

Problèmes de Désambiguïsation Interactive en TAO Personnelle.

H. BLANCHON

GETA, Institut IMAG
(UJF & CNRS)
BP 53X, 38041 Grenoble Cedex, France
e-mail : blanchon@imag.fr

Colloque de Mons, 25—27 avril 1991

“L’environnement traductionnel : la station de travail du traducteur de l’an 2000”

Résumé.

Nous nous plaçons ici dans un contexte de recherche autour d'un système de "TAO personnelle". L'utilisateur d'un tel système est un rédacteur unilingue qui souhaite pouvoir rédiger dans sa langue maternelle et être traduit pour un coût et dans un délai raisonnables. Pour la réalisation, on doit respecter trois contraintes très fortes. D'abord, la qualité de la traduction obtenue doit être suffisante pour ne pas devoir réviser (le rédacteur n'est pas un réviseur). Ensuite, on ne peut pas utiliser de base de connaissances pour obtenir automatiquement une compréhension explicite du document (il est impossible de connaître a priori le domaine sémantique de celui-ci). Enfin, on ne peut pas contraindre le rédacteur à utiliser un langage contrôlé, ni même un sous-langage déterminé (la rédaction doit être la plus libre possible).

Le respect des deux dernières contraintes ne permet pas la réalisation d'un système autonome de traduction avec lequel les analyses (syntaxique, logico-sémantique) du texte en langue source reflètent l'intention de l'auteur. Le système produira en effet, à certains stades du traitement, des représentations qui refléteront des ambiguïtés que les connaissances grammaticales “fortes” et les connaissances sémantiques “faibles” du système ne permettront pas de résoudre automatiquement. Or, si l'on veut obtenir une traduction de qualité, il faut obtenir une représentation désambiguïsée de ce que l'on veut traduire. La solution que nous envisageons au GETA est de réaliser, dans le cadre du projet LIDIA, la maquette d'un système de traduction qui utilise le dialogue avec le rédacteur pour obtenir une représentation désambiguïsée de ce que l'on veut traduire. Dans un tel système, le dialogue occupe donc une place privilégiée puisque c'est lui qui va permettre d'obtenir une traduction de qualité.

Dans le présent article, nous décrirons comment, à partir d'un résultat d'analyse, nous percevons l'organisation des traitements qui nous permettront d'obtenir une représentation débarrassée de trois types d'ambiguïtés. Pour chacun de ces types, nous montrerons quel genre de dialogue peut proposer le système. Nous verrons alors que l'on peut adapter la forme du dialogue à l'utilisateur. Pour un utilisateur novice on peut envisager un dialogue très explicite. Car, en effet, nous allons parfois surprendre le rédacteur en lui posant certaines questions. Il faut alors que l'explicitation du phénomène linguistique soit très précise et suffisamment simple pour être comprise sans avoir à faire appel à des connaissances linguistiques pointues. Ce type de dialogue doit aider le rédacteur à reconnaître les schémas ambigus pour éviter leur emploi ou bien être conscient qu'il devra répondre à des questions les concernant. Ensuite, avec une certaine habitude et une amélioration de ses compétences, il pourra choisir un mode de dialogue plus synthétique, qui rendra la tâche de désambiguïsation plus rapide.

Situation : l'analyseur de LIDIA.

Les traitements interactifs dans le projet LIDIA se divisent en deux catégories : les traitements de standardisation et les traitements de clarification [Blanchon 90, Boitet 90]. Lorsque les traitements de standardisation ont été effectués sur une unité de traduction, une première étape d'analyse produit une structure "concrète" qui est un arbre décoré "multi-niveaux" [Vauquois 78]. La structure est dite "concrète", car on y retrouve directement le texte (mot des feuilles). Elle est "multi-niveaux" car les nœuds portent des décorations complexes qui représentent divers niveaux d'interprétation : le niveau des classes syntaxiques et des classes syntagmatiques, le niveau des fonctions syntaxiques et enfin le niveau des relations logiques et sémantiques. Une première phase de désambiguïsation interactive s'applique à cette structure.

Pour préciser les choses, voici une brève description de ces divers niveaux d'interprétation qui permettra de mieux comprendre notre stratégie de désambiguïsation.

Le niveau des classes syntaxiques (classes terminales telles que : "NOM COMMUN", "ADJECTIF", "PRÉPOSITION", etc) et des classes syntagmatiques (classes non terminales telles que : "GROUPE NOMINAL", "PHRASE INFINITIVE", "PHRASE VERBALE", etc), donne le parenthésage de la phrase en groupes syntagmatiques. Les feuilles de l'arbre portent les classes syntaxiques et les nœuds non terminaux portent les classes syntagmatiques.

Le niveau des fonctions syntaxiques donne le rôle syntaxique de chaque nœud dans le groupe auquel il est attaché directement (par exemple : "GOUVERNEUR", "RÉGISSEUR", "SUJET", etc). Avec ce niveau, on peut très simplement transformer l'arborescence en une arborescence de "dépendance syntaxique".

Le niveau des relations logiques et sémantiques peut être considéré comme celui de la construction logique de la phrase, en mettant en évidence les places des arguments attachés aux unités lexicales prédicatives et donnant l'interprétation sémantique des compléments et circonstanciels (par exemple : "DÉTERMINATION", "CIRCONSTANCIEL INSTRUMENTAL", "PREMIER ARGUMENT D'UNE UNITÉ LEXICALE PRÉDICATIVE", etc).

L'arbre d'analyse de la phrase "l'élève calcule cette intégrale par la méthode de le rés" qu'il est classiquement calculé au GETA, est fourni pour clarification.

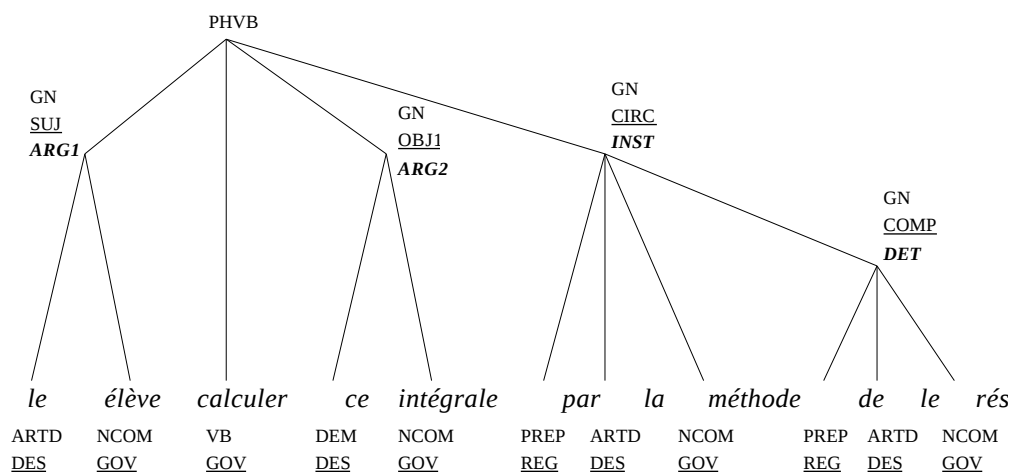
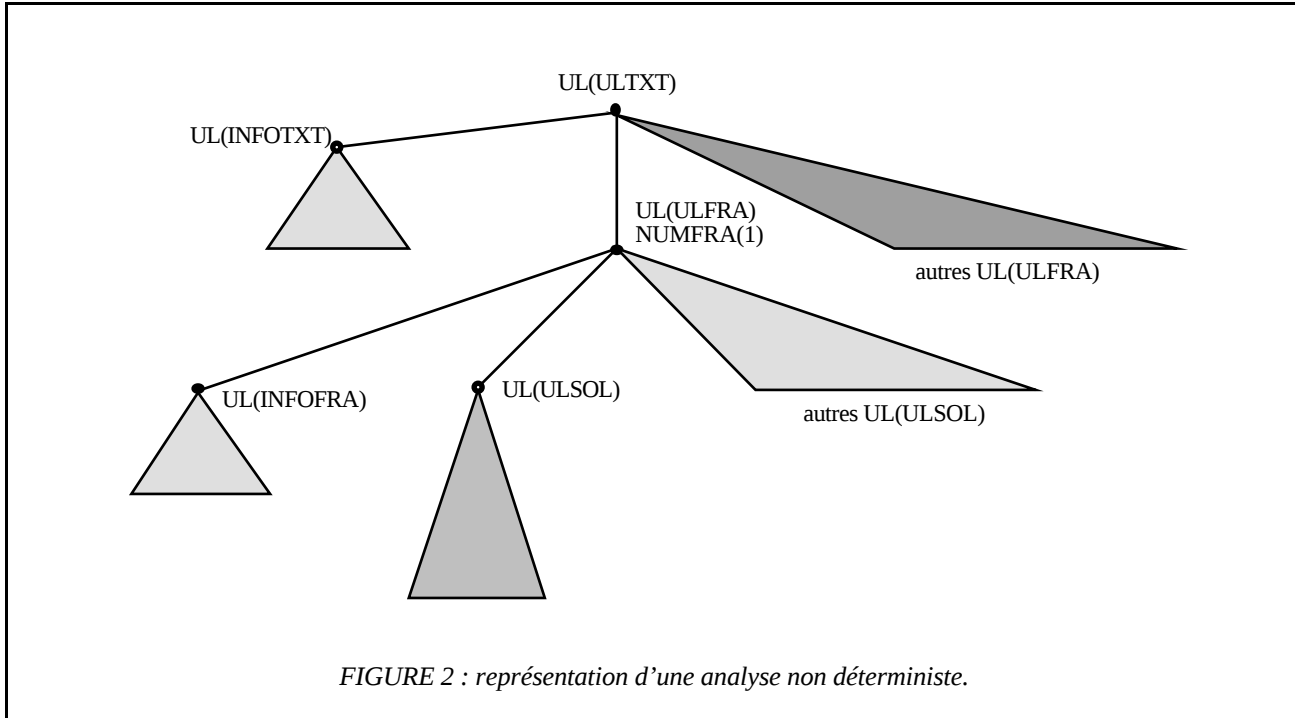


FIGURE 1 : un arbre d'analyse multi-niveaux

légende : NIVEAU 1, NIVEAU 2, NIVEAU 3

L'analyseur utilisé dans le cadre du projet LIDIA est un analyseur non-déterministe "multiple" qui, pour une phrase donnée, fournit toutes les analyses vérifiant le modèle syntagmatique, syntaxique et logico-sémantique de l'analyseur. La figure 2 montre la forme d'une analyse produite pour une unité textuelle selon le modèle non déterministe. Pour la première phrase, il y a différentes analyses : le nœud père "UL(ULFRA),NUMFRA(1)" possède plusieurs fils UL(ULSOL) et, sous chaque nœud UL(ULSOL), on dispose d'un arbre d'analyse licite de la phrase.



On impose aussi, par rapport au modèle habituel, une contrainte supplémentaire à l'analyseur : le mot qui appartient à la décoration d'une feuille de l'arbre (*en italique dans la figure 1*) doit être une occurrence de la phrase analysée, et la projection du mot des feuilles de l'arbre selon cette composante mot de la décoration fournit la phrase analysée. Le respect de cette contrainte nous permet d'avoir un processus de génération des questions plus simple à mettre en œuvre et donc moins coûteux en logiciel et en temps. Le processus de génération du dialogue étant exécuté par la station de rédaction, on perçoit l'intérêt qu'il y a à minimiser l'importance des tâches à réaliser pour que leur l'exécution soit la plus transparente possible pour l'utilisateur. Il ne faut pas que des tâches de fond trop importantes ralentissent le travail du rédacteur de façon significative.

Pour procéder à la désambiguïsation interactive en utilisant le résultat d'analyse comme support pour la préparation des questions, nous avons choisi d'utiliser trois propriétés qui permettent, pour une phrase, de différencier les divers arbres obtenus. Ces trois propriétés sont : l'étiquetage en classe syntaxique du mot des feuilles, l'étiquetage en fonction syntaxique (qui induit parfois des géométries différentes) et l'étiquetage en relation logique et sémantique des nœuds non-terminaux de l'arbre. On va maintenant donner une définition de ces trois propriétés agrémentée d'exemples.

Propriétés choisies de la structure produite par l'analyseur.

- La première propriété qui différencie les arbres d'analyse licite d'une phrase est l'étiquetage en classe syntaxique du mot des feuilles (désigné dans la suite par étiquetage en classe syntaxique) qui est, rappelons-le, le premier niveau d'interprétation de la structure profonde.

Le schéma suivant (figure 3) montre les différences entre les deux arbres produits par l'analyse de la phrase "Devant cette somme, il hésite" dans laquelle 'devant' peut être interprété comme gérondif du verbe devoir ou bien comme préposition.

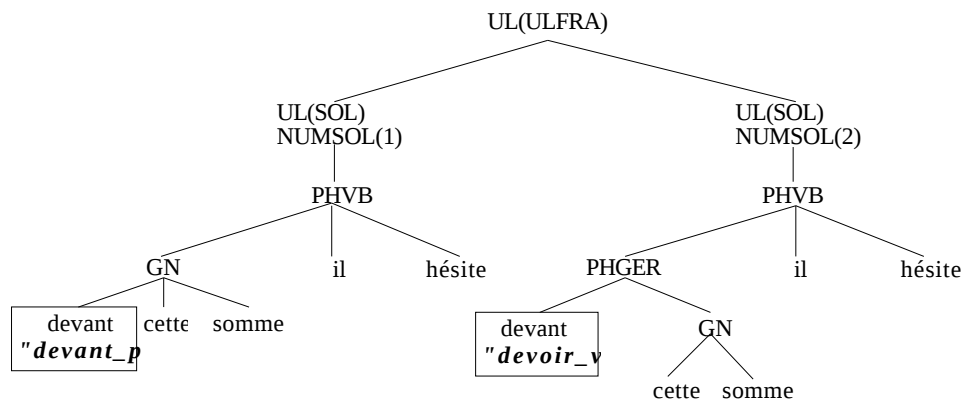


FIGURE 3 : Résultat d'analyse de la phrase "Devant cette somme, il hésite".
On observe un problème au niveau du choix de la classe syntaxique pour l'occurrence 'devant'.

Comme on peut le voir, les arbres ont une géométrie différente (étiquetage en fonction syntaxique différent), mais ce qui nous intéresse ici c'est seulement la différence au niveau de l'étiquetage en classe syntaxique de l'occurrence 'devant'.

• La seconde propriété qui différencie les arbres d'analyse licite d'une phrase est la géométrie et l'étiquetage en fonction syntaxique des nœuds non-terminaux (désigné dans la suite par géométrie et étiquetage en fonction syntaxique). Les arbres ont maintenant tous le même étiquetage en classe syntaxique. Cette propriété correspond au deuxième niveau d'interprétation de la structure profonde.

Le schéma suivant (figure 4) montre les différences entre les deux arbres produits par l'analyse de la phrase "L'évolution de la structure du réseau et des investissements est étudiée" où on peut comprendre que l'on étudie l'évolution de la structure du réseau **et** l'évolution des investissements (NUMSOL(1)) ou bien que l'on étudie l'évolution de la structure du réseau **et** l'évolution de la structure des investissements (NUMSOL(2)).

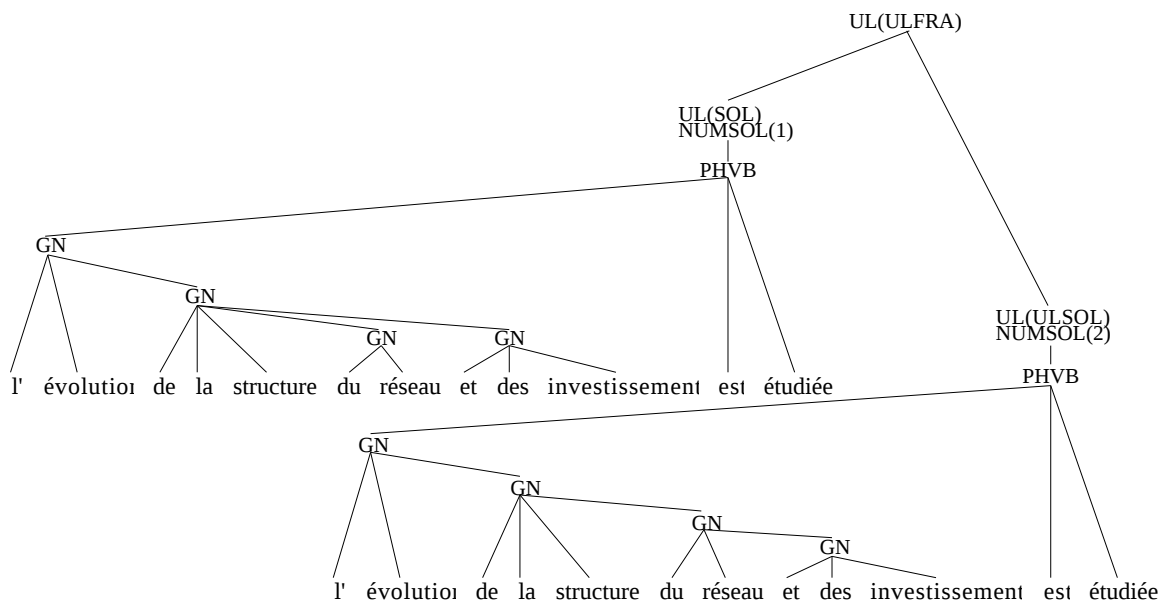
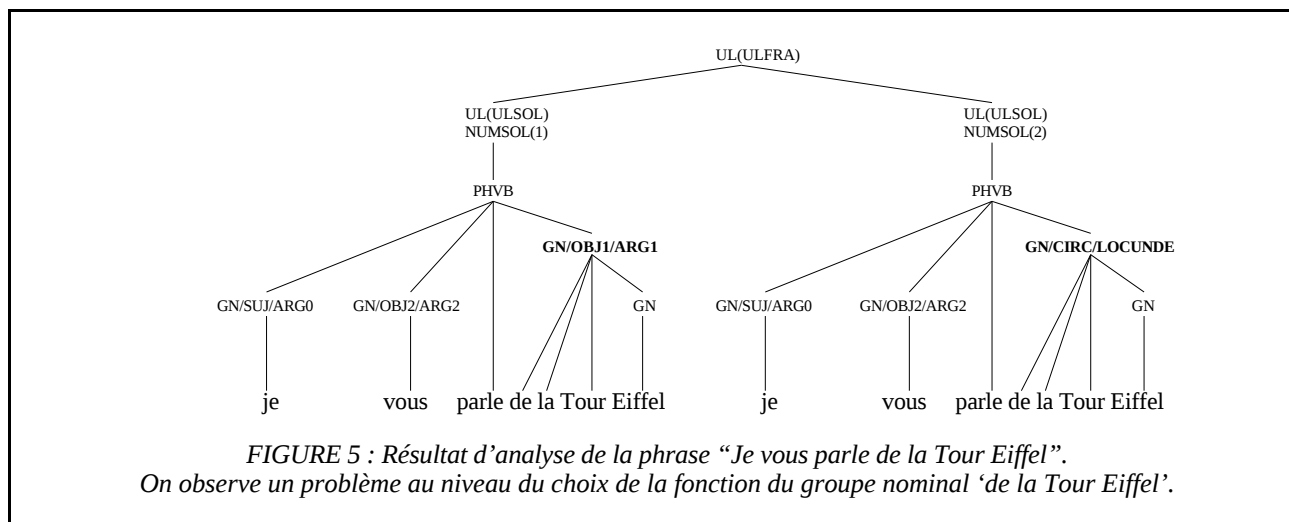


FIGURE 4 : Résultat d'analyse de la phrase "L'évolution de la structure du réseau et des investissements est étudiée".
On observe un problème au niveau du choix de la géométrie des arbres.

• La troisième propriété qui différencie les arbres d'analyse licite d'une phrase est l'étiquetage en relation logique et relation sémantique des nœuds non-terminaux (désigné dans la suite par étiquetage en relation logico-sémantique). Elle correspond au troisième niveau d'interprétation de la structure profonde.

Le schéma suivant (figure 5) montre les différences entre les deux arbres produits par l'analyse de la phrase "Je vous parle de la Tour Eiffel" où on peut comprendre que l'on nous parle depuis la Tour Eiffel (NUMSOL(2)) ou bien que l'on nous parle au sujet de la Tour Eiffel (NUMSOL(1)).



Après avoir décrit les propriétés qui nous intéressent, montrons comment, en exploitant leur présence dans le résultat d'analyse d'une phrase, nous pouvons résoudre certains types d'ambiguïtés.

Intérêt des propriétés choisies.

Relation entre propriété et type d'ambiguïté.

Si l'analyse indique un problème d'étiquetage en classe syntaxique, c'est l'ambiguïté de classe syntaxique que nous avons à résoudre. Cela signifie que l'analyseur, pour une occurrence donnée, n'a pas su déterminer quelle est sa classe syntaxique. Par exemple, dans la phrase "Le vétérinaire prend la patte de l'oiseau et la bague", on ne sait pas, a priori, si l'occurrence 'bague' joue le rôle d'un verbe (la est alors pronom) ou d'un substantif (la est alors article).

Par contre, si l'analyse indique un problème de géométrie et étiquetage en fonction syntaxique, c'est l'ambiguïté de structure de constituants grammaticaux que nous devons résoudre. Cette ambiguïté de constituants grammaticaux est essentiellement liée à deux problèmes : la coordination et la subordination. La phrase "Il prend des classeurs et des cahiers noirs" présente un problème de géométrie pure, puisque l'on ne sait pas s'il faut comprendre que seuls les cahiers sont noirs ou bien que les classeurs sont noirs et les cahiers aussi. Pour la subordination, on peut citer la phrase : "Il est inscrit dans une école de cuisine lyonnaise" où on ne sait pas s'il s'agit d'une 'école située à Lyon où on apprend la cuisine' ou bien d'une 'école sise on ne sait où, où on apprend la cuisine de Lyon'.

Enfin, lorsque l'analyse indique un problème d'étiquetage en relation logico-sémantique, on a à résoudre deux types de problèmes : un problème de place des arguments attachés aux unités lexicales prédicatives (ex : "quel hôtel gère ce bureau ?"), ou bien un problème d'interprétation sémantique des compléments et circonstants. Comme exemple de problème d'interprétation, on peut citer la phrase "Je vous parle de la Tour Eiffel" où 'de la Tour Eiffel' peut être considéré comme circonstant de localisation (depuis la Tour Eiffel), ou comme premier argument et premier objet du verbe parler (au sujet de la Tour Eiffel).

Combinaisons possibles des propriétés.

Bien sûr, une phrase peut donner lieu à plusieurs ambiguïtés, de même type, ou de types différents. Tous les croisements sont effectivement possibles, comme le montrent les exemples suivants.

- différents étiquetages en classes syntaxiques du mot des feuilles :

– avec différentes géométries et étiquetages en fonctions syntaxiques et avec différents étiquetage en relations logico-sémantiques :

Il nous parle de l'école de cuisine lyonnaise, prend le livre et le lit.

On voudrait savoir comment interpréter

'de', 'de' : 'depuis', 'au sujet de'

'école de cuisine lyonnaise' : 'école lyonnaise', 'cuisine lyonnaise'

'le lit' : 'pronom + verbe', 'article + substantif'

– avec différentes géométries et étiquetages en fonctions syntaxiques et un seul étiquetage en relations logico-sémantiques :

Il regarde le livre de l'école de cuisine lyonnaise et le lit.

– avec une seule géométrie et étiquetage en fonctions syntaxiques et avec différents étiquetages en relations logico-sémantiques :

Il observe la photo de Paul et la classe.

On voudrait savoir comment interpréter

'de' : 'prise par Paul', 'représentant Paul', 'appartenant à Paul'

'la classe' : 'pronom + verbe', 'article + substantif'

– avec une seule géométrie et étiquetage en fonctions syntaxiques et un seul étiquetage en relations logico-sémantiques :

Le vétérinaire prend la patte de l'oiseau et la bague.

On voudrait savoir comment interpréter

'la bague' : 'pronom + verbe', 'article + substantif'

- différentes géométries et étiquetages en fonctions syntaxiques :

– avec différents étiquetages en relations logico-sémantiques :

Il parle de l'école de cuisine lyonnaise.

On voudrait savoir comment interpréter

'de' : 'depuis', 'au sujet de'

'école de cuisine lyonnaise' : 'école lyonnaise', 'cuisine lyonnaise'

– avec un seul étiquetage en relations logico-sémantiques :

Il est inscrit dans une école de cuisine lyonnaise.

- différents étiquetages en relations logico-sémantiques seulement :

Quel hôtel gère ce bureau ?

On voudrait savoir comment interpréter les arguments du verbe

'un hôtel gère-t-il le bureau' ou 'le bureau gère-t-il un hôtel'

Définition d'un ordre d'utilisation des propriétés.

- Comme plusieurs propriétés peuvent apparaître dans un résultat d'analyse, nous allons exploiter, si elles se présentent, les propriétés dans l'ordre dans lequel nous les avons définies. C'est à dire l'étiquetage en classes syntaxiques, puis la géométrie et l'étiquetage en fonctions syntaxiques et enfin l'étiquetage en relations logico-sémantiques. La figure 6 donne une idée de la façon dont nous allons procéder. Le but du processus que nous allons décrire est la production d'une série de questions organisées hiérarchiquement qui permettent d'obtenir, en partant d'une liste d'analyses licites, l'analyse qui correspond aux intentions du rédacteur. Cet ordre d'exploitation des propriétés pourra, si nécessaire, être remis en question car, comme le montre la figure 6, l'organisation des traitements est fortement modulaire.

- On est libre d'examiner la présence de chacune des propriétés dans n'importe quel ordre. Mais il nous semble que l'organisation proposée a quelques fondements psychologiques et ergonomiques.

Pour le point de vue psychologique, lorsque l'on se place dans la situation de levée interactive d'ambiguïté, le rédacteur sera amené à réexaminer son texte sans se servir du contexte, comme le fait le système. Supposons par exemple qu'il soit occupé à rédiger un document qui parle du classement d'un certain nombre de clichés représentant des acteurs célèbres ou en train de le devenir et que dans un tel contexte il écrive la phrase "Il examine la photo de Pierre Latouche et la classe". Dans ce texte, cette phrase ne pose aucun problème d'interprétation. Mais, sortie de tout contexte, elle est ambiguë, puisqu'on peut comprendre qu'il classe des photos ou qu'il examine une photo et une classe (un groupe d'élèves par exemple).

Hors-contexte, la fabrication du sens (par un humain) semble procéder par restriction des classes syntaxiques des termes, puis des géométries et relations de coordination et subordination qu'ils établissent ou non avec d'autres termes et enfin, par la qualité de ces relations. D'un point de vue psychologique, il semble donc convenable de procéder de manière analogue.

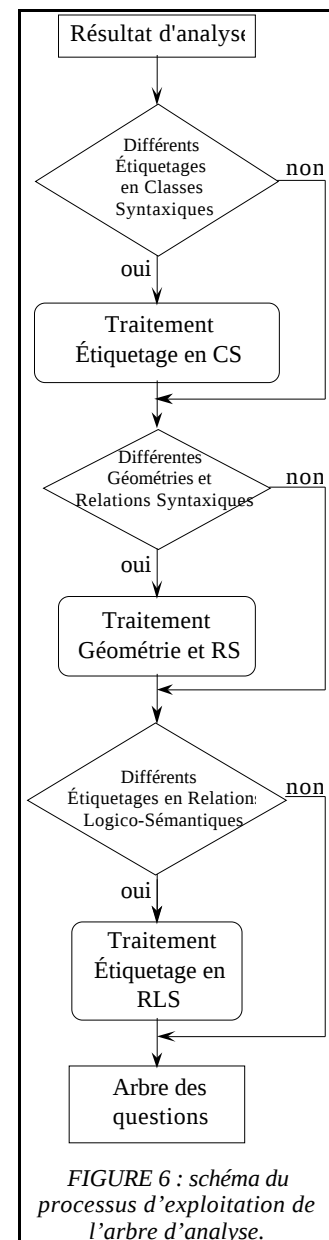
Le point de vue ergonomique renforce cette idée : il est préférable de réussir rapidement la désambiguïsation en posant les questions par ordre d'immédiateté de leurs réponses. C'est dire que la sélection de la bonne analyse doit procéder par "filtrages" successifs en utilisant d'abord des critères simples et des questions courtes, et ensuite seulement des critères plus fins et donc des questions plus complexes.

Ainsi, pour trouver la bonne interprétation de la phrase "Il atteint la grange et la ferme", il est plus naturel de demander si l'occurrence 'ferme' doit être comprise comme le verbe 'fermer' ou comme substantif que de demander si 'et la ferme' doit être compris comme un groupe nominal coordonné au verbe atteindre ou bien comme une phrase verbale coordonnée au verbe atteindre. Il est donc normal de commencer par chercher le bon étiquetage en classes syntaxiques puis la bonne géométrie et le bon étiquetage en fonctions syntaxiques, et enfin le bon étiquetage en relations logico-sémantiques.

Utilisation des propriétés choisies en désambiguïsation.

Généralités

Le système partitionnera donc l'ensemble des solutions en trois étapes successives, d'abord selon l'étiquetage en classes syntaxiques, puis selon la géométrie et l'étiquetage en relations syntaxiques et enfin, selon l'étiquetage en relations logico-sémantiques. On aboutira donc à un arbre des questions dont les niveaux successifs correspondront à cette organisation tripartite.



La figure 7 donne un exemple de découpage en classes et montre quel est l'arbre de décision que l'on peut construire pour hiérarchiser les questions.

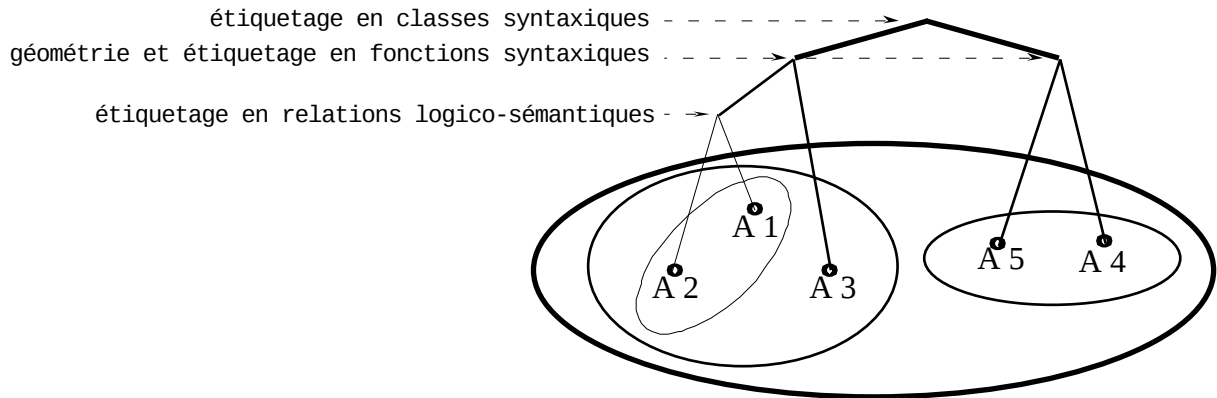


FIGURE 7 : Arbre de décision permettant de sélectionner la bonne analyse en fonction de la partition des analyses

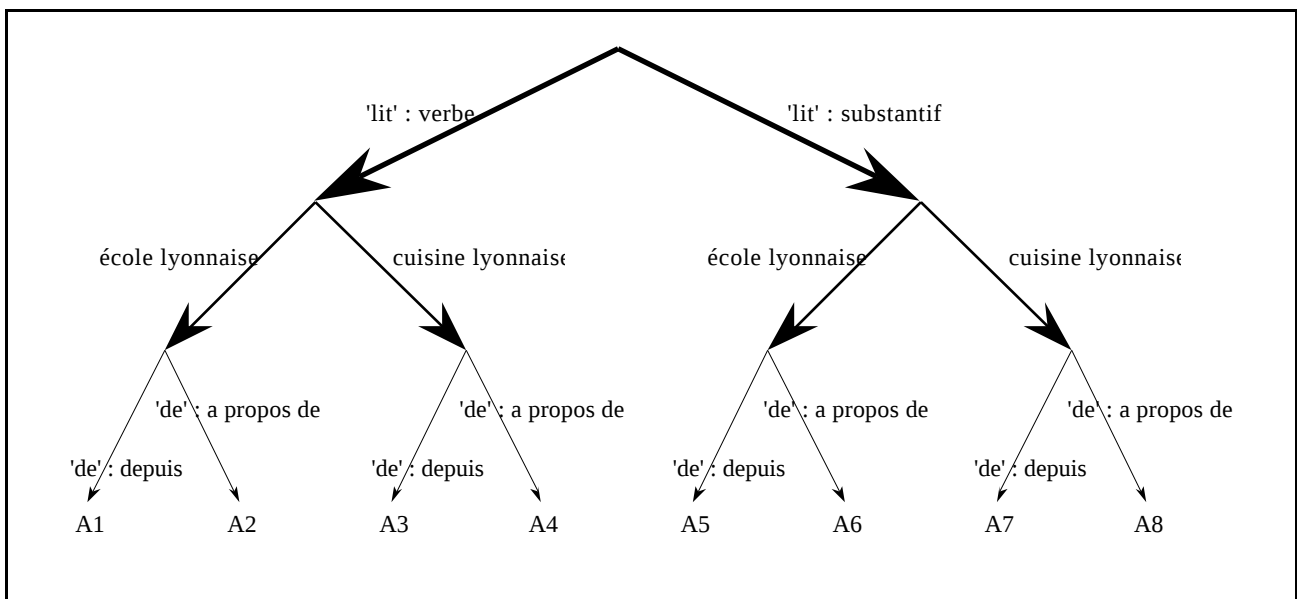
Cette figure montre que les analyses A 1, A 2 et A 3 ont le même étiquetage en classes syntaxiques et que A 1 et A 2, tout en ayant le même étiquetage en fonctions syntaxiques ont des étiquetages en relations logico-sémantiques différents ; A 3, lui, a un autre étiquetage en fonctions syntaxiques. De même les analyses A 4 et A 5 ont le même étiquetage en classes syntaxiques, mais des étiquetages différents en relations syntaxiques.

Illustrons le processus de désambiguisation sur un exemple.

Exemple.

Examinons la phrase “Le professeur de l'école de cuisine lyonnaise parle de la Tour d'Argent, prend le menu et le lit”. Cette phrase a seulement la prétention d'être un exemple démonstratif.

- Sans factorisation, l'analyse de cette phrase fournit huit solutions.
- On a schématiquement l'arbre de questions suivant :



- Pour les questions on a les dialogues suivants :

— Pour la résolution des ambiguïtés de classes.

On propose trois types de dialogue :

Présentation grammaticale explicite du problème, pour se familiariser avec le système.

problème de classe	
L'analyseur a rencontré un problème d'identification de classe dans la phrase : Le professeur de l'école de cuisine lyonnaise parle de l d'Argent, prend le menu <u>dit</u>	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒	<u>lit</u> : verbe 'lire' (il <u>lit</u>) <u>le</u> : pronom.
☐	<u>lit</u> : substantif 'lit' (nom masculin) <u>le</u> : article.
OK	

Paraphrasage de la partie de la phrase qui pose problème, l'explication grammaticale est implicite.

problème de classe	
L'analyseur a rencontré un problème d'identification de classe dans la phrase : Le professeur de l'école de cuisine lyonnaise parle de l d'Argent, prend le menu <u>dit</u>	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒	Il <u>le</u> <u>lit</u> (peut-être 'le menu ?).
☐	Il prend <u>le</u> <u>lit</u>
OK	

Projection des classes syntaxiques de la partie de la phrase qui pose problème, destinée à un utilisateur qui a voulu se familiariser avec une présentation grammaticale des problèmes.

problème de classe	
L'analyseur a rencontré un problème d'identification de classe dans la phrase : Le professeur de l'école de cuisine lyonnaise parle de l d'Argent, prend le menu <u>dit</u>	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒	 et <u>le</u> (pronom) <u>lit</u> (verbe).
☐	 et <u>le</u> (article) <u>lit</u> (substantif).
OK	

— Pour la résolution des ambiguïtés de constituants grammaticaux.

On peut proposer des présentations de type paraphrastique qui tiennent plus ou moins compte du contexte.

problème de subordination	
L'analyseur a rencontré un problème de subordination dans la phrase : Le professeur <u>décôle de cuisine lyonnaise</u> de la Tour d'Argent, prend le menu et le lit.	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒ (école lyonnaise) de cuisine.	
☐ école de (cuisine lyonnaise).	
OK	

problème de subordination	
L'analyseur a rencontré un problème de subordination dans la phrase : Le professeur <u>décôle de cuisine lyonnaise</u> de la Tour d'Argent, prend le menu et le lit.	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒ école lyonnaise.	
☐ cuisine lyonnaise.	
OK	

Ou une présentation plus compacte qui exige peut être un peu plus d'expérience du système.

problème de subordination	
L'analyseur a rencontré un problème de subordination dans la phrase : Le professeur <u>décôle de cuisine lyonnaise</u> de la Tour d'Argent, prend le menu et le lit.	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒ Il parle de (l'école de cuisine) lyonnaise.	
☐ Il parle de l'école (de cuisine lyonnaise).	
OK	

— Pour la résolution des ambiguïtés de relation logico-sémantique.

On peut ici proposer des niveaux très différents de dialogue :

Un niveau où on remplace l'occurrence qui permet de mettre en évidence la place des arguments ou d'introduire les compléments ou les circonstants. Ici, il s'agit de remplacer 'de' qui permet d'introduire, soit l'argument 1 du verbe parler soit un circonstant de localisation.

problème de relation logico-sémantique	
L'analyseur a rencontré un problème de relation dans la phrase : Le professeur de l'école de cuisine lyonnaisedis la parole d'Argentprend le menu et le lit.	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒	Il parle <u>au sujet de</u> la Tour d'Argent.
☐	Il parle <u>depuis</u> la Tour d'Argent.
OK	

Un niveau qui précise en plus l'étiquetage de l'occurrence 'de' qui permet d'apprendre les représentations du système.

problème de relation logico-sémantique	
L'analyseur a rencontré un problème de relation dans la phrase : Le professeur de l'école de cuisine lyonnaisedis la parole d'Argentprend le menu et le lit.	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒	Il parle <u>au sujet de</u> la Tour d'Argent. <u>de</u> + ARGument1 du verbe parler
☐	Il parle <u>depuis</u> la Tour d'Argent.. <u>de</u> + CIRConstant de LOCalisation du verbe parler
OK	

Un niveau pour un utilisateur spécialiste.

problème de relation logico-sémantique	
L'analyseur a rencontré un problème de relation dans la phrase : Le professeur de l'école de cuisine lyonnaisedis la parole d'Argentprend le menu et le lit.	
On propose les possibilités suivantes, sélectionnez la bonne.	
☒	Il parle (<u>de</u> la Tour d'Argent) {OBJ1/ARG1}.
☐	Il parle (<u>de</u> la Tour d'Argent) {CIRC/LOCUNDE}.
OK	

Conclusion.

Nous avons défini trois propriétés de la structure arborescente concrète, mais profonde (en fait, multi-niveaux) produite par un analyseur non déterministe (dans le formalisme du GETA). Nous avons vu comment exploiter ces propriétés dans le cadre d'un système de "TAO personnelle" pour produire un dialogue de désambiguïsation qui soit réglable en fonction des compétences de l'utilisateur. Les ambiguïtés que l'on résout sont, on le rappelle, les ambiguïtés de classe syntaxique, les ambiguïtés de coordination et de subordination, et enfin les ambiguïtés de relation logique et sémantique. Mais, pour obtenir une représentation totalement désambiguïsée il restera d'autres problèmes à régler.

C'est ainsi qu'il faut encore résoudre les problèmes de polysémies (pures et induites par la langue cible), les problèmes de référence pronominale ou adjectivale et les problèmes d'ellipse, soit pure (toutes les fois qu'un terme de phrase ou de groupe n'est pas repris, souvent dans la coordination ou la comparaison), soit dans les termes composés (lorsque l'on ne reprend pas complètement un terme composé pour abrégé ou alléger le texte, le terme est remplacé par son gouverneur).

Les phases de désambiguïsation qui traitent de ces problèmes sont réalisées soit en amont soit en aval des traitements que nous avons décrits ici. Pour information, on peut dire que pour les polysémies, la notation du problème sera inscrite dans l'arbre dans l'étiquette de sens associée à chaque occurrence, et les informations nécessaires à la désambiguïsation seront fournies par la base lexicale. Pour les problèmes de référence pronominale, il faudra conserver une liste de toutes les occurrences qui peuvent être reprises en associant à chacune d'elles des informations grammaticales (genre, nombre ...) et des informations lexicales provenant du dictionnaire (les rections syntaxiques et traits sémantiques). Ces informations permettront de retrouver les occurrences pertinentes candidates à la reprise. Quant au problème de l'ellipse dans les termes composés, une liste des termes composés déjà rencontrés ou des informations à propos du domaine du document permettront de trouver une liste des candidats possibles.

-O-O-O-O-O-O-

Références.

- BLANCHON H. (1990b) *HyperAriane : architecture de la partie "micro" d'un système de TAO personnelle*. Mémoire de DEA, juin 1990, 63p.
- BLANCHON H. (1990a) *LIDIA-1 : Un prototype de TAO personnelle pour rédacteur unilingue*. Proc. Avignon-90, conférence spécialisée Le Traitement des Langues Naturelles et ses Applications, 28 mai-1 juin, 1990, 51-60.
- BOITET Ch. (1990b) *Vers la TAO personnelle : le projet LIDIA du GETA*. À paraître dans la tribune sur les industries de la langue, 12 p.
- BOITET Ch. (1990a) *Towards Personal MT : on some aspects of the LIDIA project*. Proc. COLING-90, Helsinki, 20-25 août 1990, 10 p.
- BOITET Ch. (1989b) *Speech Synthesis and Dialog Based Machine Translation*. Proc. ATR Symp. on Basic Research for Telephone Interpretation, Kyoto, Dec. 1989, 6-5-1—6-5-22.
- BOITET Ch. (1989a) *Motivation and Architecture of the LIDIA project*. Proc. MTS-89, Munich, 16-18 août 1989, 5 p.
- BOITET CH., BLANC E. /eds. (1990) *DBMT-90 : Post-COLING seminar on Dialogue-Based MT (Machine Translation of/with Dialogues)*, Le Sappey (France), August 26-28, 1990, 328 p.
- BROWN R. D., NIRENBURG S. (1990) *Human-Computer Interaction for Semantic Disambiguation*. Proc. COLING-90, Helsinki, 20-25 août 1990, 6 p.
- HUTCHINS W. J. (1989) *Out of the shadows, A retrospect of machine translation in the eighties*. Presented at "Computer and Translation 89", Tbilissi, Georgia, November-December 1989, 13p.
- LEPAGE Y. (1988) *Ambiguities and Second Generation MT Systems*. First European Conference on Information Technology for Organisational Systems. Athens, May 16-20, 1988, 6 p.

- MARUYAMA H., WATANABE H., OGINO S, (1989) *An Interactive Japanese Parser for Machine Translation*. Proc. COLING-90, Helsinki, 2-25 août 1990, 7p.
- PECKHAM J (ed) (1989) *Recent Developments and Applications of Natural Language Processing*. Unicom Seminars Limited, dans la collection Applied Information Technology Reports, 300 p.
- TOMITA M. (1985) *Feasibility Study of Personal/Interactive Machine Translation System*. Center for Machine Translation, Carnegie Mellon University, May 1985, 13 p.
- TOMITA M. (1984) *Disambiguating Grammatically Ambiguous Sentences by Asking*. Proc. COLING-84, 476-480.
- VAUQUOIS B. (1978) *Description de la structure intermédiaire*. Communication présentée au colloque de Luxembourg, 17 et 18 avril 1978, 25 p.
- WHITELOCK P. J., WOOD M. M., CHANDLER B. J., HOLDEN N., HORSFALL H. J. (1986) *Strategies for Interactive Machine Translation : the experience and implications of the UMIST Japanese project*. Proc. COLING-86, 329-334.
- ZAJAC R. (1986) *Etude des Possibilités d'Interaction Homme-Machine dans un Processus de Traduction Automatique*. Thèse, INPG, Grenoble, juillet 1986, 259 p.
- ZAJAC R. (1988) *Interactive Translation : a new approach*. Report GETA & ATR, Osaka, May 1988, and Proc. COLING-88, Budapest, August 1988.