

## Vers une méthodologie générique de contrôle basée sur la combinaison de sources de jugement

Grégory SMITS<sup>1</sup>, Christine CHARDENON<sup>2</sup>

<sup>1</sup> GREYC–Université de Caen, F-14032 CAEN cedex

<sup>2</sup> France Télécom R&D TECH/EASY/LN,

2, avenue Pierre Marzin, 22307 Lannion Cedex

gsmits@info.unicaen.fr

christine.chardenon@orange-ftgroup.com

**Résumé.** Le contrôle des hypothèses concurrentes générées par les différents modules qui peuvent intervenir dans des processus de TALN reste un enjeu important malgré de nombreuses avancées en terme de robustesse. Nous présentons dans cet article une méthodologie générique de contrôle exploitant des techniques issues de l’aide multicritère à la décision. À partir de l’ensemble des critères de comparaison disponibles et la formalisation des préférences d’un expert, l’approche proposée évalue la pertinence relative des différents objets linguistiques générés et conduit à la mise en place d’une action de contrôle appropriée telle que le filtrage, le classement, le tri ou la propagation.

**Abstract.** The control of concurrent hypotheses generated by the different modules which compose NLP processes is still an important issue despite advances concerning robustness. In this article, we present a generic methodology of control inspired from multicriteria decision aid methods. Based on available comparison criteria and formalized expert knowledge, the proposed approach evaluate the relevancy of each generated linguistic object and lead to the decision of an appropriate control action such as filtering, ordering, sorting or propagating.

**Mots-clés :** méthodologie de contrôle, aide multicritère à la décision, apprentissage automatique de métriques.

**Keywords:** control methodology, multicriteria decision aid, metrics automatic learning.

## 1 Introduction

De nombreuses avancées en termes de formalisme, d’algorithmique et de développement de ressources ont permis aux systèmes de traitement automatique des langues naturelles (TALN) d’atteindre une couverture très satisfaisante des différents phénomènes linguistiques observables. Cependant, quelle que soit la tâche ou le niveau d’analyse concerné, les différentes approches envisagées sont de manière récurrente confrontées au manque de précision des résultats générés. Ce phénomène se matérialise par la présence d’objets linguistiques concurrents de différentes natures. Bien que justifiable en présence d’ambiguïtés locales “naturelles”, ces indéterminations concernent la plupart du temps des erreurs d’interprétations souvent qualifiées d’ambiguïtés “ar-

tificielles”. Le contrôle du processus d’analyse a pour objectif d’identifier le plus tôt possible ces ambiguïtés “artificielles” afin notamment d’éviter leur propagation vers les étapes suivantes de l’analyse.

Bien que des propositions d’architectures de TALN conformes vis à vis des modèles cognitifs aient été proposées afin d’éviter la génération d’objets linguistiques erronés (Rady, 1983) (Sabah, 1990), la mise en place de stratégies spécifiques de contrôle semble indispensable.

Nous verrons dans un premier temps que le contrôle des indéterminations, appelés “points d’embarras”<sup>1</sup> par (Sabah, 1989) repose sur la prise en compte d’informations distinctives hétérogènes. En s’appuyant sur un exemple de chaîne de TALN et un cas concret d’indétermination, nous verrons dans un second temps que le contrôle peut être considéré comme une démarche décisionnelle basée sur plusieurs critères de comparaison. Nous avons ainsi développé une méthode complète et générique inspirée des approches d’aide multicritère à la décision. Avant de présenter les premiers résultats obtenus, nous verrons que la prise d’une décision repose sur la formalisation des connaissances d’un expert sur la tâche à contrôler.

## 2 Définition de la notion de contrôle

### 2.1 Le TALN comme une succession de “points d’embarras” potentiels

De nombreux systèmes de TALN peuvent être représentés comme un ensemble de modules d’analyse qui appliqués successivement composent le processus complet d’interprétation linguistique. L’objectif de chacun de ces composants est de construire ses propres interprétations à partir de ressources linguistiques et des informations générées par les étapes précédentes. Ces interprétations que nous nommerons par la suite objets linguistiques peuvent correspondre à :

- des unités lexicales ou de sens ;
- à un découpage en constituants suite à une analyse syntaxique de surface ;
- à un arbre de dépendance syntaxique suite à l’application d’une grammaire de dépendances ;
- à des graphes sémantiques ;
- à des suggestions de corrections orthographiques ;
- à des rubriques sémantiques d’indexation correspondant à une requête.

Fréquemment appliqués séquentiellement, ces modules sont sources de “points d’embarras” et affectent donc la complexité et l’efficacité du processus d’analyse et nuisent également à l’utilisabilité des résultats générés. Il devient alors indispensable de contrôler l’apparition de ces indéterminations, notamment en comparant la pertinence relative de chaque objet linguistique instancié et en mettant en place une action de contrôle appropriée pour éviter la propagation d’interprétations erronées.

### 2.2 Des “points d’embarras” aux “points de décision”

En conservant la terminologie proposée par (Sabah, 1989), l’enjeu du contrôle de ces processus réside dans la transformation des “points d’embarras” en “points de décision”. Ce dernier désigne un état du processus de traitement où plusieurs objets linguistiques concurrents ont été générés, mais qui est également caractérisé par la disponibilité de connaissances et de critères de

<sup>1</sup>L’IA et le langage tome 2 page 121

comparaisons à partir desquels une stratégie de contrôle peut être déployée. Ainsi, pour pallier le manque d'informations distinctives nécessaires à la comparaison des différents objets linguistiques, de nombreux travaux ont proposé d'intégrer des connaissances supplémentaires de différentes natures (probabiliste (Blache & Rauzy, 2006), statistiques (Charniak & M.Johnson, 2005), heuristique (Uszkoreit, 1991), symbolique (Bourigault & Frérot, 2004)) pour qualifier la pertinence de chaque candidat et permettre la mise en place d'une action de contrôle.

On constate également que pour un "point d'embarras" identifié, plusieurs sources indépendantes de connaissances peuvent être exploitées pour affecter à chacun des objets linguistiques un critère de comparaison. Il apparaît cependant que pour chaque contexte de contrôle, aucune connaissance ne permet individuellement de caractériser et d'évaluer pleinement la validité de chaque interprétation envisagée.

L'indéterminisme lié à l'attachement d'un groupe prépositionnel constitue un exemple très illustratif. En effet, l'application d'une grammaire donnée sur une phrase "jouet" comme : "Je possède la statue de bois de rose de Charles", entraînerait sans doute la construction d'au moins 5 arbres syntaxiques concurrents. Afin de déterminer leur pertinence, chaque attachement peut être jugé par rapport :

- à sa conformité vis à vis des informations prosodiques (rarement disponibles) ;
- à son respect des heuristiques d'attachement droit ou minimal ;
- à des données de sous-catégorisations syntaxiques (Bourigault & Frérot, 2004) ;
- à sa fréquence d'apparition observée sur corpus (Gala, 2003) ;
- à des préférences d'usages liées à la sémantique des propositions (Whittemore *et al.*, 1990) ;
- etc.

Chacune des sources de jugement citées précédemment apporte ainsi une information discriminante permettant d'identifier pour certains cas "ambiguïtés" d'attachement envisageables les erreurs d'interprétation à filtrer, ou réciproquement les attachements à privilégier, mais également une erreur d'appréciation pour les autres cas. Afin d'augmenter la robustesse et la crédibilité apportée à l'évaluation de la pertinence de chaque objet linguistique, il semble indispensable de combiner et d'exploiter différents points de vue de jugement. Bien que des arguments psycholinguistiques (Altmann, 1998) (Gibson & Pearlmutter, 1998) aient déjà été avancés en faveur de l'usage combinée de sources de connaissances, peu de stratégies de contrôle exploitent la complémentarité de différents critères de comparaison (Rosso *et al.*, 2003) (Rigau *et al.*, 1997). Nos travaux se sont donc focalisés sur la mise en place d'un formalisme et d'une méthodologie générique de contrôle exploitant l'information apportée par chaque source de jugement pour comparer la pertinence d'objets linguistiques concurrents.

## 2.3 Le cas de TiLT

TiLT est une boîte à outils de TALN développée par l'équipe Langues Naturelles de France Télécom R&D. Le processus d'analyse paramétrable est composé d'un ensemble de modules de traitements qui, appliqués séquentiellement, construisent de manière itérative des interprétations linguistiques de différentes natures. En fonction des caractéristiques du contexte applicatif, certaines étapes de traitement constituent des "points d'embarras". Pour éviter la propagation d'objets linguistiques erronés, différents traits, scores ou probabilités calculés, méthodes spécifiques de jugement ont été intégrés au processus classique d'analyse pour être exploités comme critères de comparaison. Cependant aucune stratégie spécifique de contrôle ne permettait de les combiner ou d'évaluer leur efficacité. L'architecture modulaire d'analyse de

TiLT a été dans un premier temps modifiée (Smits, 2006) afin de simplifier l'intégration de ces critères de comparaison, de les centraliser et de permettre la mise en place de phases de contrôle entre les étapes d'analyse.

### 3 Formalisation d'une méthode d'aide multicritère à la décision

#### 3.1 Agrégation des critères et interprétation des comparaisons

L'architecture décisionnelle de TiLT centralise en tant que critères de comparaison les connaissances supplémentaires intégrées lors du contrôle d'un "point d'embarras". En plus du cadre formel, l'approche de contrôle propose une méthode complète permettant à un expert, le linguiste ou l'informaticien en charge du paramétrage de TiLT, d'exprimer ses connaissances et intuitions sur la tâche de contrôle en question et la façon dont les critères disponibles doivent être exploités. Nous nous rapprochons ainsi du domaine de l'aide multicritère à la décision qui "vise, comme son nom l'indique, à fournir à un décideur des outils lui permettant de progresser dans la résolution d'un problème de décision où plusieurs points de vue, souvent contradictoires, doivent être pris en compte." (Vincke, 1998).

La méthode envisagée (Smits, 2007) s'inspire profondément des méthodes de surclassement (en particulier ELECTRE III et ELECTRE TRI (Roy, 1990)). À partir des valeurs des critères de comparaison qui qualifient les objets comparés et des connaissances exprimées par le décideur, ces méthodes établissent entre les différentes hypothèses candidates des relations de surclassement. Une telle relation notée  $oSo$  est établie entre deux objets linguistiques  $o$  et  $o'$  si "il y a suffisamment d'arguments pour admettre que  $o$  est au moins aussi bonne que  $o'$ , sans qu'il y ait de raison importante de refuser cette affirmation." Ces relations "abstraites" de surclassements permettent d'établir des relations de préférence, d'indifférence (utile pour la factorisation) ou d'incomparabilité (exploitées pour le filtrage).

Pour répondre à une problématique de classement, on exploite l'ensemble des relations de surclassement établies entre les différents objets linguistiques pour établir un pré-ordre partiel (avec *ex aequo*) ou total (Fig. 1). Quant aux problématiques de tri et filtrage, les différents objets linguistiques concurrents ne sont plus comparés entre eux, mais par rapport à des profils d'acceptabilité qui définissent, pour chaque classe considérée, les performances à atteindre sur chacun des critères pour faire partie de la classe en question (cf. cas d'expérimentation Fig. 2).

#### 3.2 Formalisation du problème et construction des relations

Soit  $O : o_1, o_2, \dots, o_n$  l'ensemble des objets linguistiques comparés et  $G : g_1, g_2, \dots, g_m$  les  $m$  critères utilisés lors du contrôle, où  $g_j(o_i)$  correspond à la valeur obtenue par l'objet  $o_i$  pour le  $j^{\text{ème}}$  critère.

Les préférences et connaissances du décideur se matérialisent dans un premier temps à travers les critères choisis pour la tâche de contrôle, mais également par un ensemble de paramètres qui peuvent être associés à chaque critère  $g_j$  :

- un poids d'importance  $w_j$  ;

- un seuil de préférence  $p_j$  ;  
il correspond à la plus petite différence de valeur à partir de laquelle une situation de préférence peut être établie entre deux objets.
- un seuil d'indifférence  $q_j$  ( $q_j \leq p_j$ ) ;  
correspond à la plus grande différence préservant l'indifférence entre 2 objets sur le critère  $j$ .
- un seuil de veto  $v_j$  ( $p_j \leq v_j$ ).  
correspond à la différence de valeur à partir de laquelle un objet devient incomparable vis à vis d'un autre, car jugé trop faible sur un critère important. Ce paramètre permet notamment de définir des conditions de filtrage.

Ces préférences expertes interviennent dans le calcul d'un indice de surclassement  $S(o, ot)$ , quantifiant la crédibilité du surclassement de  $ot$  par l'objet  $o$ , où  $ot$  peut correspondre à un objet concurrent de  $o$  ou à un profil d'acceptabilité d'une classe pour les problématiques de tri. Cet indice de surclassement  $S(o, ot)$  repose sur le produit d'un indice de concordance  $C(o, ot)$ , représentant la majorité des critères en faveur de  $o$ , et d'un indice de discordance  $D(o, ot)$ , représentant la minorité des critères refusant le surclassement de  $ot$  par  $o$  :  $S(o, ot) = C(o, ot) \cdot D(o, ot)$ , où  $C(o, ot) = \frac{1}{\sum_{j \in G} w_j} \sum_{j \in G} w_j c_j(o, ot)$

$$c_j(o, ot) = \begin{cases} 1, & \text{si } g_j(o) - g_j(ot) \geq p_j \\ 0, & \text{si } g_j(o) - g_j(ot) \leq q_j \\ \frac{p_j - g_j(o) - g_j(ot)}{p_j - q_j}, & \text{si } q_j < g_j(o) - g_j(ot) < p_j \end{cases}$$

, et  $D(o, ot) = \prod_{j \in \bar{G}} \frac{1 - d_j(o, ot)}{1 - C(o, ot)}$ ,  $\bar{G} = j \in G / d_j(o, ot) > C(o, ot)$

$$d_j(o, ot) = \begin{cases} 1, & \text{si } g_j(ot) - g_j(o) \geq v_j \\ 0, & \text{si } g_j(ot) - g_j(o) \leq p_j \\ \frac{g_j(ot) - g_j(o) - p_j}{v_j - p_j}, & \text{si } p_j < g_j(ot) - g_j(o) < v_j \end{cases}$$

L'interprétation des relations de surclassement permet de définir des situations de préférence stricte :  $oPot$  si  $S(o, ot) \wedge \neg S(ot, o)$ , d'indifférence :  $oIot$  si  $S(o, ot) \wedge S(ot, o)$  ou d'incomparabilité :  $oRot$  si  $\neg S(o, ot) \wedge \neg S(ot, o)$ . Les relations établies, regroupées dans une structure de préférences (Fig. 1), sont exploitées pour établir un pré-ordre complet ou partiel (avec *ex aequo*) des objets comparés.

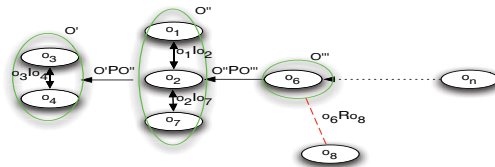


FIG. 1 – Représentation agrégée des relations de préférences établies entre les hypothèses

### 3.3 Vers une élicitation des préférences et connaissances de l'expert

La pertinence des relations de comparaison établies entre les objets concurrents dépend à la fois de l'information distinctive portée par les critères qui les qualifient, mais également des préférences du décideur. Il est cependant difficile et peu naturel pour un décideur de déterminer clairement et par l'intermédiaire de valeurs numériques, les valeurs de ces différents paramètres décisionnels. Même si l'externalisation de l'ensemble de ces informations dans un module spécifique de contrôle facilite les expérimentations itératives, l'impact de chaque paramètre sur le résultat final est difficilement quantifiable *a priori*. Il est en revanche nettement plus évident pour un expert de se prononcer sur la pertinence des objets linguistiques générés. Cette démarche de validation des résultats générés par un expert est fortement exploitée pour la construction de corpus de références, servant ensuite de données pour l'apprentissage de ressources ou l'évaluation de systèmes.

Nous proposons de considérer l'identification des objets linguistiques de référence comme l'expression de l'expertise de l'annotateur sur une tâche de contrôle à automatiser. Ainsi, qualifier une hypothèse comme valide revient à qualifier les performances qu'elle a obtenu sur les critères exploitées comme discriminants. Une hypothèse annotée comme correcte ou incorrect et les performances calculées sur les critères concernés par l'étape de contrôle constituent un enregistrement de ce que nous nommons un tableau de performances : Nous exploitons ce tableau de

Objets annotés.	vecteur de performances				annotation
	critère 1	critère 2	...	critère $m$	
$o_1$	4.2	Vrai	...	36	correct
$o_2$	5.0	Vrai	...	16	correct
$o_3$	2.6	Faux	...	24	incorrect
$o_4$	1.2	Faux	...	42	correct
...	...	...	...	...	...
$o_{p-1}$	4.0	Vrai	...	4	incorrect
$o_p$	0	Faux	...	17	incorrect

TAB. 1 – Tableau de performances construit à partir du corpus de références

performances pour évaluer et quantifier la pertinence de chacun des critères disponibles. Une distribution de performances d'un critère est jugée pertinente, si elle permet de caractériser un certain type d'annotation (i.e. la classe des hypothèses correctes ou la classes des hypothèses incorrectes). La méthode d'apprentissage de métriques RELIEF (Kononenko, 1994) permet d'atteindre ce but, en associant à chaque critère un poids normé sur  $[-1, 1]$ , où une valeur négative caractérise un critère non représentatif de la classe des hypothèses correctes. Les résultats de la méthode sont ensuite exploités lors de la construction de l'indice de crédibilité en tant que vecteur de poids des critères.

Nous proposons également de considérer la performance minimale obtenue sur un critère par les hypothèses annotées comme correctes en tant que limite d'acceptabilité de ce critère.

## 4 Expérimentation

Nous présentons dans cette section les premiers résultats obtenus sur un des nombreux cas de contrôle envisageables. Il s’agit de répondre à une problématique de classement des couples antécédent/reprise-anaphorique candidats extraits d’un corpus.

### 4.1 Contrôle des couples antécédent/reprise-anaphorique candidats

Cette expérimentation s’inscrit dans le cadre d’une collaboration et d’une extension des travaux réalisés par OLIVIER TARDIF (Tardif, 2006). Un algorithme extrait à partir d’un texte un ensemble de couples d’expressions constituant potentiellement des patrons antécédent/reprise-anaphorique. Une expression correspond à une entité nommée (NPR : Mickaël Gordbatchev, l’URSS, Vilnius), un pronom (PRON : il, celui-là) ou un groupe nominal (NCOM : le dirigeant soviétique, le parlement). L’enjeu du contrôle est de construire une classes des candidats valides à partir des performances qui leurs sont associées sur différents critères, tels que :

- des mesures de distances (en mots, phrases, etc.) ;
- la correspondance des classes sémantiques et des fonctions syntaxiques ;
- des marques morphologiques (indéfini, possessif) ;
- des mesures de distance et de similarité alphabétiques ;
- des propriétés d’accords de genre et de nombre.

Pour constituer la classe des reprises anaphoriques valides, les relations de surclassement ne sont pas construites entre les couples candidats, mais entre chaque candidat et un profil d’acceptabilité (Fig. 2). Ce vecteur de limites d’acceptabilité constitue une nouvelle préférence mise en place par l’expert pour contrôler la validité des hypothèses comparées. Ainsi, un candidat qui surclasse ce profil est considéré comme un cas de reprise anaphorique.

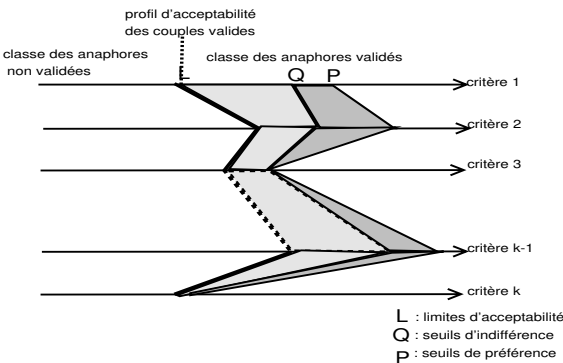


FIG. 2 – Profil d’acceptabilité des reprises anaphoriques candidates

## 4.2 Entre expertise et apprentissage automatique

Nous disposons d'un corpus de 80 textes journalistiques (Le Monde de 1989-1990) annoté automatiquement par TILT afin de disposer d'informations morphologiques, syntaxiques et sémantiques sur les expressions et leur rôle dans la phrase. Les liens de coréférences entre les expressions ainsi que la classe sémantique (personne, lieux, organisation) de celles-ci ont ensuite été marqués manuellement.

Les couples d'expressions constituées à partir du corpus ont été partitionnés en fonction de leur type : NPR-NPR, NPR-NCOM, NPR-PRON, NCOM-NCOM, NCOM-PRON, PRON-PRON. Nous nous sommes restreint pour cette évaluation aux couples ayant pour antécédent un nom propre.

Dans un premier temps, nous avons, à travers une démarche interactive, demandé à un expert du domaine de constituer trois profils de paramètres décisionnels (voir Sec. 3.2) pour les trois cas de reprise anaphorique traités. L'expert devait ainsi identifier les critères qu'il jugeait pertinents dans chacun des cas, ainsi que leur importance relative dans l'agrégation, un profil d'acceptabilité et éventuellement des seuils de préférence, indifférence et veto. Dans un second temps, nous avons exploité une partie du corpus annoté pour apprendre automatiquement les poids des différents critères ainsi que les seuils délimitant la classe des couples valides (Sec. 3.3). Ce corpus est constitué de 950 paires npr-npr, 3400 paires npr-ncom et 620 paires npr-pron, composé respectivement de 90, 46 et 48 cas valides (positifs) de coréférence. Nous avons ensuite évalué les différents profils de paramètres sur un corpus d'évaluation extrait du corpus de référence, constitué de 120 paires candidates NPR-NPR (19 positives), 80 paires NPR-PRON (12 positives) et de 414 paires NPR-NCOM (9 positives).

Profil	Expressions	Précision	Rappel	F-mesure
Manuel	NPR-NPR	0.9	0.92	0.91
	NPR-NCOM	0.4	0.2	0.26
	NPR-PRON	0.4	0.25	0.3
Automatique	NPR-NPR	0.94	0.96	0.95
	NPR-NCOM	0.35	0.17	0.23
	NPR-PRON	0.6	0.39	0.47

L'apprentissage automatique des poids des critères ainsi que des limites d'acceptabilité nous permet d'améliorer sensiblement les résultats bien que le corpus soit principalement composé d'exemples négatifs (à plus de 90% sur le corpus de test et à plus de 95% sur le corpus d'apprentissage). Inférer automatiquement ces paramètres décisionnels nous permet d'identifier et de quantifier l'utilité des différents critères disponibles, contredisant parfois les intuitions de l'expert qui exploitait des critères n'apportant que du bruit. Par exemple, pour le cas des couples NPR-PRON, l'expert a sélectionné quatre critères comme pertinents et a formulé les préférences suivantes concernant l'importance relative de chacun d'eux :

1. accord en nombre entre l'antécédent et la reprise
2. accord en genre entre l'antécédent et la reprise
3. nombre d'occurrences de l'antécédent dans le texte
4. nombre d'expressions séparant l'antécédent de la reprise
5. antécédent et reprise ont la fonction sujet



Cependant, par apprentissage sur corpus de référence, de nouveaux critères ont été identifiés comme pertinents et l'ordre d'importance de l'ensemble des critères utilisé a été modifié, ce qui explique l'amélioration des résultats (les autres paramètres de seuils restants identiques au profil décisionnel de l'expert) :

1. antécédent est l'expression la plus proche
2. antécédent et reprise sont dans la même phrase
3. nombre d'occurrences de l'antécédent dans le texte
4. accord en nombre entre l'antécédent et la reprise
5. nombre d'expressions séparant l'antécédent de la reprise
6. nombre de mots séparant l'antécédent de la reprise
7. antécédent et reprise ont la fonction sujet
8. accord en genre entre l'antécédent et la reprise

Cette tâche d'identification des couples antécédent/reprise-anaphorique avait dans un premier été traitée à l'aide de classifieurs bayésiens naïfs. Outre des améliorations des valeurs de précision et de rappel, notre approche offre à l'expert la possibilité d'intervenir sur le comportement de la méthode de classification mais également une meilleure compréhension des décisions émises.

## 5 Perspectives et conclusion

Nous proposons une méthode générique de contrôle des points d'embarras apparaissant lors d'un processus de TALN. Cette méthode inspirée de l'aide multicritère à la décision se base sur l'agrégation de critères de comparaison hétérogènes. Les différents paramètres externalisés dans un profil décisionnel permettent à un expert d'exprimer ses connaissances et intuitions sur le problème traité. Pour valider ou inférer automatiquement les préférences émises par un expert, nous utilisons des méthodes d'apprentissage supervisé exploitant un corpus de référence.

La méthode d'apprentissage de métriques RELIEF s'avère efficace pour quantifier la représentativité d'un critère vis à vis d'un ensemble d'exemples annotés comme valides. Nous envisageons cependant d'exploiter une variante de cette méthode pour réduire l'impact de la forte proportion d'exemples négatifs lors de l'apprentissage. Nous travaillons actuellement sur la mise en place de méthodes de seuillage pour inférer automatiquement les autres paramètres décisionnels.

L'approche proposée est en cours de validation sur un autre cas concret d'expérimentation : le classement des arbres syntaxiques concurrents. L'automatisation d'une procédure d'aide à la décision basée sur la comparaison deux à deux des hypothèses concurrentes, pour les problématiques de classement, pose cependant des problèmes de complexité. Ainsi, en présence d'un grand nombre d'hypothèses concurrentes, il ne semble pas judicieux de construire un classement de tous les candidats. Nous proposons donc d'effectuer un premier filtrage en exploitant notamment les critères jugés comme les plus pertinents par la méthode d'apprentissage automatique des poids. Sur le sous-ensemble d'hypothèses restant, des relations de surclassement peuvent être établies et interprétées pour obtenir un classement des  $N$  meilleurs candidats.

## Références

- ALTMANN G. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, **2**(4).
- BLACHE P. & RAUZY S. (2006). Mécanismes de contrôle pour l'analyse en grammaires de propriétés. In *in proceedings of TALN*.
- BOURIGAULT D. & FRÉROT C. (2004). Ambiguïté de rattachement prépositionnel : introduction de ressources exogènes de sous-catégorisation dans un analyseur syntaxique de corpus endogène. In *in proceedings of TALN*.
- CHARNIAK E. & M. JOHNSON (2005). Coarse-to-fine n-best parsing and maxent discriminative reranking. In *ACL '05 : Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, p. 173–180, Morristown, NJ, USA : Association for Computational Linguistics.
- GALA N. (2003). Une méthode non supervisée d'apprentissage sur le web pour la résolution d'ambiguïtés structurelles liées au rattachement prépositionnel. In *in proceedings of TALN*.
- GIBSON E. & PEARLMUTTER N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, **2**(7).
- KONONENKO I. (1994). Estimating attributes : Analysis and extensions of relief. In *In proceedings of the European Conference on Machine Learning*.
- RADY M. (1983). *L'ambiguïté du langage naturel est-elle la source du non-déterminisme des procédures de traitement ?* PhD thesis, Université de Paris VI.
- RIGAU G., ATSERIAS J. & AGIRRE E. (1997). Combining unsupervised lexical knowledge methods for word sense disambiguation. *Proceedings of the 35th annual meeting on Association for Computational Linguistic*.
- ROSSO P., MASSULLI F. & BUSCALDI D. (2003). Word sense disambiguation combining conceptual distance, frequency and gloss. *IEEE*.
- ROY B. (1990). Decision-aid and decision-making. In *European Journal of Operational Research*, volume 45, p. 324–331.
- SABAH G. (1989). *L'IA et le langage (tome 2)*. Hermes.
- SABAH G. (1990). Caramel : A flexible model for interaction between the cognitive processes underlying natural language understanding. In *Proceedings of the Ninth European Conference on Artificial*.
- SMITS G. (2006). Contrôle dynamique multicritère des résultats d'une chaîne de tal. In *in proceedings of RECITAL*.
- SMITS G. (2007). Méthodologie d'aide multicritère à la décision pour le contrôle d'une chaîne de traitement automatique des langues naturelles. In *in proceedings of ROADEF'07*.
- TARDIF O. (2006). Résoudre la coréférence à l'aide d'un classifieur bayésien naïf. In *in proceedings of RECITAL*.
- USZKOREIT H. (1991). Strategies for adding control information to declarative grammars. In A. FOR COMPUTATIONAL LINGUISTICS, Ed., *Proceedings of the 29th annual meeting on Association for Computational Linguistics*, p. 237–245.
- VINCKE P. (1998). *Aide multicritère à la décision*. Ellipses Marketing.
- WHITTEMORE G., FERRARA K. & BRUNNER H. (1990). Empirical study of predictive powers of simple attachment schemes for post-modifier prepositional phrases. In *Proceedings of the 28th annual meeting on Association for Computational Linguistics*.