

Analyse syntaxique robuste de dialogues transcrits : peut-on vraiment traiter l'oral à partir de l'écrit ?

Narjes Boufaden¹, Sylvain Delisle² Bernard Moulin^{1,3}

(1) Département d'informatique, Université Laval, Québec, Canada

(3) Centre de recherche en géomatique de l'Université Laval

Courriel : {boufaden, moulin}@ift.ulaval.ca

(2) Département de mathématiques et d'informatique, Université du Québec à

Trois-Rivières, Québec, Canada

Courriel : Sylvain_Delisle@uqtr.quebec.ca

Résumé

L'analyseur syntaxique robuste que nous décrivons dans cet article s'insère dans le cadre des travaux relatifs au traitement du langage oral. Nous montrons à partir d'une étude menée sur des dialogues retranscrits qu'il est avantageux de traiter le langage oral avec un analyseur syntaxique robuste faisant appel à un analyseur syntaxique conçu pour l'écrit. Pour ce faire, nous avons réalisé un système qui se base sur une architecture à deux couches : le noyau et la périphérie, l'une interagissant avec l'autre via un superviseur. La première couche, le noyau, est dédiée à l'analyse des constituants d'énoncés respectant la grammaire standard : c'est un analyseur syntaxique de l'écrit. La deuxième couche, la périphérie, se charge de "corriger" les constituants de l'énoncé oral ayant subis des distorsions. Cette couche intervient lorsque le noyau ne parvient plus à progresser dans son analyse. L'ordre d'intervention de ces deux couches dans le processus d'analyse est déterminé par l'occurrence de marques de surface qui signalent la présence d'une distorsion ou d'une construction particulière (interrogative, relative, etc.). Le système ainsi conçu nous a permis de traiter les bruits et différents types de répétitions caractéristiques de l'oral. La première version de l'analyseur nous a fourni des résultats encourageants qui nous permettent de confirmer l'interdépendance entre le traitement du langage oral et le traitement du langage écrit. Ces résultats nous invitent aussi à reconsidérer le rôle et l'utilité, pour le traitement de l'oral, des ressources d'informatique linguistique développées pour l'écrit.

1. Introduction

La plupart des analyseurs syntaxiques conçus pour le traitement du langage naturel sont construits pour des textes écrits qui par définition sont conformes à une grammaire bien définie (la grammaire standard). Les entrées prévues pour ces analyseurs doivent être dépourvues de toutes extra-grammaticalités. Cependant, une telle conception rend ces analyseurs fragiles et non robustes face aux différentes distorsions que l'on trouve dans le langage oral. Cette constatation a suscité un grand nombre de questions qui se résument à savoir comment situer l'oral par rapport à l'écrit. Au-delà des modes de communication différents, y a-t-il d'autres dissimilarités entre l'oral et l'écrit ? Faut-il considérer que le traitement de l'oral est une discipline indépendante nécessitant ses propres solutions ? Ou, au

contraire, faut-il traiter l'oral dans une perspective d'adaptation des solutions conçues pour l'écrit ?

Deux points de vue se sont développés concernant l'élaboration d'un analyseur syntaxique robuste pour le traitement des dialogues oraux. Le premier point de vue porte sur la conception d'un analyseur syntaxique robuste des dialogues oraux à partir d'une grammaire de l'oral. Ainsi, en augmentant les règles de grammaire standard par des règles générant différentes constructions de l'oral, il serait possible d'envisager l'automatisation de l'analyse syntaxique des dialogues oraux. Cependant comme l'explique Kong *et al.* (1995), une telle démarche peut engendrer une augmentation du nombre de règles de manière disproportionnée, ce qui rendrait leur gestion impossible.

Le deuxième point de vue prévoit la conception d'un analyseur syntaxique robuste à partir de règles de grammaire standard et d'heuristiques gérant les distorsions de l'oral. Concrètement, ceci s'effectue en concevant un analyseur syntaxique de l'écrit (non robuste) et en ajoutant des procédures permettant le traitement des extra-grammaticalités de l'oral. Ces procédures sont généralement basées sur les opérations d'ajouts, d'effacements, de transpositions ou de substitutions de constituants. Cette perspective est actuellement celle qui est la plus utilisée. Toutefois, elle présente aussi un inconvénient lié au fait que ces méthodes sont moins performantes pour les applications indépendantes d'un domaine (Rosé et Lavie 1997).

D'autres alternatives sont en voie de développement. Ces alternatives s'inspirent de la deuxième approche et préconisent l'ajout d'une deuxième phase de traitement. Elle consiste à combiner les analyses partielles obtenues pour un énoncé afin d'en extraire la combinaison la plus plausible. Cette stratégie permettrait de résoudre le problème de la deuxième approche lié à l'utilisation de connaissances dépendantes du domaine pour la définition des traitements des distorsions de l'oral (Rosé et Lavie 1997).

Dans cet article, nous nous intéressons plus à la deuxième approche. Nous visons à montrer qu'il est possible de traiter des dialogues oraux orientés tâches en exploitant une théorie de la grammaire et une technique d'analyse essentiellement utilisée pour la langue écrite. Le corpus sur lequel nous travaillons est une retranscription orthographique d'enregistrements authentiques constitué par Ozkan (Ozkan 1994) et analysé par Colineau (Colineau 1997). Ce corpus met en scène un instructeur et un manipulateur, qui n'étaient pas dans la même pièce, mais communiquaient entre eux grâce à des microphones et partageaient les mêmes informations sur leurs écrans (les actions exécutées par l'un étaient visibles par l'autre). L'expérience s'est déroulée dans le contexte de l'utilisation d'un logiciel de dessin. L'instructeur avait sous les yeux la scène à représenter et donnait des directives au manipulateur qui ignorait ce qu'il devait dessiner. Les dialogues ont été enregistrés et filmés et leurs transcriptions orthographiques ont été faites par Ozkan. Voici un extrait de ce corpus (i=instructeur, m=manipulateur) :

- (i) alors tu prends un carré rouge euh le petit
- (i) tu mets à gauche du du rond
- (i) voilà
- (m) *geste de déplacement*
- (i) un peu plus en haut
- (i) ba c'est bon

Le projet MAREDI (MARqueurs et REprésentation des DIscours), dans lequel s'inscrivent nos travaux, vise à développer une approche et un outil d'analyse de discours qui s'appuie essentiellement sur la détection et l'interprétation de marques de surface afin d'élaborer un modèle conceptuel du discours sous la forme d'un ensemble d'états mentaux structurés (Moulin et Rousseau 1997). Pour construire automatiquement le modèle conceptuel du discours à partir de l'analyse des énoncés contenus dans un dialogue, nous développons

actuellement un système constitué de quatre modules principaux : un réseau neuronal qui sert à déterminer les types d'actes de dialogue accomplis à partir d'indices syntaxiques relevés à la surface des énoncés (Colineau et Moulin 1996 ; Colineau 1997) ; un analyseur syntaxique robuste qui doit pouvoir bien réagir aux problèmes d'extragrammaticalité qui sont fréquents dans l'oral ; et un analyseur sémantique mettant en œuvre une approche à base de cas sémantiques et ayant recours à l'utilisation de templates (Boufaden *et al.* 1997) ; et, finalement, un intégrateur qui, à partir des structures produites par les trois autres composantes, procède graduellement à la construction du modèle conceptuel du discours analysé. Dans ce qui suit, l'analyseur syntaxique robuste constitue le thème principal de nos propos : nous faisons abstraction des liens éventuels que pourrait avoir un tel analyseur avec un module de reconnaissance de la parole.

2. Analyseur syntaxique robuste

Dans le contexte de la deuxième approche, nous nous basons sur l'hypothèse que l'oral et l'écrit ont en commun une même grammaire qui est indépendante des phénomènes de périphérie de l'oral—sur la notion de périphérie, voir (Chanod 1993). Ces derniers phénomènes sont considérés comme des déformations du langage qui sont dues au caractère spontané de l'oral. De ce fait, l'utilisation d'outils conçus pour le traitement de l'écrit est envisageable. Selon ce point de vue, nous percevons les transcriptions de dialogues comme étant des textes où se mêlent les constituants bien formés et moins bien formés.

En partant de ces faits, nous proposons un système basé sur une stratégie hybride. L'idée de base consiste à élaborer un système à deux composants autonomes. Le premier composant, le noyau, effectue une analyse des constituants bien formés selon un procédé classique d'analyse de textes écrits. Tandis que le deuxième composant, le module de recouvrement, se charge de traiter les constituants moins bien formés (appelés distorsions dans la suite de l'article). Ce composant garantit la robustesse de l'analyseur syntaxique. Les deux composants interagissent au niveau de l'analyse grâce à un module de supervision, appelé aussi superviseur, qui gère les interventions de chacun des deux composants. La gestion des interventions s'effectue grâce à un système qui détecte des marques de surface identifiant une construction particulière ou une distorsion.

3. Noyau de l'analyseur syntaxique robuste

Le noyau de l'analyseur joue le rôle d'un analyseur syntaxique de l'écrit. Il a été conçu de manière à satisfaire deux conditions : traiter les constituants bien formés mais, surtout, faciliter l'intervention du module de recouvrement des distorsions. Dans cette perspective, nous avons utilisé des règles de la grammaire française (Figure 1) que nous avons extraites de Dubois (1969). La description des règles s'est effectuée selon le schéma X-barre (Chomsky 1981) : voir la Figure 1 à la page suivante. La stratégie d'analyse que nous avons implantée est une stratégie déterministe ascendante s'inspirant de la méthode LR(k) et de certains éléments de l'analyse par îlots (Satta et Stock 1991 ; Woods 1982). Le lexique utilisé est un lexique général (BDLEX¹) du français standard écrit. L'utilisation de la Théorie du Gouvernement et du Liage et le choix d'une analyse s'inspirant de la méthode LR(k) visent principalement à faciliter l'intervention du module de recouvrement des distorsions dans le processus d'analyse, comme nous l'expliquons dans les sous-sections 3.1 à 3.3.

¹ Le dictionnaire que nous utilisons pour notre analyseur syntaxique du français est BDLEX, obtenu grâce à l'aimable collaboration des gens du CLIPS de l'IMAG à Grenoble.

<p>Règles de grammaire associées au syntagme nominal N'' :</p> <p>$N'' \rightarrow \text{Det} + N'$ $N' \rightarrow (A_1'') + N' + (A_2'')$ $N' \rightarrow N' + (C'')$ $N' \rightarrow N + (P'')$</p> <p>Règles de grammaire associées au syntagme verbal V'' :</p> <p>$V'' \rightarrow (\text{Adv}_1'') + V' + (\text{Adv}_2'')$ $V' \rightarrow V + (N'') + (P'')$ $V' \rightarrow V + (C'') + (P'')$ $V' \rightarrow \text{Aux} + (N'') + (P'') + (A'')$</p> <p>Règles de grammaire associées à l'inflexion I'' :</p> <p>$I'' \rightarrow (\text{Adv}'') + I''$ $I'' \rightarrow (N'', C'') + I''$ $I' \rightarrow I + V''$ $I \rightarrow (\text{Neg}) + \text{Tps} + (\text{Aux}) + (M) + (\text{Aux})$</p>
--

Figure 1 : Extrait des règles de grammaire utilisées par le noyau.

3.1 Théorie du Gouvernement et du Liage

La théorie du Gouvernement et du Liage (Chomsky 1981) a soulevé de nombreuses critiques concernant le caractère inné de la grammaire universelle—par exemple, voir (Abeillé 1993). Néanmoins, nous avons retenu deux aspects avantageux par rapport à notre démarche : le cadre de sous-catégorisation des têtes de syntagmes et la concept de transformation. Dans notre description du noyau, nous utilisons une grammaire qui reconnaît les structures profondes (projections des cadres de sous-catégorisation des catégories majeures). Nous avons choisi de modéliser l'appareil transformationnel en tant que module autonome intervenant de la même façon que le module de recouvrement, c'est-à-dire grâce à la détection de marques de surface identifiant les différentes constructions. Cette approche vient renforcer notre vision de noyau indépendant de tous les phénomènes de périphérie. Le deuxième avantage concerne l'encodage de la structure hiérarchique de la phrase. La régularité du schéma X-barre permet le contrôle de la progression de l'analyse et facilite la détection des regroupements (de constituants) non conformes à ce schéma.

3.2 Méthode d'analyse

L'approche déterministe revêt un intérêt particulier dans notre stratégie d'analyse. En effet, dans le contexte du traitement de l'oral, le problème de l'ambiguïté de l'analyse prend encore plus d'ampleur que pour l'écrit. À l'ambiguïté syntaxique qui caractérise les langues naturelles, vient s'ajouter un deuxième problème, soit : le traitement des distorsions de l'oral. Lorsque l'analyse échoue à une étape donnée, il est difficile d'identifier la cause de cet échec. Celui-ci peut être causé par un "mauvais" choix de règle de grammaire ou encore par une distorsion. Ainsi, afin d'isoler le traitement des distorsions, nous cherchons à minimiser l'ambiguïté syntaxique.

Plusieurs analyseurs syntaxiques ont utilisé une approche déterministe de type LR(k) pour le traitement de l'écrit comme, par exemple, GLR (Tomita 1987) et PARSIFAL (Marcus 1980). Nous avons opté pour une analyse de droite à gauche, donc RL(k), car elle tend à diminuer l'ambiguïté syntaxique, comme nous l'indique notre expérimentation sur nos corpora, et elle facilite le recouvrement de certaines distorsions tels que les répétitions.

3.3 Marques associées aux transformations et aux procédures de recouvrement

La démarche utilisée pour l'identification des transformations et des procédures de recouvrement s'inspire de la méthode proposée pour l'identification des actes de dialogue, à savoir, déterminer des indices syntaxiques rendant compte des actes de dialogues (Colineau 1997). Ces indices ou marques sont des unités lexicales ou certaines formations syntaxiques. C'est la combinaison de ces marques qui permet l'identification du type d'acte de dialogue. Parallèlement à cette démarche, nous avons identifié un ensemble de marques syntaxiques (Figure 2) capables de rendre compte des différentes constructions de l'écrit ainsi que de certaines distorsions de l'oral. Les marques que nous avons relevées pour les constructions de l'écrit permettent une identification exacte de chaque construction (et donc de chaque transformation à appliquer). En ce qui concerne les heuristiques et procédures de recouvrement, nous avons pu aussi identifier des marques caractéristiques. Actuellement, les marques que nous avons relevées nous ont permis de reconnaître de manière univoque les heuristiques. Néanmoins, d'autres expérimentations sont nécessaires pour affirmer la généralité de notre démarche. Voici un exemple de marques identifiant des transformations et procédures de recouvrement :

<p>1. Transformation relative</p> <p>$L_{1}=\{T_{relative}, \text{absence d'un argument du verbe ou du sujet, détection d'un pronom relatif, prochain constituant mineur est de même catégorie syntaxique que l'élément manquant}\}$</p> <p>2. Transformation interrogative partielle</p> <p>$L_{1}=\{T_{interrogative_partielle}, \text{prosodie, absence de l'argument d'un verbe, détection d'un constituant interrogatif (Qu-)}\}$</p> <p>3. Recouvrement du bruit</p> <p>$L_{1}=\{R_{ibruit}, \text{constituant ne vérifie pas la condition d'insertion (Problème d'insertion)}\}$</p> <p>4. Recouvrement des répétitions de syntagmes et d'unités lexicales</p> <p>$L_{1}=\{R_{irépétition_de_syntagmes_ou_d'unités_lexicales}, \text{chevauchement de syntagmes de même nature (Sauf Prépositionnel), ou de catégories mineures de même nature}\}$</p>

Figure 2 : Exemples de marques de surfaces associées aux transformations et aux procédures de recouvrements.

4. Couche périphérique de l'analyseur syntaxique robuste

La couche périphérique se charge du recouvrement des distorsions. Elle est composée d'un ensemble d'heuristiques et de mécanismes de recouvrement qui interviennent à chaque fois qu'un ensemble de marques identifiant une distorsion est détecté par le superviseur. Afin de définir les procédures de recouvrement, nous avons relevé les différentes distorsions présentes dans nos corpora et elles représentent en moyenne 65% des énoncés. Quatre types

de distorsions ont été identifiés : les bruits, les effacements, les interruptions, et les répétitions. Dans ce qui suit nous présentons les procédures de recouvrement du bruit et des répétitions.

4.1 Heuristique pour le recouvrement du bruit

En général le bruit est associé aux éléments qui ne présentent aucun intérêt pour l'analyse syntaxique et qui ne peuvent pas être insérés dans l'arbre d'analyse, comme dans l'énoncé *tu mets à gauche du du rond*, où le déterminant *du* ne peut être inséré dans l'arbre d'analyse. La condition d'insertion d'une unité lexicale permet de détecter la présence de bruit. Elle est définie comme suit : *un constituant respecte la condition d'insertion s'il répond à l'une des trois conditions suivantes : 1) il peut être inséré dans le syntagme en cours de construction ; 2) il peut représenter le début d'un nouveau syntagme ; 3) il représente à lui seul un syntagme.* Ainsi, l'heuristique du recouvrement du bruit consiste à *ignorer les unités lexicales ne respectant pas la condition d'insertion.*

4.2 Heuristique du recouvrement des répétitions

Les répétitions de syntagmes sont fréquentes dans nos corpora, comme par exemple l'énoncé *maint'nant tu prends un gros rond un gros cercle*. L'heuristique que nous proposons pour le traitement des répétitions part de l'hypothèse que la dernière information répétée est celle retenue par l'allocutaire et celle sur laquelle se bâtit le reste du dialogue. Ainsi, l'heuristique que nous retenons est la suivante : *les constituants de la répétition retenus pour l'analyse sont ceux répétés en dernier lieu.* Cependant, notons que certains cas de répétitions nécessitent un traitement particulier, comme par exemple dans l'énoncé *main'tnant tu prends le carré rouge euh le petit* où il y a répétition de syntagmes nominaux. L'application de l'heuristique du recouvrement des répétitions engendre une perte d'information puisque le syntagme retenu dans l'analyse est *le petit*, qui est une construction elliptique. En effet, en éliminant le premier syntagme *le carré rouge*, qui représente le référent de *le petit*, nous risquons de ne pas pouvoir calculer l'ellipse. Dans la section suivante, nous montrons qu'il est possible de traiter ce problème grâce à la gestion de l'ordonnancement de marques.

5. Gestion de l'ordonnancement entre transformations et mécanismes de recouvrement

Le rôle du superviseur est particulièrement important, dans la mesure où nos corpora représentent des textes où se mêlent les constituants bien formés et les distorsions. De ce fait, il nous apparaît impossible d'établir un ordre figé quant à l'intervention du noyau et de la périphérie dans le processus d'analyse. Afin de mieux expliquer ce point, reprenons l'exemple *je prends le carré rouge euh le petit*, nous pouvons remarquer la présence d'une ellipse du nom et une répétition du syntagme nominal (rappelons que l'analyse se fait de la droite vers la gauche). Deux possibilités sont à envisager. La première consiste à effectuer d'abord le recouvrement de la répétition ensuite le calcul de l'ellipse, ce qui revient à analyser l'énoncé *je prends le petit*. Nous remarquons dans ce cas que l'objet de l'ellipse a été retiré de l'énoncé après l'application du recouvrement de la répétition, ce qui implique une perte d'information. La deuxième possibilité consiste à calculer d'abord l'ellipse puis à appliquer la procédure de recouvrement de la répétition de syntagme. Le résultat de cet enchaînement revient à analyser l'énoncé *je prends le petit carré rouge*, ce qui permet de corriger l'énoncé sans pour autant perdre de l'information. Nous pouvons constater que l'enchaînement des procédures joue un rôle important dans l'analyse syntaxique. En

respectant l'ordre d'occurrence des marques associées aux traitements de l'écrit et ceux de l'oral il est possible d'effectuer une analyse sans perte d'informations.

Ainsi, le superviseur a pour rôle de gérer l'ordre d'intervention du noyau et de la périphérie dans le processus d'analyse. Il mémorise toutes les marques rencontrées afin de déclencher la transformation ou la procédure de recouvrement appropriée. Lorsqu'une transformation et une procédure de recouvrement peuvent s'exécuter simultanément, c'est l'ordre d'occurrence de leurs marques qui fixe l'ordre de leur exécution. Dans la Figure 3, nous présentons l'algorithme détaillé du superviseur.

6. Premiers résultats expérimentaux

L'implémentation actuelle de notre analyseur syntaxique robuste ne couvre pas tous les phénomènes linguistiques, ni de l'écrit, ni de l'oral. La première phase de ces travaux visait d'ailleurs à démontrer la valeur potentielle de l'approche ; nous estimons avoir atteint cet objectif. Les transformations qui ont été implantées et testées sont celles des pronoms clitics, des interrogatives, des négatives et des relatives. Une partie seulement du recouvrement de toutes les distorsions possibles est actuellement en fonction : implémentation complète pour les bruits et les répétitions ; implémentation partielle pour les effacements ; et pas encore d'implémentation pour les interruptions.

Nous avons récemment évalué cette implémentation sur deux corpora, d'une centaine d'énoncés chacun. Les résultats présentés dans le Tableau 1 (ci-dessous) indiquent les pourcentages de succès pour chaque type de distorsion et cela de manière intrinsèque. Les résultats pour "zéro distorsions" (énoncés corrects) sont plus faibles qu'attendus. Après examen plus détaillé de leurs causes, nous avons constaté qu'ils étaient dus principalement des lacunes au niveau des phénomènes linguistiques couverts par le noyau de l'analyseur. Cependant, nous considérons satisfaisants les résultats obtenus pour le recouvrement des distorsions. En effet, les points importants qui ont été vérifiés sont le recouvrement des différents phénomènes de l'oral avec succès et sans perte d'information, et le respect de l'ordre d'intervention du noyau et de la périphérie dans le processus d'analyse.

En particulier, l'heuristique pour le recouvrement du bruit a prouvé son efficacité quant à l'élimination du bruit dans les énoncés puisque nous enregistrons une moyenne de 87,5% de corrections réussies. Pour le recouvrement des répétitions nous avons enregistré 86% de corrections réussies et cela sans perte d'information. Jusqu'à maintenant, il est clair que nous avons d'avantage concentré nos efforts sur le traitement des distorsions. C'est pourquoi nous reconsidérons la couverture grammaticale du noyau dans la suite immédiate de nos travaux.

Phénomène	Corpus C11	Corpus C5	Moyenne
zéro distorsion	63%	33%	48%
bruit	100%	75%	87,5%
répétition	86%	n/a	86%

Tableau 1 : Taux de succès du traitement des énoncés comportant une occurrence de chacun des phénomènes considérés par l'analyseur syntaxique robuste. 'n/a' signifie qu'il n'y avait pas d'énoncé de ce type.

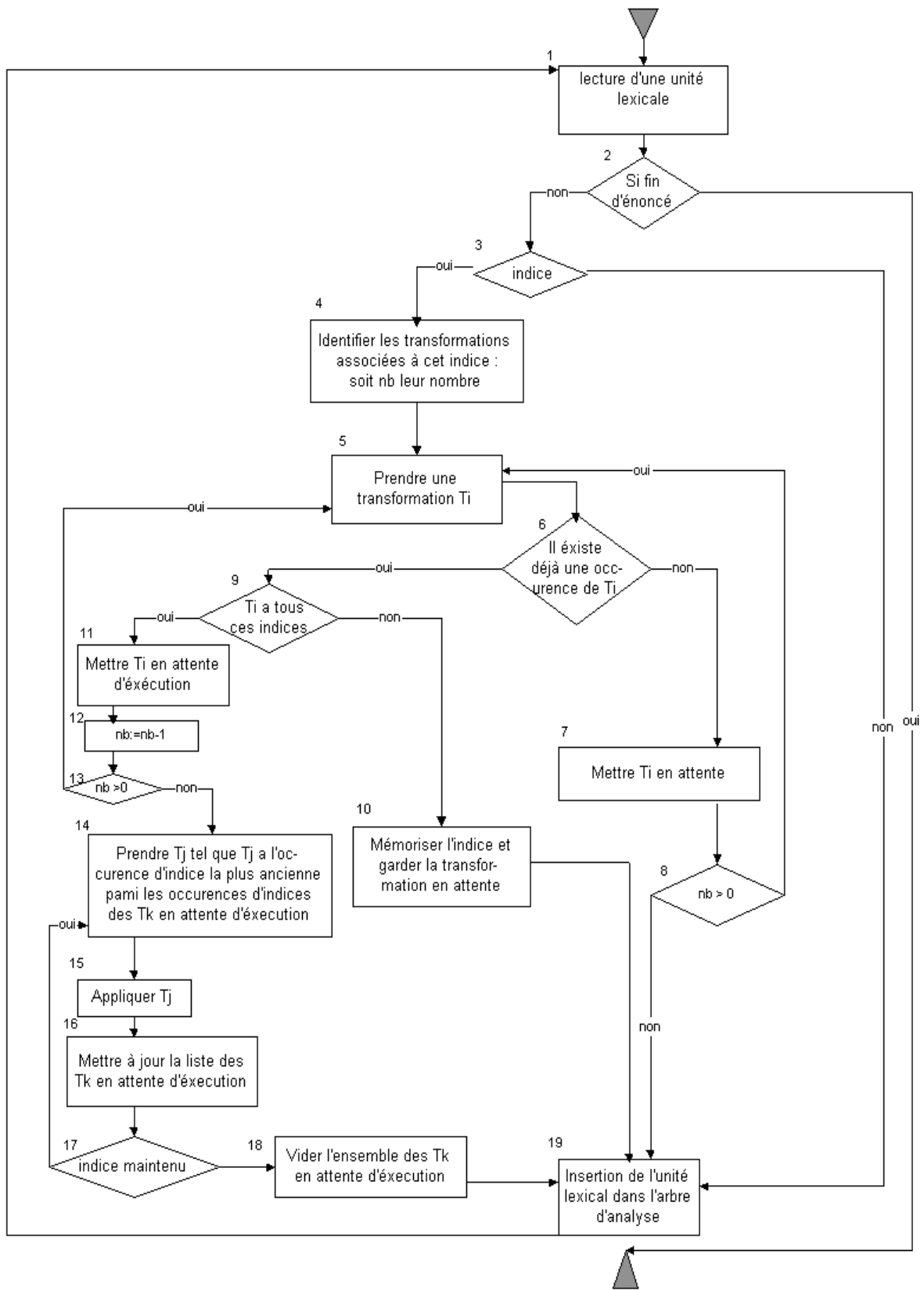


Figure 3 : Algorithme de gestion de l'ordonnement des transformations et des procédures de recouvrements.

7. Brève comparaison avec d'autres travaux

Nous signalons au lecteur deux recherches en rapport direct avec nos propres travaux. Premièrement, Langer (1990) qui s'intéresse spécifiquement à l'analyse syntaxique d'énoncés oraux dits "mal-formés" et utilise un concept de normalisation qu'il définit comme étant la relation entre un énoncé mal-formé et sa version bien formée. Pour ce faire, il augmente un analyseur de l'écrit d'un certain nombre d'heuristiques qui permettent de traiter quelques phénomènes comme les déviations morpho-syntaxiques, certaines réparations et certaines répétitions. Et deuxièmement, Nakano *et al.* (1994) qui augmentent un analyseur d'abord conçu pour l'écrit d'un ensemble de règles qui permettent d'élargir la couverture des phénomènes syntaxiques en ajoutant ceux qui sont spécifiques à la langue orale Japonaise. Par rapport à ces travaux, notre approche présente le net avantage d'établir une distinction claire, à la fois du point de vue théorique et du point de vue de la réalisation informatique, entre ce que nous désignons comme étant le noyau et la périphérie de la grammaire. De plus, le contrôle de la coordination de ces deux couches est effectué par un module dédié à cette tâche, le superviseur : il s'agit donc d'une approche qui semble plus modulaire et flexible, que celles citées ci-dessus.

8. Conclusion et travaux futurs

Nous avons présenté les principales caractéristiques de notre approche d'analyse syntaxique robuste d'énoncés transcrits de l'oral. Les caractéristiques originales en sont l'exploitation judicieuse des marques de surface ; l'adoption d'une perspective selon laquelle la langue orale est une variation de la langue écrite; et la réalisation d'un analyseur syntaxique robuste à partir des concepts de noyau et de périphérie, et d'un superviseur qui contrôle ces deux couches. Bien que nous n'ayons pas encore terminé notre évaluation, les résultats obtenus actuellement nous laissent croire à la pertinence d'une conception modulaire et complémentaire pour le traitement de l'oral.

Nos travaux actuels s'attachent principalement à augmenter la couverture de la grammaire noyau et à augmenter le nombre de distorsions traitées.

Remerciements

Nous remercions le FCAR (Fonds pour la formation de Chercheurs et l'Aide à la Recherche) ainsi que le CRSNG (Conseil de Recherche en Sciences Naturelles et Génie du Canada) pour l'aide financière apportée à nos travaux de recherche.

Références

- Abeillé A. (1993), *Les nouvelles syntaxes : grammaires d'unifications et analyse du français*, Paris : Colin, Collection Linguistique.
- Boufaden N., Delisle S., Moulin B., Gouiaa M. (1997), "Linguistique informatique et robustesse : Analyse syntaxique et sémantique pour la modélisation de dialogues finalisés", *Actes du colloque du colloque international FRACTAL'97*, pp.75-84.

- Rosé C.P., Lavie A. (1997), "An Efficient Distribution of Labor in a Two Stage Robust Interpretation Process", *In Proceedings of Empirical Methods in Natural Language Processing EMNLP'97*, Rhode Island U.S.A.
- Chanod J.P. (1993), "Problèmes de robustesse en analyse syntaxique", *Actes de la Conférence Informatique et Langue Naturelle ILN'93*, pp.223-244.
- Chomsky N. (1981), *Lectures on Government and Binding*, Dordrecht : Foris, Collection : Studies in generative grammar.
- Colineau N., Moulin B. (1996), "Un modèle connexionniste pour la reconnaissance d'actes de dialogues", *Actes de la conférence Informatique et Langue Naturelle ILN'96*, Nantes, France, pp.157-174.
- Colineau N. (1997), Étude des marqueurs discursifs dans le dialogue finalisé, Thèse de 3^e cycle de l'Université Joseph Fourier, Spécialité Sciences Cognitives, décembre 1997, 283p.
- Dubois J. (1969), *Grammaire structural du français*, Paris : Librairie Larousse, Collection Langue et langage.
- Langer H. (1990), "Syntactic Normalization of Spontaneous Speech ", *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)*, Helsinki, Finland, Vol. 3, pp.180-183.
- Marcus M.P. (1980), *A Theory of Syntactic Recognition for Natural Language*, Cambridge, Mass :MIT Press.
- Moulin B., Rousseau D. (1997), "An Approach for Modelling and Simulating Conversations", à paraître dans D. Vanderveken et S.Kubo (éditeurs), *Essays in Speech Act Theory*, John Benjamins. Aussi rapport de recherche Université Laval.
- Nakano M., Shimazu A., Kogure K. (1994), "A Grammar and a Parser for Spontaneous Speech ", *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, Kyoto, Japan, Vol. 2, pp.1014-1020.
- Ozkan N. (1994), Vers un modèle dynamique du dialogue : analyse de dialogue finalisés dans une perspective communicationnelle, Thèse de Doctorat, Institut National Polytechnique de Grenoble.
- Satta G., Stock O. (1991), "A Tabular Method for Island-Driven Context-Free Grammar Parsing", *Proceedings of AAAI-91*, pp.143-148.
- Tomita M. (1987), "An Efficient Augmented-Context-Free Parsing Algorithm", *Computational Linguistics*, Vol 13(1-2), pp.31-46.
- Woods W.A. (1982), "Optimal Search Strategies for Speech Understanding Control", *Artificial Intelligence* 18, pp.295-326.