
Une grammaire partagée multitâche pour le traitement de la parole : application aux langues romanes

Pierrette Bouillon* — **Manny Rayner*** — **Bruna Novellas***
— **Marianne Starlander*** — **Marianne Santaholma***
— **Yukie Nakao**** — **Nikos Chatzichrisafis***

* *University of Geneva, TIM/ISSCO*
40, bd du Pont-d'Arve, CH-1211 Geneva 4, Switzerland
Pierrette.Bouillon@issco.unige.ch, Emmanuel.Rayner@issco.unige.ch

** *National Institute for Communications Technology*
3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto, Japan 619-0289
yukie-n@khn.nict.go.jp

RÉSUMÉ. Aujourd'hui, l'approche la plus courante en traitement de la parole consiste à combiner un reconnaiseur vocal statistique avec un analyseur sémantique robuste. Pour beaucoup d'applications cependant, les reconnaiseurs linguistiques basés sur les grammaires offrent de nombreux avantages. Dans cet article, nous présentons une méthodologie et un ensemble de logiciels libres (appelé Regulus) pour dériver rapidement des reconnaiseurs linguistiquement motivés à partir d'une grammaire générale partagée pour le catalan, le français et l'espagnol.

ABSTRACT. Today, the most common architecture for speech understanding consists of a combination of statistical recognition and robust semantic analysis. For many applications, however, grammar-based recognisers can offer concrete advantages. In this paper, we present a methodology and an Open Source platform (Regulus), which together permit rapid derivation of linguistically motivated recognisers, in either language, from a bilingual grammar of Catalan, French and Spanish.

MOTS-CLÉS : Traduction automatique de la parole, modélisation du langage, grammaire d'unification, reconnaissance linguistique, grammaires multilingues.

KEYWORDS: Speech to speech translation, language modelling, unification grammar, grammar-based recognition, multilingual grammars.

1. Introduction

Aujourd'hui, l'approche la plus courante en traitement de la parole consiste à combiner un reconnaiseur vocal statistique avec un analyseur sémantique robuste. Pour beaucoup d'applications cependant, les reconnaiseurs linguistiques basés sur les grammaires offrent de nombreux avantages. On sait tout d'abord qu'entraîner un reconnaiseur statistique requiert des données importantes, qui ne sont pas toujours disponibles en début de projet. Certaines comparaisons ont aussi montré que, pour les applications où la précision est plus importante que la robustesse, les reconnaiseurs linguistiques tendent à donner de meilleurs résultats ((Knight *et al.*, 2001) et (Rayner *et al.*, 2005a)). De plus, comme ces reconnaiseurs reposent sur une grammaire qui peut directement être utilisée pour l'analyse, ils évitent par là même le développement d'un analyseur indépendant. La cohérence globale du système est ainsi assurée : si une phrase peut être reconnue, elle sera aussi analysée !

De nos jours, la plupart des plateformes vocales commerciales permettent les deux types de reconnaissance vocale, statistique et linguistique. Malgré les avantages mentionnés plus haut, la préférence va cependant toujours aux reconnaiseurs statistiques. La raison en est simple : les grammaires requises par l'approche linguistique doivent, aujourd'hui encore, être décrites dans le formalisme des grammaires indépendantes de contexte (CFG), voire même parfois dans des formalismes moins expressifs. Or personne n'ignore qu'avec ce type de langage, les grammaires importantes sont difficiles à développer. Même les petites grammaires *ad hoc* pour des domaines limités deviennent vite redondantes et impossibles à maintenir, surtout si elles doivent couvrir des domaines liés.

Pour tirer parti de la précision des reconnaiseurs linguistiques, tout en évitant le formalisme CFG, différentes solutions peuvent être envisagées. La plus répandue consiste à développer des modèles du langage dans des formalismes grammaticaux plus puissants, typiquement celui des grammaires d'unification (GU), et à les compiler ensuite dans le formalisme CFG requis par les reconnaiseurs ((Moore, 1998), (Dowding *et al.*, 2001) et (Bos, 2002)). Regulus est l'un des outils conçus dans ce but. Il s'agit d'une plateforme de logiciels libres ((Regulus, 2006) et (Rayner *et al.*, 2006)), développée dans le cadre du projet MedSLT, un système de traduction automatique de la parole dans le domaine médical, spécialisé pour le diagnostic d'urgence de patients étrangers (Bouillon *et al.*, 2005).

Avec Regulus, nous voulons atteindre le plus haut niveau d'abstraction possible. Notre but est d'avoir une seule grammaire générale, qui pourra ensuite être réutilisée pour différents domaines et tâches, et partagée entre plusieurs langues. Pour un projet de traduction comme MedSLT, le gain est non négligeable. MedSLT utilise en effet un reconnaiseur et un système de traduction linguistiques. Il requiert ainsi une grammaire par sous-tâche (reconnaissance/analyse et génération), de préférence pour chaque sous-domaine du diagnostic (pour l'instant, les maux de tête, douleurs abdominales, douleurs thoraciques et maux de gorge) et, ceci, dans les sept langues du

système (français, catalan, finnois, anglais, japonais, espagnol, arabe). Sans Regulus, nous arriverions ainsi déjà à un total de cinquante-six grammaires différentes !

Dans d'autres articles (Rayner *et al.*, 2006), nous avons déjà décrit comment dériver avec Regulus des reconnaisseurs linguistiques à partir des grammaires générales monolingues. Nous ne nous y attarderons donc plus ici. En résumé, nous partons d'une grammaire d'unification générale, motivée linguistiquement et d'un lexique, qui contient, dans notre cas, le vocabulaire des sous-domaines du diagnostic pour chaque langue. Nous spécialisons ensuite cette grammaire d'unification générale dans des grammaires d'unification spécialisées par domaine avec des méthodes d'apprentissage basées sur la méthode de l'EBL (Explanation Based Learning ; (Rayner *et al.*, 2006, chap. 10)). Cette méthode est paramétrée : 1) par un corpus du domaine d'où le système va apprendre le vocabulaire et les types de phrases qui doivent figurer dans la grammaire spécialisée et 2) par des critères d'opérationnalité (« operatinality criteria ») qui contrôlent la granularité de la grammaire spécialisée (à quelles généralisations la grammaire doit-elle aboutir à partir du corpus ?). Ces grammaires d'unification spécialisées sont ensuite compilées dans le format CFG, requis par notre plateforme vocale (Nuance, pour le projet MedSLT). Grâce à cette méthodologie, nous combinons la précision et la performance des systèmes *ad hoc*, développés pour un domaine spécifique, avec tous les avantages bien connus des grammaires générales : facilité de maintenance et de développement.

Dans cet article, nous montrons comment pousser encore plus loin le degré d'abstraction. Dans la lignée des travaux de (Bender *et al.*, 2002) et de (Kim *et al.*, 2003), nous proposons d'écrire une grammaire d'unification partagée pour différentes langues. La difficulté dans ce contexte particulier est que cette grammaire doit pouvoir être compilée efficacement dans des reconnaisseurs pour les différentes langues. Pour cette première expérience, nous travaillons avec trois langues très proches (le français, le catalan et l'espagnol), mais il ne s'agit ici que du préalable obligé d'une recherche plus ambitieuse dont le but est de développer une méthodologie générale pour construire des grammaires partagées pour la parole. Dans la suite, nous présentons d'abord les spécificités des grammaires Regulus, par rapport à des grammaires conçues pour l'écrit et l'analyse syntaxique. Nous décrivons ensuite la grammaire générale partagée pour les trois langues, puis nous montrons comment cette grammaire partagée est spécialisée pour les trois tâches de reconnaissance, d'analyse et de génération. Dans la dernière section, nous évaluons cette grammaire pour ces trois tâches.

2. Les grammaires d'unification Regulus : spécificités

Avec Regulus, nous voulons arriver à une seule grammaire d'unification générale, motivée linguistiquement, qui sera compilée pour la reconnaissance, l'analyse et la génération. Cette grammaire générale doit ainsi combiner différentes spécificités, propres à ces différentes tâches.

Du fait que cette grammaire doit pouvoir être transformée pour la reconnaissance dans des grammaires CFG, il faut tout d’abord que tous les traits aient un nombre fini de valeurs, le plus limité possible. En pratique, ceci signifie que les attributs ne peuvent pas prendre des valeurs complexes et que l’approche lexicaliste, défendue par LFG ou HPSG, devient difficile à mettre en œuvre. Par exemple, la rection ne peut pas être traitée ici avec des schémas de règles généraux, comme dans HPSG. Il faut au contraire multiplier les règles syntagmatiques pour chaque type de verbes (intransitif, transitif, etc.). Bien que cette première contrainte conduise clairement à des grammaires moins élégantes et plus répétitives, ceci ne semble pas un frein au développement de grammaires de la complexité requise pour ce type d’application.

Comme il s’agit avant tout d’une grammaire pour la reconnaissance, elle doit aussi intégrer toutes les informations susceptibles d’améliorer ce processus. Par exemple, différentes évaluations ont montré que les restrictions de sélection ne peuvent pas être omises sans dégrader considérablement la reconnaissance (Rayner *et al.*, 2006, chap. 11). En pratique donc, toutes les grammaires Regulus générales contiennent plusieurs traits pour gérer ce type de contrainte. Par exemple, les noms sont typés sémantiquement ; les entrées verbales contiennent quant à elles des traits qui précisent le type de compléments, en fonction de leur rection, etc. Ces types, difficiles à définir de manière cohérente pour le vocabulaire général, ne posent normalement pas de problèmes pour les applications liées à la parole où le domaine est bien cerné et le vocabulaire assez limité. Ils n’ont d’ailleurs aucune influence sur la structure globale de la grammaire générale, puisqu’ils proviennent des lexiques spécialisés pour les différents domaines.

Dans Regulus, la reconnaissance et l’analyse (syntaxique et sémantique) sont réalisées par la plateforme Nuance, après conversion de la grammaire d’unification dans le format requis par Nuance. La représentation sémantique est donc, elle aussi, particulière à ce type d’applications. Nuance construit en effet la sémantique de la phrase de manière compositionnelle, par concaténation de la sémantique des mots ou des phrases. Regulus (et Nuance) permettent différents types de représentations (Rayner *et al.*, 2006, chap. 7), mais, dans le cadre de ce projet où les domaines sont très contraints, nous pouvons nous limiter à la plus simple possible. Il s’agit d’une représentation plate, formée d’une liste de paires attributs-valeurs, avec un seul niveau d’enchâssement pour les phrases subordonnées (Rayner *et al.*, 2005b). Par exemple, *avez-vous mal à la tête quand vous buvez ?* sera représenté de la manière suivante¹ :

```
[[sc, quand], [clause, [[pronoun, vous], [voice, active],
                        [tense, present], [action, boire]]],
 [path_proc, avoir], [pronoun, vous], [symptom, mal],
 [tense, present], [utterance_type, ynq],
 [voice, active], [locative_prep, à], [body_part, tête]]
```

Ces représentations plates présentent en effet différents avantages. Dans le contexte de la traduction automatique de la parole, elles facilitent le transfert vers l’interlangue

1. *path_proc* signifie ici « pathological process » et *sc* « subordinate clause ».

ou à partir de celle-ci lors de la traduction (Rayner *et al.*, 2005b). Elles sont aussi compatibles avec le résultat d'un analyseur sémantique robuste, ce qui permet la comparaison avec les reconnaisseurs statistiques. Dans le même souci de simplicité, nous avons aussi choisi de ne pas faire figurer les déterminants dans la représentation. Les évaluations ont en effet montré que ces derniers sont difficiles à reconnaître correctement et à traduire, une fois reconnus. Il en va de même pour l'information sur le nombre. Ce choix aura évidemment un impact sur la manière de dériver la grammaire de génération puisqu'elle devra être beaucoup plus contrainte que la grammaire générale dont elle est pourtant issue – notamment, elle devra associer à chaque nom le déterminant et le nombre les plus appropriés.

Dans l'approche défendue ici, la compilation de la grammaire de génération se fait par une méthode similaire à celle de la « Semantic-Head-Driven Generation » (Shieber *et al.*, 1990). Comme nous venons de le suggérer, la génération et la reconnaissance ont cependant des exigences différentes. Pour la génération, la grammaire doit être très contrainte de manière à ne produire qu'un seul résultat pour chaque représentation ; pour la reconnaissance, nous voulons rester beaucoup plus souple puisque nous ne pouvons pas prédire la structure utilisée par l'utilisateur. Il y aurait évidemment différentes manières de poser la question plus haut : *Avez-vous ces maux de tête quand vous buvez ? Avez-vous des maux de tête quand vous buvez ? Avez-vous vos maux de tête quand vous buvez ? Avez-vous mal de tête quand vous buvez ?*, etc. Pour répondre à ces exigences différentes, les grammaires générales Regulus décrivent donc toutes les tournures utiles pour notre domaine. Celles-ci seront ensuite spécialisées différemment pour les deux tâches de reconnaissance et de génération, grâce à des corpus et des règles de spécialiation différentes. Dans la suite, nous illustrons cette méthodologie avec la grammaire partagée pour le français, le catalan et l'espagnol.

3. La grammaire partagée

La grammaire générale partagée a été spécifiquement conçue pour MedSLT. Elle décrit pour l'instant le français, le catalan et l'espagnol et couvre les différents domaines du diagnostic vus plus haut. Elle comprend trois types de règles : les règles syntagmatiques communes aux trois langues, des règles spécifiques, ainsi que des règles communes à l'espagnol et au catalan.

Comme les trois langues sont très proches, les règles spécifiques sont très peu nombreuses, ce qui motive entièrement l'approche proposée ici. En français, par exemple, elles décrivent principalement les questions oui-non, formées avec *est-ce que* (*Est-ce que vous avez mal à la tête ?*), par inversion du pronom sujet avec ajout d'un tiret (*Avez-vous mal à la tête ?*) ou par inversion complexe (*La douleur est-elle latérale ? La douleur est-elle causée par le vin rouge ? En combien de temps la douleur irradie-t-elle la nuque ?*) ; la partie commune au catalan et à l'espagnol regroupe une dizaine de règles pour les phrases déclaratives et interrogatives sans sujet (par exemple, *Té mal de cap ?*, « avez mal de tête ? » en catalan), ainsi que les phrases avec inversion du pronom sujet sans tiret (*Quan té vostè dolor ?*, « Quand avez vous douleur », toujours

en catalan) et du nom sujet (*Quan apareix el dolor ?*, « Quand apparaît la douleur ? »). Ces dernières sont bien sûr aussi possibles en français, notamment quand la proposition commence par un mot interrogatif attribut ou un complément d'objet direct (*Qu'a dit le docteur ?*), mais ces constructions n'existent pas dans notre corpus – elles ne font donc pas encore partie de la description commune.

La grammaire commune comprend pour l'instant quatre-vingt-dix-neuf règles syntagmatiques, paramétrables pour les trois langues. Le paramétrage se fait avec des macros. Celles-ci permettent d'instancier une valeur différente pour un attribut dans les différentes langues. Par exemple, dans la règle suivante :

```
vp_comps:[..., subcat=trans, ...] -->
  np:[@french_catalan_or_spanish(has_spec=yes,
    has_spec=_,
    has_spec=_), ...].
```

la macro (caractérisée par le symbole @) rend compte du fait que, en français, le syntagme nominal objet des verbes transitifs requiert normalement un déterminant (sauf s'il s'agit d'un verbe support traité par une autre règle), alors qu'en catalan et en espagnol cette contrainte est moins forte (la valeur de l'attribut `had_spec=_` n'est donc pas spécifiée). De la même manière, une macro similaire permet de spécifier que, en français, tous les verbes réfléchis sont formés avec le pronom enclitique *se* (sauf à l'impératif qui n'est pas inclus ici dans les verbes finis), alors que, en espagnol et en catalan, *se* se trouve avant le verbe quand celui-ci est fini ou au participe présent, mais après le verbe à l'infinitif et à l'impératif, par exemple *se levanta* en espagnol (« il se lève ») vs *levantarse* (« se lever ») :

```
vbar:[@french_catalan_or_spanish(
  vform=(finite\infinitive\pres_participle),
  vform=(finite\pres_participle),
  vform=(finite\pres_participle)), vform=Vform] -->
  reflexive_pronoun:[...],
  vbar:[..., vform=Vform].
```

Les macros sont aussi utilisées pour instancier des lexèmes différents dans les trois langues. Par exemple, ici, la macro permet de réaliser différemment le pronom interrogatif sujet :

```
wh_question:[...] -->
  @french_catalan_or_spanish(
    ('qu'est-ce', qui),
    què,
    qué),
  vp:[inv=uninverted, ...].
```

Malgré ses spécificités pour la parole, la grammaire commune propose un traitement général des phénomènes les plus importants dans les langues romanes. Les clitiques ne sont pas traités avec des règles de mouvement, comme dans beaucoup de grammaires, par exemple (Rayner *et al.*, 2000) et (Sportiche, 1996). Comme nous ne pouvons pas avoir recours à des structures complexes, cette approche nous obligerait en effet à multiplier les traits, de manière à pouvoir faire passer les informations du constituant avec le verbe et le clitique à son *gap*. Pour éviter de trop complexifier la grammaire CFG, nous nous inspirons donc plutôt de (Miller *et al.*, 1997) et nous proposons une approche lexicaliste où une règle de grammaire est utilisée comme une règle lexicale qui a ici la forme :

```
vbar: [] -->
    pronoun: [],
    vbar: [] .
```

Ainsi, la règle suivante forme un *vbar* transitif à partir d'un pronom clitique et d'un *vbar* de type *di* transitif (comme *donner* en français) :

```
vbar: [subcat=trans] -->
    pronoun: [pron_type=clitic, ...],
    vbar: [subcat=di trans, ...].
```

Ce *vbar* (par exemple, *vous donne*) pourra ensuite se combiner avec un objet direct, puis un sujet pour former une phrase complète.

Pour les mêmes raisons, l'inversion n'est pas non plus considérée comme un type de mouvement. À nouveau, nous exploitons le constituant *vbar* :

```
yn_question: [] --> vp: [inv=inverted, ...].

vp: [inv=...] -->
    vbar: [inv=inverted\uninverted, ...],
    vp_comps: [...],
    optional_vp_postmods: [...].
```

Nous considérons qu'une question oui-non (*yn_question*) est formée d'un *vp* (inversé). Celui-ci est constitué d'un *vbar* et de son complément (*vp_comps*). Quand il est inversé, ce *vbar* peut se réécrire de différentes manières, en fonction de la langue. En français, il peut s'agir d'un verbe suivi d'un tiret, d'un pronom et d'un éventuel adverbe :

```
vbar: [inv=inverted, ...] -->
    verb: [], hyphen: [], pronoun: [], optional_adverb: [] .
```

En catalan et en espagnol, nous aurons la variante suivante :

```
vbar: [inv=inverted, ...] -->
      verb: [], (pronoun: [], np: []), optional_adverb: [] .
```

Cette approche ne complexifie pas la grammaire. Comme on le voit plus haut, la règle pour le vp s'applique en effet aux vbar inversé et noninversé. De plus, elle permet de rendre facilement compte du fait que, en français, le vbar inversé peut être coordonné (*As-tu ou avez-vous déjà eu la grippe ?*).

Le traitement des questions *wh* est plus traditionnel que celui de l'inversion : il est impossible de les traiter sans au moins simuler le mouvement. Nous considérons donc que le pronom interrogatif est déplacé de sa position initiale (objet direct, PP, etc.), laissée vide, en position initiale (*quels médicaments_i prenez-vous [i]*). Pour traiter le mouvement, nous utilisons le mécanisme standard du *gap threading*, proposé déjà par (Pereira, 1981). Le lien entre le constituant vide [*i*] et le constituant déplacé (*quels médicaments_i* dans notre exemple) se fait avec les deux attributs *gapsin* et *gapsout*, qui apparaissent dans toutes les catégories impliquées par le mouvement. Par exemple, dans la règle suivante, ces attributs indiquent que le *wh_np* n'est possible que si le vp contient un np vide (*gapsin=np_gap*).

```
wh_question: [] -->
      wh_np: [...],
      vp: [inv=inverted, gapsin=np_gap, gapsout=null] .
```

Finalement, pour le passif, nous appliquons une règle lexicale. Celle-ci produit, à partir d'une entrée transitive au participe passé, une entrée passive avec la rection *passivised_trans* :

```
vbar: [voice=passive, subcat=passivised_trans, ...] -->
      verb: [vform=past_participle, subcat=trans, voice=active, ...],
      optional_adverb: [sem=Advp, position=post] .
```

Ce vbar pourra ensuite se combiner avec un autre vbar (pour former un vbar fini) et un vp_comps (ici, un complément d'agent introduit par *par*, *por* ou *per* en fonction des langues) pour former une phrase passive. Dans la suite, nous voyons comment cette grammaire générale peut être spécialisée différemment pour la reconnaissance et la génération.

4. Spécialisation de la grammaire générale pour la reconnaissance et la génération

L'un des principaux avantages de l'approche adoptée ici est que la grammaire générale peut être spécialisée pour différentes tâches et domaines, de manière à obtenir

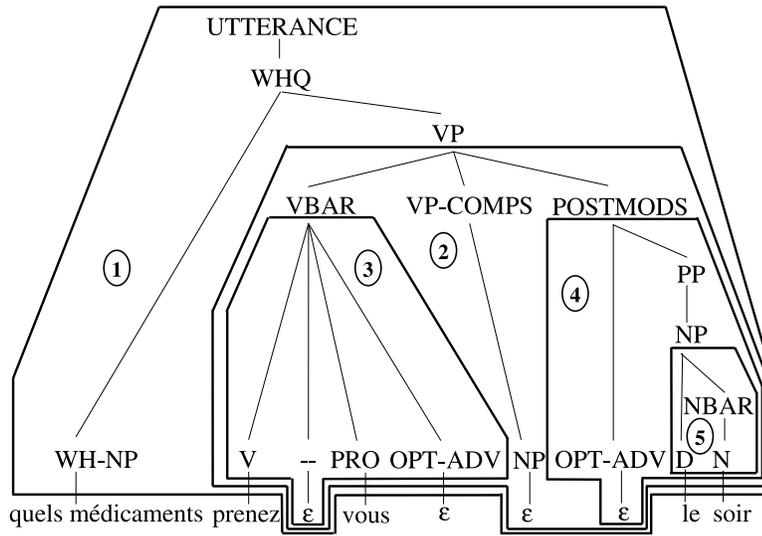


Figure 1. *Processus de spécialisation de la grammaire générale à partir d'un exemple*

la grammaire la plus performante dans chaque cas. Comme la grammaire générale est encore assez limitée, il est techniquement possible de la compiler directement pour la reconnaissance et la génération. L'évaluation montrera cependant qu'elle est déjà assez complexe pour causer de graves problèmes d'efficacité. La spécialisation est en effet essentielle pour toute application basée sur les grammaires. Elle permet de les simplifier, limitant ainsi l'espace de recherche au niveau de la reconnaissance, l'ambiguïté de l'analyse, ainsi que la surgénération.

Dans l'approche proposée ici, la spécialisation se fait automatiquement avec la méthode d'EBL citée plus haut, sur base des critères d'opérationnalité et des corpus. Les corpus donnent des exemples de mots et des types de structures qui doivent figurer dans la grammaire spécialisée ; les critères déterminent, par des règles, quels sont les constituants à garder dans les grammaires spécialisées et comment les restructurer.

Les figures 1 et 2 illustrent le processus de dérivation des grammaires spécialisées avec un exemple concret. La première d'entre elles reprend l'arbre d'analyse de la phrase *Quels médicaments prenez-vous le soir?* (légèrement simplifié ici), avec la grammaire générale. Les lignes en gras indiquent les constituants (sous-arbres) extraits par les critères d'opérationnalité (tels qu'ils ont été définis pour les langues romanes dans le contexte de MedSLT). C'est cette information qui sert de point de départ à l'algorithme EBL pour dériver les grammaires spécialisées à partir des exemples du corpus : chaque sous-arbre est transformé dans une règle de la grammaire spécialisée – la partie gauche de la règle dérivée correspond à la racine de l'arbre et la partie droite

à l'ensemble des catégories feuilles du sous-arbre. Pour le sous-arbre 1, le nœud WHQ est donc éliminé et nous obtenons schématiquement la règle dérivée suivante :

```
utterance --> wh_np, vp.
```

Comme les règles spécialisées conservent tous les traits de la grammaire générale (suivant nos critères), celle-ci hérite en plus des traits comme `gapsin` et `gapsout` (cf. section 3), ce qui donne plus précisément :

```
utterance:[] --> wh_np:[], vp:[gapsin=np_gap, gapsout=null, ...].
```

L'algorithme s'applique de la même manière aux autres sous-arbres. Le sous-arbre 2 inclut une dérivation vide pour la trace du NP. Comme celle-ci fait partie du sous-arbre, les nœuds NP et VP-COMPS disparaissent dans la règle spécialisée qui correspond donc ici à :

```
vp --> vbar, postmods.
```

Le sous-arbre 3 est le plus complexe. Le tiret, qui n'est pas nécessaire pour la reconnaissance, domine une chaîne vide qui fait partie du sous-arbre et qui disparaît donc de la règle dérivée, comme le NP plus haut. Par contre, la dérivation pour l'adverbe optionnel OPT_ADV n'est pas incluse dans l'arbre. Ce constituant est donc retenu dans la règle spécialisée qui a la forme :

```
vbar --> v, pro, opt_adv.
```

Les arbres 4 et 5 sont similaires aux précédents. Ils produisent les deux dernières règles :

```
postmods --> np.
```

```
et
```

```
np --> d, n.
```

La figure 2 montre la même phrase, mais analysée maintenant avec la grammaire spécialisée que nous venons de dériver. Chaque sous-arbre y est devenu une règle dans l'arbre dérivé, ce qui conduit évidemment à des structures simplifiées et moins ambiguës, dont nous montrerons l'intérêt plus tard pour l'efficacité du système (cf. section 5).

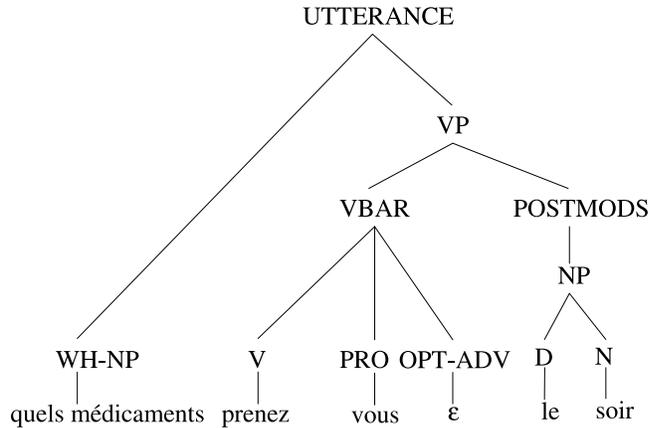


Figure 2. Même exemple que dans la figure 1, mais analysé avec la grammaire spécialisée

Pour arriver aux grammaires les plus adéquates possibles pour la reconnaissance et la génération, nous faisons donc varier les deux paramètres vus plus haut : les corpus d'apprentissage et les critères d'opérationnalité. De manière générale, il est intéressant de constater que des critères d'opérationnalité similaires peuvent être utilisés pour les deux types de grammaire, conduisant presque à la même granularité. Pour la génération, ce niveau suffit pour produire les bonnes structures, sans en surgénérer de mauvaises ; pour la reconnaissance, nous verrons dans la dernière partie sur l'évaluation que, d'une part, les grammaires plus granulaires allourdissent le processus de décodage et que, d'autre part, les grammaires moins granulaires ne permettent plus assez de généralisations.

La seule différence importante entre les deux grammaires spécialisées au niveau de la structure concerne en effet le traitement des NPs. Dans la grammaire pour la reconnaissance, nous généralisons, à partir du corpus et du lexique, l'ensemble des déterminants et des noms. La grammaire spécialisée contient ensuite une règle qui forme de manière traditionnelle un NP à partir d'un déterminant et d'un nom. Celle-ci nous permet de reconnaître toutes les combinaisons des noms et des déterminants qui se trouvent dans le corpus et le lexique, ce qui est bénéfique pour la reconnaissance. Pour la génération en revanche, les critères d'opérationnalité diffèrent : de manière à produire le déterminant plus approprié pour chaque nom sur base du corpus, nous apprenons des NP complets (lexicalisés). Si nous compilons donc l'exemple de la

figure 1 en vue de dériver la grammaire de génération, le sous-arbre 5 (à droite) couvrirait aussi les nœuds lexicaux *le* et *soir*, ce qui conduirait à la règle spécialisée :

np --> le, soir.

Comme les déterminants sont difficiles à reconnaître et ensuite à traduire correctement, la spécialisation par domaine et par tâches offre donc ici une excellente solution : nous générons simplement les combinaisons *déterminant* + *nom* attestées dans nos corpus de référence. De cette manière, nous pouvons très facilement rendre compte de toutes les spécificités des différents sous-domaines. Dans le sous-domaine des maux de tête par exemple, le déterminant le plus approprié pour le nom *maux de tête* est le possessif *vos* puisqu'il s'agit de l'objet du diagnostic (*Depuis quand avez-vous vos maux de tête ?*, *Vos maux de tête sont-ils fréquents ?*); dans le domaine des douleurs abdominales, il s'agira en revanche de l'article indéfini (*Avez-vous eu des maux de tête la nuit passée ?*).

Cette approche a évidemment un impact sur les corpus d'entraînement et explique l'une des différences essentielles entre les corpus pour la génération et la reconnaissance; le corpus de génération doit énumérer toutes les séquences de déterminants et de noms qui doivent pouvoir être générées, ce qui n'est pas le cas du corpus de reconnaissance puisque les critères d'opérationnalité intègrent tous les noms du lexique général dans la grammaire spécialisée. En revanche, celui-ci devra donner plus d'exemples puisqu'il doit couvrir toutes les variantes possibles au niveau des structures. Il contient, par exemple, des phrases interrogatives avec *est-ce que* (*Est-ce que vous avez mal sur le devant de la tête ?*) ou sans inversion (*Vous avez mal sur le devant de la tête ?*), auxquelles on préférera l'inversion lors de la génération (*Avez-vous mal sur le devant de la tête ?*). De même, ce corpus reprend aussi des phrases avec des déterminants élidés (*l'*, *s'*, etc.) et des prépositions ou déterminants contractés (*au*, *aux*, *del*, etc.), qui pourront ainsi être reconnus par le système. Au niveau de la génération, ces différents phénomènes sont traités avec des règles orthographiques qui s'appliquent après la génération de la phrase. Cette approche nous évite de complexifier la grammaire générale avec les nombreux traits potentiellement nécessaires pour contraindre des phénomènes comme l'éllision ou la contraction au niveau des règles elles-mêmes.

Le tableau 1 résume les différences entre les grammaires générales et spécialisées au niveau du nombre de règles d'unification (UG) et CFG (après conversion avec Regulus) et du nombre de mots (en terme de lemmes, de formes fléchies et de formes de surface²). Si l'on regarde le français, on voit directement que, comme escompté, la grammaire de génération compte moins de règles CFG que celle de reconnaissance. Le

2. Par formes de surface, nous entendons ici le nombre total de chaînes de caractères impliquées dans les entrées lexicales. Par exemple, l'entrée *résonance magnétique nucléaire* correspond à un lemme, une forme fléchiée et trois formes de surface. Le nombre de formes de surface a évidemment une influence sur la reconnaissance puisqu'il détermine l'espace de recherche pour le reconnaissseur.

nombre de règles UG est plus élevé dans les grammaires spécialisées que les générales parce que, lors du processus de spécialisation, les règles sont combinées et résolues de différentes manières en fonction des corpus d'entraînement. Les différences entre les langues au niveau du nombre de règles et de mots ne sont pas vraiment significatives ici et reflètent simplement l'état d'avancement et la taille du corpus d'entraînement. En espagnol et en catalan, nous avons en effet surtout développé le système pour la génération, ce qui fait que la grammaire de génération est plus élaborée que celle de reconnaissance.

	UG	CFG	Lemmes	F. fléchies	F. surface
Français					
Générale	84	15418	727	1685	1254
Reconnaissance	144	8397	–	1464	1111
Génération	142	4137	–	1552	1188
Espagnol					
Générale	99	24706	461	1743	1297
Reconnaissance	119	5344	–	431	359
Génération	167	5514	–	1563	1149
Catalan					
Générale	90	24602	863	2024	1503
Reconnaissance	78	2954	–	835	673
Génération	112	3853	–	1639	1201

Tableau 1. Statistique pour les différentes grammaires

Mais comment évaluer la qualité des grammaires spécialisées pour les différentes tâches ? Il nous faut montrer que, grâce à la spécialisation, il nous est possible de dériver la meilleure grammaire pour une application et une langue données.

5. Évaluation

Pour évaluer les grammaires spécialisées de reconnaissance, nous avons d'abord mesuré la qualité de la reconnaissance avec ces grammaires dans les trois langues ; nous avons ensuite aussi effectué la reconnaissance avec la grammaire générale et de génération, pour mieux comparer les résultats entre eux.

Nous avons donc commencé par collecter des données dans les différentes langues. En français, nous avons simulé un véritable diagnostic avec des étudiants de dernière année en médecine et des patients standardisés, suivant une procédure qui a été décrite dans (Chatzichrisafis *et al.*, 2006) ; en espagnol et en catalan, les données ont été collectées auprès d'étudiants et de collègues qui jouaient le rôle du docteur, en se basant sur le corpus existant.

Ces données ont ensuite été divisées en trois groupes : (1) les questions de diagnostic qui font partie du corpus d'entraînement (*In training*), (2) les autres questions couvertes par la grammaire (*In coverage*) et (3) les questions non couvertes par la grammaire (*Out of coverage*). Cette distinction entre données couvertes et non couvertes est importante dans le contexte de notre application où toute l'architecture est conçue pour diminuer le nombre de phrases non couvertes par la grammaire, grâce à un système d'aide. Celui-ci aide l'utilisateur à apprendre la couverture du système, en lui fournissant pour chaque phrase d'entrée une liste de phrases similaires qui sont couvertes par le système. Pour ce faire, il utilise la reconnaissance statistique, plus robuste par définition que la reconnaissance linguistique et compare le résultat avec des phrases préenregistrées du même domaine pour extraire les plus similaires. L'utilisateur peut lire ces phrases, s'en inspirer pour la requête suivante ou même sélectionner l'une d'entre elles ((Rayner *et al.*, 2005a) et (Chatzichrisafis *et al.*, 2006)). En d'autres termes, nous supposons que l'utilisateur pourra reformuler les phrases qui ne peuvent pas être reconnues avec le système d'aide et notre but est donc surtout d'obtenir des résultats compétitifs pour les données couvertes par la grammaire.

Sur la base de ces données, nous avons ainsi pu calculer la qualité de la reconnaissance avec la grammaire spécialisée pour la reconnaissance dans les trois langues. Pour chacun des trois groupes de données, nous avons mesuré, comme on le fait traditionnellement, le taux d'erreurs au niveau des mots (*WER*) et des phrases (*SER*). Comme ces mesures sont connues pour être peu fiables (Wang *et al.*, 2003), nous avons aussi compté le nombre d'erreurs sémantiques (*SemER*), c'est-à-dire le pourcentage de phrases qui ne préservent pas le sens. Ce sont ici les phrases qui n'ont pas été envoyées à la traduction par les sujets lors de la collecte des données. Les résultats sont résumés dans le tableau 2.

À première vue, si on ne regarde que les mesures traditionnelles (*WER* et *SER*), les résultats sont moins bons en français que dans les deux autres langues, ainsi qu'en anglais (Bouillon *et al.*, 2005). On pourrait s'en étonner. Après tout, la grammaire a d'abord été conçue pour le français, puis adaptée aux autres langues. On sait aussi que la version catalane de Nuance est de mauvaise qualité : elle n'a été entraînée que sur le catalan de Barcelone, ce qui aurait dû avoir une influence négative sur les résultats.

Ces chiffres doivent cependant être relativisés. D'abord, nous avons vu plus haut (tableau 1) que le nombre de mots et de règles dans la grammaire spécialisée est beaucoup plus important en français, ce qui peut avoir une légère influence sur les résultats, et en particulier sur le *WER* et *SER* (Rayner *et al.*, 2006, chap. 11). Mais il faut surtout remarquer ceci : si nous comparons maintenant les résultats du *WER* et *SER* avec ceux du *SemER*, nous voyons clairement que ceux-ci s'inversent : le *SemER* est nettement plus faible en français que dans les deux autres langues (5,5 % versus 7,7 % et 9,9 %), ce qui signifie que **la plupart des erreurs faites en français sont minimales et n'entravent pas la traduction**. Comme nous l'espérons, le reconnaiseur français est donc meilleur que le catalan, et ceci contrairement à ce que laissait suggérer le *WER* et le *SER*. Un bref examen des erreurs de reconnaissance en français confirme d'ailleurs cette conclusion : la plupart des fautes concernent effectivement des éléments qui ne

	#Utts	WER	SER	SemER	xRT
Français					
Total	427	15,5 %	45,5 %	19,9 %	0,06
Out of coverage	100	38,8 %	99,9 %	67,7 %	0,04
In training	119	3,3 %	12,2 %	2,2 %	0,05
In coverage	208	10,0 %	38,8 %	7,7 %	0,04
In coverage (total)	327	8,8 %	28,8 %	5,5 %	0,05
Espagnol					
Total	578	31,1 %	57,7 %	46,6 %	0,01
Out of coverage	295	57,7 %	100,0 %	83,3 %	0,05
In training	138	6,6 %	16,6 %	8,8 %	0,09
In coverage	145	4,4 %	18,8 %	5,5 %	0,08
In coverage (total)	283	5,5 %	17,7 %	7,7 %	0,09
Catalan					
Total	463	32,2 %	63,3 %	32,2 %	0,03
Out of coverage	230	57,7 %	99,9 %	76,6 %	0,05
In training	178	7,7 %	27,7 %	8,8 %	0,00
In coverage	55	8,8 %	32,2 %	14,4 %	0,02
In coverage (total)	233	7,7 %	28,8 %	9,9 %	0,02

Tableau 2. Reconnaissance avec la grammaire spécialisée pour la reconnaissance

figurent pas dans la représentation sémantique de la phrase (section 2), comme le *-t-* euphonique ou les déterminants, d'ailleurs plus nombreux en français que dans les deux autres langues (tableau 3). Il arrive aussi souvent que l'on obtienne une variante synonyme de la phrase de départ – comme le corpus d'entraînement contient très peu de structures interrogatives avec *est-ce que*, le reconnaiseur a, par exemple, tendance à reconnaître l'équivalent sans inversion (tableau 4), ce qui n'a aucune influence sur la représentation, ni sur la traduction.

Phrases reconnues	Input
avez vous mal à le arrière de la tête	avez vous mal à l arrière de la tête
douleurs irradiant l épaule	la douleur irradie t elle l épaule
la douleur augmente elle graduellement	la douleur augmente t elle graduellement

Tableau 3. Exemples de phrases mal reconnues au niveau du déterminant ou du « *-t-* » euphonique

L'examen du tableau 2 permet par ailleurs de tirer une autre conclusion intéressante : on voit que, de manière générale, les résultats pour les données d'entraînement sont assez comparables à ceux pour les données couvertes par la grammaire (surtout en catalan). En espagnol, le pourcentage d'erreurs est même meilleur pour les phrases

Phrases reconnues	Input
la douleur dure plus de une heure	est-ce que la douleur dure plus d une heure
la douleur irradie elle la nuque	est-ce que la douleur irradie vers la nuque
avez vous des vertiges quand vous avez mal à la tête	avez vous des vertiges quand vous avez mal
avez vous mal derrière la tête	avez vous mal à l arrière de la tête

Tableau 4. Exemples où la phrase reconnue est une variante synonyme de la phrase de départ

couvertes par la grammaire que pour les données d’entraînement elles-mêmes. Les généralisations de la grammaire à partir des corpus semblent donc utiles et correctes et ceci dans les trois langues.

Pour mieux évaluer la qualité des grammaires spécialisées pour la reconnaissance, nous avons, dans un second temps, effectué la reconnaissance avec la grammaire de génération et la grammaire générale, et ceci pour le français. Les résultats dans les tableaux 5 et 6 confirment entièrement nos intuitions de départ. Du fait que la grammaire de génération est plus contrainte (notamment au niveau des groupes nominaux, cf. section 4), le *WER* est à première vue amélioré pour les données couvertes par la grammaire. Il faut cependant directement noter que le *SemER* ne diminue presque pas (5,1 % *versus* 5,5 %) : le pourcentage de phrases traduisibles reste donc quasi le même, les erreurs évitées concernent essentiellement les déterminants, ce qui n’a aucune influence sur le système. Le gain est d’autant plus réduit pour l’utilisateur que cette faible amélioration du *SemER* est corrélée à une augmentation importante des phrases qui ne sont pas couvertes par la grammaire (270 *versus* 100), pour lesquelles le *SemER* est par définition mauvais³. C’est d’ailleurs précisément le type de questions que nous voulons éviter dans ce projet grâce au système d’aide, dont nous avons parlé plus haut.

Avec la grammaire générale, les résultats sont encore plus tranchés : la reconnaissance se dégrade clairement à tous les niveaux, même si le nombre de phrases non couvertes diminue un peu, puisque la grammaire permet plus de généralisations. L’efficacité du système est aussi affectée : le temps de la reconnaissance (*xRT*) augmente considérablement (0,07 avec la grammaire générale *versus* 0,05 avec la grammaire spécialisée pour la reconnaissance) ; le pourcentage de phrases ambiguës est aussi plus important, ce qui est toujours une source potentielle d’erreurs. En français, nous avons une moyenne de 1,17 analyses pour les phrases d’entraînement avec la grammaire générale, mais seulement 1,15 avec la grammaire spécialisée pour la reconnaissance ; en espagnol, le contraste est plus important encore (4,43 *versus* 1,18 analyses), no-

3. Quelle que soit d’ailleurs la méthode de reconnaissance utilisée, linguistique ou statistique, si nous utilisons le même corpus d’entraînement pour les deux types de reconnaissseurs (Bouillon *et al.*, 2005).

tamment parce que le lexique est moins contraint au départ. Ce qu'il est important de noter ici, c'est que la spécialisation estompe les différences (en partie historiques) au niveau des grammaires générales et les rend plus uniformes, une fois spécialisées.

Nous arrivons ainsi à la conclusion que, dans le type d'application comme Med-SLT, la diminution du *WER* et *SER* ne va pas de pair avec une amélioration globale du système. Il est, de plus, essentiel de pouvoir, grâce à une méthode comme la nôtre, dériver des grammaires différentes et les comparer entre elles pour arriver à la meilleure grammaire pour une application donnée.

	#Utts	WER	SER	SemER	xRT
Total	427	34,4 %	68,8 %	40,0 %	0,07
Out of coverage	270	48,8 %	100,0 %	60,0 %	0,06
In training	89	2,2 %	10,0 %	2,2 %	0,07
In coverage	68	5,5 %	26,6 %	8,8 %	0,06
In coverage (total)	157	3,3 %	14,4 %	5,5 %	0,06

Tableau 5. *Reconnaissance en français avec la grammaire spécialisée pour la génération*

	#Utts	WER	SER	SemER	xRT
Total	427	31,1 %	63,3 %	33,3 %	0,05
Out of coverage	83	49,9 %	98,8 %	72,2 %	0,04
In training	121	24,4 %	43,3 %	21,1 %	0,00
In coverage	223	28,8 %	61,1 %	26,6 %	0,08
In coverage (total)	344	27,7 %	55,5 %	24,4 %	0,07

Tableau 6. *Reconnaissance en français avec la grammaire générale*

6. Conclusion

Dans cet article, nous avons présenté une méthode et un ensemble d'outils (Regulus) pour dériver facilement et rapidement des reconnaisseurs linguistiques et des générateurs à partir de grammaires générales, partagées pour plusieurs langues. Plus précisément, nous avons montré que :

- il est possible de développer des grammaires linguistiquement motivées pour la parole,
- ces grammaires peuvent être partagées entre plusieurs langues sans dégrader la performance,
- de ces grammaires, il est possible de dériver des grammaires spécialisées spécifiques en fonction des données et de la tâche

– et ces grammaires spécialisées peuvent ainsi être comparées entre elles de manière à obtenir la meilleure grammaire possible pour un ensemble de données.

Nous pensons ainsi avoir contribué à montrer l'intérêt de l'approche linguistique pour la parole et à en définir les bases de manière plus rigoureuse. Dans l'approche proposée ici, les règles ne sont plus des ensembles *ad hoc*, dont les performances sont difficilement prévisibles et varient en fonction de la personne qui écrit la grammaire. Il s'agit au contraire de modèles construits et dérivés à partir de modèles généraux de la langue selon une méthodologie reproductible et bien définie.

Remerciements

Ce projet est financé par le Fonds National de la Recherche Suisse, que nous remercions ici.

7. Bibliographie

- Bender E., Flickinger D., Oepen S., « The Grammar Matrix : An Open Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars », *Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan, p. 8-14, 2002.
- Bos J., « Compilation of Unification Grammars with Compositional Semantics to Speech Recognition Packages », *Proceedings of the 19th International Conference on Computational Linguistics*, Taipei, Taiwan, 2002.
- Bouillon P., Rayner M., Chatzichrisafis N., Hockey B., Santaholma M., Starlander M., Nakao Y., Kanzaki K., Isahara H., « A Generic Multi-Lingual Open Source Platform for Limited-Domain Medical Speech Translation », *Proceedings of the 10th Conference of the European Association for Machine Translation (EAMT)*, Budapest, Hungary, p. 50-58, 2005.
- 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.
- Dowding J., Hockey B., Gawron J., Culy C., « Practical Issues in Compiling Typed Unification Grammars for Speech Recognition », *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, France, p. 164-171, 2001.
- Kim R., Dalrymple M., Kaplan R., King T., Masuichi H., Ohkuma T., « Multilingual Grammar Development via Grammar Porting », *Proceedings of the ESSLLI Workshop on Ideas and Strategies for Multilingual Grammar Development*, Vienna, Austria, 2003.
- 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.
- Miller P. H., Sag I., « French Clitic Movement Without Clitics or Movement », *Natural Language and Linguistic Theory*, vol. 15, p. 573-639, 1997.

- Moore R., « Using Natural Language Knowledge Sources in Speech Recognition », *Proceedings of the NATO Advanced Studies Institute*, p. 115-129, 1998.
- Pereira F., « Extraposition Grammars », *American Journal of Computational Linguistics*, vol. 7, p. 243-256, 1981.
- 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, 2005a.
- Rayner M., Bouillon P., Santaholma M., Nakao Y., « Representational and Architectural Issues in a Limited-Domain Medical Speech Translator », *Proceedings of TALN 2005*, Dourdan, France, p. 163-172, 2005b.
- Rayner M., Carter D., Bouillon P., Digalakis V., Wirén M. (eds), *The Spoken Language Translator*, Cambridge University Press, 2000.
- Rayner M., Hockey B., Bouillon P., *Putting Linguistics into Speech Recognition : The Regulus Grammar Compiler*, CSLI Press, Chicago, 2006.
- Regulus, *Regulus project website*, <http://sourceforge.net/projects/regulus/>. 2006, As of 1 Nov 2006.
- Shieber S., van Noord G., Pereira F., Moore R., « Semantic-Head-Driven Generation », *Computational Linguistics*, 1990.
- Sportiche D., « Clitic Constructions », in J. Rooryck, L. Zaring (eds), *Phrase Structure and The Lexicon*, IULC Press, Bloomington, Indiana, p. 213-276, 1996.
- Wang Y.-Y., Acero A., Chelba C., « Is Word Error Rate a Good Indicator for Spoken Language Understanding Accuracy », *Proceedings of Eurospeech 2003*, Geneva, Switzerland, p. 609-612, 2003.