas reçu une interlangue identique à la référence (transcription de la phrase reconnue) et n'auront donc pas une traduction correcte. Les résultats sont résumés dans les tableaux 4 et 5. Ils montrent clairement que, même si la version oracle ne peut pas être atteinte comme dans les autres expériences de ce type (Rayner *et al.*, 1994), la prise en compte de l'interlangue améliore globablement les résultats, en particulier on peut faire quatre constatations.

1) Au niveau du *SemER*, les résultats sont toujours meilleurs avec la version *6-best* avec une sélection fondée sur l'interlangue que dans la version *1-best*, quelle que soit la langue.

2) Plus les résultats de reconnaissance sont bons avec la version *1-best* et proches de l'oracle, moins l'amélioration est importante avec la version *6-best*. Celle-ci est donc plus significative pour le *SemER* en français, où la différence entre le *1-best* et l'oracle est relativement importante, que pour le japonais (où la différence est faible).

3) Pour le japonais, le *WER* augmente légèrement dans la version *6-best* en raison de quelques phrases bien reconnues, mais qui n'ont pas reçu de représentation interlingue pour diverses raisons (grammaire d'interlangue incomplète ; phrase bien reconnue, mais hors domaine ; erreurs au niveau des règles de traduction). Dans ces cas-là, la version *6-best* rejette en effet la première hypothèse qui est correcte, pour en sélectionner une autre incorrecte, qui augmente ainsi le *WER*. Ce type de problèmes existe bien sûr aussi dans les deux autres langues (comme le montre le taux

de faux négatifs dans l'évaluation dans la section 4.3), mais ils sont compensés par plus d'exemples pour lesquels l'interlangue a un rôle positif. En revanche, nous avons vu que le *SemER* n'augmente pas comme le *WER*, mais diminue légèrement : celui-ci n'est pas affecté de la même manière par les exemples précédents puisque, dans ces cas, les deux versions du système (*1-best* et *6-best*) échouent également et obtiennent donc un même score négatif (0 *versus* 1).

4) Même si le *WER* varie beaucoup en fonction des langues dans la version oracle, le *SemER* reste plus constant, ce qui montre que la performance potentielle du système est assez équivalente dans les trois langues. Les fautes de reconnaissance sont plus nombreuses en français, mais elles n'ont aucun impact sur l'attribution de l'interlangue (et donc la traduction), comme nous l'avions déjà constaté dans (Bouillon *et al.*, 2006c). La plupart des fautes concernent en effet les déterminants, plus nombreux en français que dans les autres langues[4], dont la sémantique ne figure pas dans l'interlangue (*cf.* section 3).

Langue	**Procédure de sélection**	**WER**
Français (1 586 énoncés)	1-best	9,34 %
	Sélection fondée sur l'interlangue (6-best)	8,55 %
	Oracle	5,54 %
Anglais (526 énoncés)	1-best	5,38 %
	Sélection fondée sur l'interlangue (6-best)	5,00 %
	Oracle	2.56 %
Japonais (343 énonces)	1-best	2,28 %
	Sélection fondée sur l'interlangue (6-best)	3,16 %
	Oracle	1,04 %

Tableau 4. *Taux d'erreurs au niveau des mots (WER)*

Après cette première évaluation qui montre l'utilité de l'interlangue pour la reconnaissance, nous avons voulu voir si cette dernière ne pourrait pas remplacer entièrement les contraintes sémantiques (restrictions de sélection) qui figurent actuellement dans la grammaire de reconnaissance (Rayner *et al.*, 2006). Ceci serait un résultat intéressant puisqu'il serait ainsi possible de simplifier les lexiques et, dans une certaine mesure, les grammaires. Grâce à la plate-forme Regulus, nous avons donc produit deux grammaires de reconnaissance spécialisées différentes à partir de la grammaire générale d'unification du français : l'une qui prend en compte les différents traits pour les restrictions de sélection et l'autre, sans ces traits. Nous avons ensuite comparé la performance de ces deux grammaires, dans nos deux configurations *1-best/6-best* et ceci sur le même ensemble de données que celui utilisé plus haut (1 586 énoncés). Les résultats figurent dans le tableau 6. Ils montrent que, dans la version *6-best*, la gram-

4. Le japonais ne possède en effet pas de détermiants, ni de marques de pluriel et de genre.

Langue	Procédure de sélection	SemER
Français (1 586 énoncés)	1-best	12,3 %
	Sélection fondée sur l'interlangue (6-best)	10,1 %
	Oracle	7,3 %
Anglais (526 énoncés)	1-best	15,6 %
	Sélection fondée sur l'interlangue (6-best)	13,5 %
	Oracle	8,7 %
Japonais (343 énoncés)	1-best	8,7 %
	sélection fondée sur l'interlangue (6-best)	7,9 %
	oracle	7,3 %

Tableau 5. *Taux d'erreurs sémantiques (SemER)*

maire **sans les restrictions de sélection** permet d'arriver à des résultats comparables à ceux de la version *1-best* **avec les restrictions de sélection**, comme on pouvait s'y attendre : la sélection fondée sur l'interlangue a le même effet que les restrictions de sélection. En revanche, s'il y a une sélection fondée sur l'interlangue, il semble toujours préférable d'utiliser une grammaire avec des restrictions de sélection : l'examen manuel des données confirme en effet clairement que la liste des *N-best* produite par la grammaire est de meilleure qualité, ce qui permet d'améliorer les résultats globaux.

Une sélection fondée sur l'interlangue ne remplace donc pas les restrictions de sélection, mais est complémentaire. Il est en tout cas indispensable, pour chaque langue, d'avoir une méthodologie pour chercher le meilleur compromis entre ces différentes connaissances, comme celle proposée ici.

Version	Procédure de sélection	SemER
Avec restrictions de sélection	1-best	12,3 %
	Sélection fondée sur l'interlangue (6-best)	10,1 %
	Oracle	7,3 %
Sans restrictions de sélection	1-best	16,6 %
	Sélection fondée sur l'interlangue (6-best)	12,2 %
	Oracle	11,7%

Tableau 6. *Taux d'erreurs sémantiques (SemER) avec et sans restrictions de sélection*

7. Vers un système hybride

Comme l'interlangue permet de distinguer les phrases sémantiquement correctes de celles qui ne le sont pas, il est possible de générer automatiquement, avec la grammaire de reconnaissance, un grand ensemble de données conformes à l'interlangue, qui peuvent être traduites par le système de TA linguistique et qui peuvent ensuite servir à entraîner un système de TA statistique (Dugast *et al.*, 2008). Il est donc intéressant de voir si un système de TA statistique, ainsi entraîné, permet d'améliorer la qualité du système de TA linguistique, voire de le remplacer. L'expérience de Rayner *et al.* (2009) démontrait que ce type de système de TA statistique reste légèrement inférieur au système linguistique, à la fois pour les données couvertes et non couvertes par le système de TA linguistique, dont il n'arrive pas à reproduire la qualité. Tel quel, il ne permet pas d'améliorer la robustesse du système linguistique : il n'est malheureusement pas assez précis pour aboutir à un système de TA hybride où la TA statistique s'appliquerait quand la TA linguistique échoue. Dans cette section, nous montrons comment nous allons exploiter la glose pour améliorer le système hybride. Puisque la glose est un langage comme un autre, mais simplifié, nous allons entraîner le système statistique sur la traduction langue source-glose. De cette manière, nous pourrons gagner en précision de trois manières.

1) Comme la traduction langue source-langue cible est factorisée en deux étapes (de la langue source à la glose et de la glose à la langue cible), il nous sera possible d'utiliser le système statistique pour traduire de la langue source à la langue cible et d'utiliser le système interlingue pour traduire de la glose à la langue cible. De cette manière, nous tirons parti des deux approches : le système statistique ajoute de la robustesse pour l'extraction de l'interlangue, mais une fois l'interlangue trouvée, le système linguistique assure la précision requise.

2) Comme le système statistique produit la glose et qu'il est possible de produire une phrase à partir de la glose, nous pouvons fournir une rétrotraduction à l'utilisateur (voir section 2). Celle-ci permettra de vérifier si la phrase a été bien comprise par le système de TA statistique avant de la faire traduire, comme dans le système interlingue.

3) Nous pourrons finalement recourir à la même approche que pour la reconnaissance de parole. Plutôt que de produire une seule traduction (*1-best*) avec le système statistique, nous allons en produire une liste (*15-best*), parmi laquelle nous allons sélectionner la première traduction avec une interlangue correcte.

Dans la suite, nous montrons l'apport de ces deux méthodes sur les données réelles non couvertes par le système interlingue français-anglais, extraites de l'ensemble des questions utilisées pour l'expérience de la section 4.3 (384 phrases). Nous décrivons d'abord l'expérience, puis les résultats.

7.1. *Expérience*

Pour produire le système statistique, nous avons d'abord généré automatiquement 500 000 phrases en français avec la grammaire de reconnaissance française, que nous avons fait traduire en anglais par le système interlingue. Les 162 564 phrases résultantes, ainsi que leur glose, ont ensuite servi à entraîner deux modèles de traduction statistiques, en utilisant Giza++, Moses et SRILM (Och et Ney, 2000 ; Koehn *et al.*, 2007 ; Stolcke, 2002) :

1) **modèle statistique A :** entraîné avec les traductions français → anglais ;

2) **modèle statistique B :** entraîné avec les traductions français → glose.

Nous avons ensuite testé ces deux modèles de traduction sur les 384 questions non couvertes par le système interlingue dans les trois configurations suivantes :

1) **version 1 :** modèle statistique A ;

2) **version 2 :** modèle statistique B pour la traduction langue source-glose + système interlingue pour la traduction glose-langue cible ;

3) **version 3 :** modèle statistique B pour la traduction langue source-glose + sélection *n-best* + système interlingue pour la traduction glose-langue cible.

L'évaluation s'est faite de la manière suivante : nous avons fait traduire les 384 questions non couvertes par le système interlingue par les trois systèmes et nous avons demandé à deux personnes de juger la rétrotraduction (quand elle existait) et, ensuite, la traduction.

7.2. *Résultats*

Le tableau 7 résume les résultats, avec un système qui ne traduirait que les phrases dont la rétrotraduction a été jugée correcte (**avec une rétrotraduction**) et un autre, qui traduirait toutes les phrases (**sans rétrotraduction**). Nous avons considéré que la rétrotraduction et la traduction étaient correctes si les jugements des deux juges convergeaient (95 % des cas). Dans le cas contraire, les phrases n'ont pas été prises en considération pour l'évaluation. Pour la version 1 du système, il n'y pas pas de résultats avec une rétrotraduction, puisque ce type de système entièrement statistique ne permet pas d'obtenir l'interlangue.

Ces résultats montrent l'importance de la rétrotraduction pour le système statistique. Celle-ci, combinée avec une sélection *n-best*, nous permet d'obtenir un rappel sur les phrases non couvertes par le système interlingue de 17,7 % (pourcentage de phrases traduites), sans détériorer la précision qui reste à 100 % pour les phrases dont la rétrotraduction a été jugée correcte. Comme il s'agit de phrases non couvertes par le système interlingue de base, l'augmentation globale du rappel avec le système hybride pour l'ensemble des phrases couvertes et non couvertes (1 586 + 384 = 1 970) est donc de 68 phrases sur 1 970, soit 3,5 %. En d'autres termes, le système hybride traduit correctement 17,7 % des phrases qui n'avaient pas pu être traduites avec le sys-

Rétrotraduction	Version	Rappel	Précision
Sans rétrotraduction	Version 1	97,2 %	20,9 %
	Version 2	28,5 %	60,7 %
	Version 3	37,9 %	51,1 %
Avec une rétrotraduction	Version 1	–	–
	Version 2	16,0 %	100 %
	Version 3	17,7 %	100 %

Tableau 7. *Évaluation de la traduction*

tème interlingue de base. Sans la glose et la rétrotraduction, la traduction statistique ne serait pas envisageable ici comme complément du système linguistique, quand celui-ci échoue : le rappel certes augmente, mais avec une précision trop faible pour notre tâche. L'interlangue lisible et vérifiable nous permet donc une fois encore d'améliorer la performance globale du système de manière statistiquement significative, sans coût supplémentaire. Pour l'utilisateur du système, la manière dont se fait la traduction n'a pas d'impact : il vérifie la rétrotraduction quelle que soit la manière dont se fait la traduction (*via* le système par interlangue ou hybride). De toute manière, un médecin ne ferait pas confiance à un système qui traduirait sans qu'il puisse en vérifier le sens. C'est d'ailleurs la grande différence entre un système pour lequel la précision est importante et un système généraliste.

8. Conclusion

Dans cet article, nous avons présenté une méthodologie pour construire des interlangues lisibles et vérifiables, qui combinent à la fois expressivité et simplicité. Comme celles-ci sont représentées dans le même formalisme que celui utilisé pour l'analyse et la génération, leur bonne formation peut être décrite par des grammaires d'unification qui permettent naturellement (1) de générer une glose lisible par l'être humain d'une représentation structurée, (2) pour déterminer si cette dernière est bien formée sémantiquement. Bien plus que de simples pivots pour la traduction, les interlangues participent ainsi directement à l'amélioration de la reconnaissance et de la traduction et simplifient le développement du système.

Il est naturel de se demander si ce double mécanisme pour gloser et traduire à partir de l'interlangue est vraiment nécessaire. Une possiblitité serait de se passer de la glose et de vérifier la sémanticité des phrases au niveau de la traduction. Ceci ne s'est jamais fait à notre connaissance, sauf peut-être dans des systèmes fondés sur les connaissances (KBMT), qui n'ont jamais été couplé à la parole. Pour certaines applications dans des domaines plus limités, il devrait en revanche être possible de remplacer la traduction par la glose. Ceci est notamment l'approche utilisée par le formalisme GF (*Grammatical Framework*, (Ranta, 2004)), où la traduction est pro-

duite directement de la syntaxe abstraite. Nous avons également utilisé une approche similaire dans le système d'apprentissage des langues fondé sur les jeux de traduction oraux, CALL-SLT (Rayner *et al.*, 2010). Pour un sous-domaine aussi complexe que Med-SLT, nous ne pensons cependant pas qu'il soit possible d'obtenir une traduction de qualité de cette manière, qui rende compte adéquatement de toutes les différences entre les langues. Pour l'instant, GF n'a été utilisé que pour des domaines très réduits : il sera intéressant de voir comment la plate-forme évolue pour des domaines plus importants.

De ce travail découlent différentes autres perspectives :

- sur le plan de la performance d'abord, les données générées automatiquement peuvent aussi servir à entraîner les reconnaisseurs statistiques (utilisés dans le système d'aide (section 2)). Une première expérience publiée dans (Hockey *et al.*, 2008) montre en effet qu'un reconnaisseur statistique est plus performant s'il est entraîné sur les données filtrées que non filtrées (résultats significatifs à $P = 0,05$). De manière assez intéressante, le reconnaisseur entraîné avec des données filtrées est même légèrement meilleur que celui créé avec les données manuelles qui ont servi à spécialiser les grammaires de reconnaissance (section 2) (mais de manière non significative). Ceci nous permet d'envisager un système hybride qui utiliserait aussi la reconnaissance statistique, quand la linguistique échoue ;

- pour ce qui est du développement, le vérificateur d'interlangue peut se contenter d'indiquer au développeur si une phrase est correcte ou pas, conformément à la grammaire d'interlangue. Il est possible d'aller plus loin et d'aider le développeur à trouver la raison de l'échec. Pour ce faire, nous exploitons encore une fois la grammaire d'interlangue. Nous appliquons à l'interlangue incorrecte des règles qui la transforment dans une interlangue correcte. Celles-ci, successivement, suppriment un à un les différents éléments de l'interlangue incorrecte[5], ajoutent un élément sous-spécifié à l'interlangue ou sous-spécifient l'un des éléments existants. À chaque transformation, le système essaye de générer une glose avec la grammaire d'interlangue. S'il aboutit à une représentation interlangue correcte, un message est produit qui explique la transformation utilisée, par exemple dans la figure 4 : quand on remplace l'élément `in_time=[timeunit,week]` par `in_time=[time_period,day]`, on obtient la glose correcte `yes STATEMENT doctor carry-out-thera-process throat-culture-test in-time the-last-day PAST ACTIVE`. Un premier prototype de ce type est intégré au système MedSLT et est en cours de test : 75 % des messages proposés aideraient le développeur à trouver l'erreur. Par exemple, en lisant le message de la figure 4, le développeur peut comprendre que le mauvais concept `timeunit` a été utilisé à la place de `time_period`.

Nous espérons que ces autres travaux contribueront aussi à motiver une interlangue lisible par l'être humain et vérifiable automatiquement pour la traduction automatique de la parole dans des domaines limités et où la précision est plus importante que le rappel.

5. C'est-à-dire une paire attribut-valeur.

```
Source: sí ha realizado un cultivo la semana pasada
Target: yes the doctor carried out a throat culture test in the last week
interlingua = [(null=[action,carry_out_thera]), (arg1=[agent,doctor]),
               (null=[interjection,yes]), (in_time=[spec,the_last]),
               (null=[tense,past]),
               (arg2=[therapeutic_process,throat_culture_test]),
               (in_time=[timeunit,week]),
               (null=[utterance_type,dcl]), (null=[voice,active])]

interlingua_surface = WARNING: INTERLINGUA REPRESENTATION FAILED
STRUCTURE CHECK.
APPLYING MODIFICATION
(in_time=[timeunit,week]-->in_time=[time_period,day]) GIVES "yes
STATEMENT doctor carry-out-thera-process throat-culture-test in-time
the-last-day PAST ACTIVE
```

Figure 4. *Exemple de correction proposée par le vérificateur d'interlangue*

Remerciements

Le projet MedSLT a été financé par le Fonds national de la recherche suisse, que nous remercions ici.

9. Bibliographie

Alshawi H. (ed.), *The Core Language Engine*, MIT Press, Cambridge, Massachusetts, 1992.

Arnold D., Balkan L., Meijer S., Humphreys R., Sadler L., *Machine Translation : An Introductory Guide*, Blackwell, Oxford, 1994.

Belvin R. S., « On manual generation from an interlingua », *Proceedings of seventh interlingua workshop on Determining Interlingua Utility for Machine Translation*, Washington, DC, 2004.

Blanchon H., « HLT Modules Scalability within the Nespole Project », *Proceedings of ICSLP*, Jeju Island, Korea, 2004.

Bouillon P., Ehsani F., Frederking R., Rayner M. (eds), *Proceedings of the workshop on Medical Speech Translation, HLT/NAACL-06*, New York, New York, 2006a.

Bouillon P., Flores G., Georgescul M., Halimi S., Hockey B., Isahara H., Kanzaki K., Nakao Y., Rayner M., Santaholma M., Starlander M., Tsourakis N., « Many-to-Many Multilingual Medical Speech Translation on a PDA », *Proceedings of The Eighth Conference of the Association for MachineTranslation in the Americas*, Waikiki, Hawaii, 2008.

Bouillon P., Rayner M., Starlander M., Santaholma M., « Les ellipses dans un système de Traduction Automatique de la Parole », *Proceedings of TALN/RECITAL*, Toulouse, France, 2007.

Bouillon P., Rayner M., Vall B. N., Nakao Y., Santaholma M., Starlander M., Chatzichrisafis N., « Une grammaire multilingue partagée pour la traduction automatique de la parole », *Proceedings of TALN 2006*, Leuwen, Belgium, 2006b.

Bouillon P., Rayner M., Vall B. N., Starlander M., Santaholma M., Nakao Y., Chatzichrisafis N., « Une grammaire partagée multi-tâche pour le traitement de la parole : application aux langues romanes », *Traitement automatique des langues*, 2006c.

Chatzichrisafis N., Bouillon P., Rayner M., Santaholma M., Starlander M., Hockey B. A., « Evaluating Task Performance for a Unidirectional Controlled Language Medical Speech Translation System », *Proceedings of the workshop on Medical Speech Translation, HLT/NAACL-06*, New York, NY, 2006.

Dorr B., Green R., Habash N., Levin L., Frawell D., Helmreich S., Hovy E., Miller K., Mitamura T., Rambow O., Reefer F., Siddhrathan A., « IAMTC Report : Background on IAMTC Interlingua and Representational Comparison for Task G », *Proceedings of seventh interlingua workshop on Determining Interlingua Utility for Machine Translation*, Washington, DC, 2004.

Dugast L., Senellart J., Koehn P., « Can we relearn an RBMT system ? », *Proceedings of the Third Workshop on Statistical Machine Translation*, Columbus, Ohio, p. 175-178, 2008.

Ehsani F., Kimzey J., Master D., Sudre K., Park H., « Speech to Speech Translation for Medical Triage in Korean », *Proceedings of the workshop on Medical Speech Translation, HLT/NAACL-06*, New York, NY, USA, 2006.

Gerlach J., *« Les Interlangues en TA : l'exemple de MedSLT »*, Master's thesis, ETI, Université de Genève, 2008.

Hockey B., Rayner M., Christian G., « Training Statistical Language Models from Grammar-Generated Data : A Comparative Case-Study », *Proceedings of the 6th International Conference on Natural Language Processing*, Gothenburg Sweden, 2008.

Huet S., Sébillot P., Gravier G., Utilisation de la linguistique en reconnaissance de la parole : un état de l'art, Technical Report n° 5917, INRIA, France, 2006.

Knight S., Gorrell G., Rayner M., Milward D., Koeling R., Lewin I., « Comparing grammar-based and robust approaches to speech understanding : a case study », *Proceedings of Eurospeech 2001*, Aalborg, Denmark, p. 1779-1782, 2001.

Koehn P., Hoang H., Birch A., Callison-Burch C., Federico M., Bertoldi N., Cowan B., Shen W., Moran C., Zens R. *et al.*, « Moses : Open source toolkit for statistical machine translation », *ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, vol. 45, p. 2, 2007.

Levin L., Gates D., Lavie A., Pianesi F., Wallace D., Watanabe T., Woszczyna M., « Evaluation of a practical interlingua for task-oriented dialogue », *Proceedings of NAACL-ANLP Workshop on Applied interlinguas : practical applications of interlingual approaches to NLP*, Philadelphia, PA, 2002.

Levin L., Gates D., Wallace D., Perterson K., Lavie A., Pianta E., Cattoni R., Mana N., « Balancing Expressiveness and Simplicity in an Interlingua for Task based Dialogue », *Proceedings of ACL 2002 workshop on Speech-to-speech Translation : Algorithms and Systems*, Seattle, WA, 2000.

Lindberg D., Humphreys B., McCray A., « The Unified Medical Language System », *Methods of Information in Medicine*, vol. 32, p. 281-291, 1993.

Och F., Ney H., « Improved statistical alignment models », *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, 2000.

Ranta A., « Grammatical framework », *Journal of Functional Programming*, vol. 14, n° 02, p. 145-189, 2004.

Rayner M., « Applying Explanation-Based Generalization to Natural-Language Processing », *Proceedings of the International Conference on Fifth Generation Computer Systems*, Tokyo, Japan, p. 1267-1274, 1988.

Rayner M., Bouillon P., Bretan I., « Transfer coverage », *in* M. Rayner, D. Carter, P. Bouillon, V. Digalakis, M. Wirén (eds), *The Spoken Language Translator*, Cambridge University Press, 2000.

Rayner M., Bouillon P., Chatzichrisafis N., Hockey B., Santaholma M., Starlander M., Isahara H., Kanzaki K., Nakao Y., « A Methodology for Comparing Grammar-Based and Robust Approaches to Speech Understanding », *Proceedings of the 9th International Conference on Spoken Language Processing (ICSLP)*, Lisboa, Portugal, p. 1103-1107, 2005.

Rayner M., Bouillon P., Hockey B., Chatzichrisafis N., Starlander M., « Comparing Rule-Based and Statistical Approaches to Speech Understanding in a Limited Domain Speech Translation System », *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation*, Baltimore, MD, 2004.

Rayner M., Bouillon P., Hockey B., Nakao Y., « Almost Flat Functional Semantics for Speech Translation », *Proceedings of COLING-2008*, Manchester, England, 2008.

Rayner M., Bouillon P., Tsourakis N., Gerlach J., Georgescul M., Nakao Y., Baur C., « A Multilingual CALL Game Based on Speech Translation », *Proceedings of LREC 2010*, Valetta, Malta, 2010.

Rayner M., Carter D., Digalakis V., Price P., « Combining Knowledge Sources to Reorder N-best Speech Hypothesis Lists », *Proceedings of the workshop on Human Language Technology*, Princeton, NJ, 1994.

Rayner M., Estrella P., Bouillon P., Hockey B., Naka o Y., « Using artificially generated data to evaluate statistical machine translation », *Proceedings of the 2009 Workshop on Grammar Engineering Across Frameworks*, Association for Computational Linguistics, Singapore, p. 54-62, 2009.

Rayner M., Hockey B., Bouillon P., *Putting Linguistics into Speech Recognition : The Regulus Grammar Compiler*, CSLI Press, Chicago, 2006.

Ritchie G., « Semantics in Parsing », *in* M. King (ed.), *Parsing Natural language*, Academic Press, London, p. 199-217, 1983.

Shieber S., van Noord G., Pereira F., Moore R., « Semantic-Head-Driven Generation », *Computational Linguistics*, 1990.

Starlander M., Bouillon P., Chatzichrisafis N., Santaholma M., Rayner M., Hockey B., Isahara H., Kanzaki K., Nakao Y., « Practising Controlled Language through a Help System integrated into the Medical Speech Translation System (MedSLT) », *Proceedings of the MT Summit X*, Phuket, Thailand, 2005.

Stolcke A., « SRILM - an extensible language modeling toolkit », *Seventh International Conference on Spoken Language Processing*, ISCA, 2002.

van Harmelen T., Bundy A., « Explanation-Based Generalization = Partial Evaluation (research note) », *Artificial Intelligence*, vol. 36, p. 401-412, 1988.