

## **Peut-on trouver la taille de contexte optimale en désambiguïsation sémantique?**

Éric Crestan (1,2), Marc El-Bèze(1) et Claude de Loupy (2)

(1) Laboratoire Informatique d'Avignon  
339, ch des Meinajaries, BP 1228  
F-84911 Avignon Cedex 9  
{eric.crestan, marc.elbeze}@lia.univ-avignon.fr

(2) Sinequa S.A.S.  
51-59 rue Ledru-Rollin  
F-94200 Ivry-sur-Seine  
{crestan, loupy}@sinequa.com

### **Résumé – Abstract**

Dans la tâche de désambiguïsation sémantique, la détermination de la taille optimale de fenêtre de contexte à utiliser, a fait l'objet de plusieurs études. Dans cet article, nous proposons une approche à deux niveaux pour répondre à cette problématique de manière automatique. Trois systèmes concurrents à base d'arbres de classification sémantique sont, dans un premier temps, utilisés pour déterminer les trois sens les plus vraisemblables d'un mot. Ensuite, un système décisionnel tranche entre ces sens au regard d'un contexte plus étendu. Les améliorations constatées lors d'expériences menées sur les données de SENSEVAL-1 et vérifiées sur les données SENSEVAL-2 sont significatives.

The determination of context length to use for Word Sense Disambiguation (WSD) has been the object of several studies. In this paper, we propose to use a monitoring system in order to select automatically the optimal window size among three possibilities. We used a two-step strategy based on Semantic Classification Trees (SCT) and on a similarity measure. Whereas SCTs are employed on a short window size of 3, 5 and 7 words, the technique based on similarity measure is applied to a 'wider' context size. The improvements observed in the SENSEVAL-1 lexical-sample task are verified on the SENSEVAL-2 data.

### **Mots clés – Keywords**

Désambiguïsation sémantique, arbres de classification sémantique.

Word sense disambiguation, semantic classification trees, monitoring system.

## 1 Introduction

Dans les années 50, Kaplan (1950) a observé, à l'occasion d'une expérimentation, que la traduction d'un mot par sept traducteurs différents n'était ni meilleure ni pire lorsque ceux-ci n'avaient à leur disposition que deux mots de contexte de chaque côté du mot à traduire, plutôt que la phrase au complet. Plus récemment, Yarowsky (1993) a déclaré que la plupart des indices utiles à la désambiguïsation sémantique se trouvent dans un micro-contexte de 6 à 8 mots. Toutefois, il faut noter que dans un contexte de « si grande taille », il est souvent difficile de discerner les éléments clés, par rapport aux éléments non porteurs d'information pour la détermination du sens d'un mot.

Au-delà de cette réduction de la fenêtre de contexte nécessaire à la désambiguïsation, il paraît évident qu'une taille fixe n'est pas adaptée à tous les mots. Pour s'affranchir de ce problème, il est possible dans le cadre de systèmes supervisés, de déterminer la taille de fenêtre optimale au regard du corpus d'apprentissage. Nous appellerons cette méthode "adaptation statique". Le désavantage d'une telle approche est qu'elle est très sensible à la qualité du corpus d'apprentissage. Une autre solution, que nous nommerons « adaptation dynamique », est de déterminer une taille optimale de contexte appropriée pour chaque test.

L'approche proposée dans ce papier est basée sur une approche mixte à deux niveaux. Dans un premier temps, la phrase contenant le mot à désambiguïser est soumise à trois systèmes identiques entraînés sur différentes tailles de contexte. Ensuite, lors de la seconde phase, le sens final est sélectionné parmi les sens proposés sur des critères plus thématiques. Les expériences, proposées dans cet article, ont été menées sur les corpus issus des campagnes d'évaluation SENSEVAL-1 et 2 (Kilgarriff et Rosenzweig, 2000). Le but de ces évaluations étant de désambiguïser un mot en contexte, pour lequel un corpus d'apprentissage était fourni.

Le papier est organisé de la manière suivante : une brève présentation du système pour la désambiguïsation sémantique à base d'arbres de classification sémantique (SCT) est tout d'abord donnée (section 2). Dans la section 3, la technique du *leave-one-out*, méthode très répandue notamment dans la modélisation statistique du langage (Ney, Martin et Wessel, 1997), est employée pour opérer une adaptation statique de la taille de fenêtre de contexte. Enfin, dans la section 4, nous démontrons que l'utilisation d'un système de détection de longueur optimale de contexte, couplé avec les SCT, améliore les résultats au-delà de ce qui aurait pu être obtenu par une adaptation statique dans le meilleur des cas.

## 2 Arbres de classification sémantique pour la désambiguïsation du sens

Kuhn et De Mori (1995) ont été les premiers à introduire l'idée des arbres de Classification Sémantique (*Semantic Classification Trees*, SCT) dans le domaine de la compréhension du langage naturel. Plus récemment, les mêmes techniques ont été utilisées par Loupy et al. (2000) en désambiguïsation sémantique dans le cadre de l'évaluation SENSEVAL-1 avec quelque succès. Dans la suite de cette section, le principe de l'approche est brièvement rappelé, et nous montrons comment elle peut

## Peut-on Trouver la Taille de Contexte Optimale en Désambiguïsation Sémantique?

être étendue pour améliorer les performances au-delà de ce qui a été obtenu jusqu'à présent.

Les SCT sont des arbres de décision binaires entraînés sur un corpus d'apprentissage annoté. Le corpus d'entraînement de la tâche *lexical sample* de l'évaluation SENSEVAL-2 a donc été utilisé pour construire un arbre de classification pour chaque mot à désambigüiser. Le processus de construction des arbres consiste à trouver à chaque étape de la construction, la question optimale qui sépare au mieux l'espace constitué par les différentes populations d'exemples. Les questions posées à chaque nœud de l'arbre sont de la forme :

*'Est-ce que l'élément (lemme, stemme ou graphie) à la position P est égale à X ?'*

Dans le cadre de cette expérience, le critère d'impureté de Gini (Breiman *et al.* 1984) a été employé comme critère de décision.

sense%1:10:00:: firebreak , <i>in sense of</i> material burning sense%1:10:00:: someone , <i>in sense of</i> appear , sense%1:10:00:: which people <i>make sense of</i> situation , sense%1:10:00:: way of <i>make sense of</i> be to sense%1:10:00:: make sort <i>of sense of</i> relatedness .	} "the meaning of a word or expression"
sense%1:09:05:: be inhibit <i>by sense that</i> government and sense%1:09:05:: Jess feel <i>such sense of</i> disappointment that sense%1:09:05:: of credo <i>underpin sense of</i> faith in sense%1:09:05:: ceiling add <i>to sense of</i> space and sense%1:09:05:: inspire reader <i>with sense of</i> presence of	} "a general conscious awareness"
sense%1:09:04:: against law <i>make sense</i> . sense%1:09:04:: PRP make <i>more sense to</i> annex quality sense%1:09:04:: PRP <i>have sense than</i> PRP have sense%1:09:04:: , and <i>make sense in</i> term of	} "sound practical judgment"
sense%1:09:02:: to ensure <i>that senses can</i> operate . sense%1:09:02:: take in <i>via senses and</i> process . sense%1:09:02:: whether PRP <i>have senses be</i> anybody 's	} "the faculty through which the external world is apprehended"

Figure 1: Échantillon du corpus d'apprentissage pour le nom 'sense'

Avant de pouvoir construire les SCT, les données ont subi un pré-traitement. Tout d'abord, les contextes ont été lemmatisés pour augmenter le pouvoir de généralisation lors de la sélection des questions. Le mot de référence (à désambigüiser) est conservé à l'état de graphie pour utiliser le maximum d'indices susceptibles d'apporter une information sur le sens de celui-ci (typiquement : mot capitalisé : "Sense", flexion : "senses"). Par la suite, les adjectifs, adverbes et déterminants/articles ont été retirés conformément aux observations faites par Loupy et El-Bèze (2000). Les pronoms, quant à eux, sont remplacés par la désignation générique *PRP*.

Un échantillon du corpus d'apprentissage utilisé pour entraîner l'arbre de décision pour le nom anglais *sense* est donné en Figure 1. Pour les quatre sens différents du mot, les exemples sont présentés avec 3 mots de contexte de chaque côté du mot à désambigüiser après pré-traitement, donc une fenêtre de 7 mots ( $K=7$ ). On observera que les ponctuations fortes provoquent, s'il y a lieu, un raccourcissement de la fenêtre. L'arbre présenté en Figure 2 a, quant à lui, été construit sur le même corpus d'apprentissage, mais sur un contexte plus court qui correspond à la partie grisée sur

l'échantillon présenté plus haut ( $K=3$ ). Considérant que le terme central est à la position 0, les questions ne peuvent porter que sur les lemmes en position  $-1$  et  $1$ , ainsi que la graphie du mot sur lequel l'arbre est construit (position 0).

Les feuilles de cet arbre contiennent le sens dominant (le référentiel étant *Wordnet 1.7*). Les nœuds de l'arbre contiennent pour leur part, les questions qui ont été jugées les plus pertinentes par le processus d'inférence. Les annotations  $y$  et  $n$  sur les arcs correspondent respectivement à une réponse positive ou négative apportée à la question sur le nœud précédent. Ainsi, la question sélectionnée comme la plus pertinente à la racine de l'arbre concerne la présence (ou absence) de la préposition *in* en position  $-1$  (terme précédant le nom *sense*). Cette question semble pertinente au vu de l'échantillon. Seul les exemples donnés pour le sens `sense%1:10:00::` (sens *WordNet*) ont une préposition *in* qui précède le mot à désambiguïser.

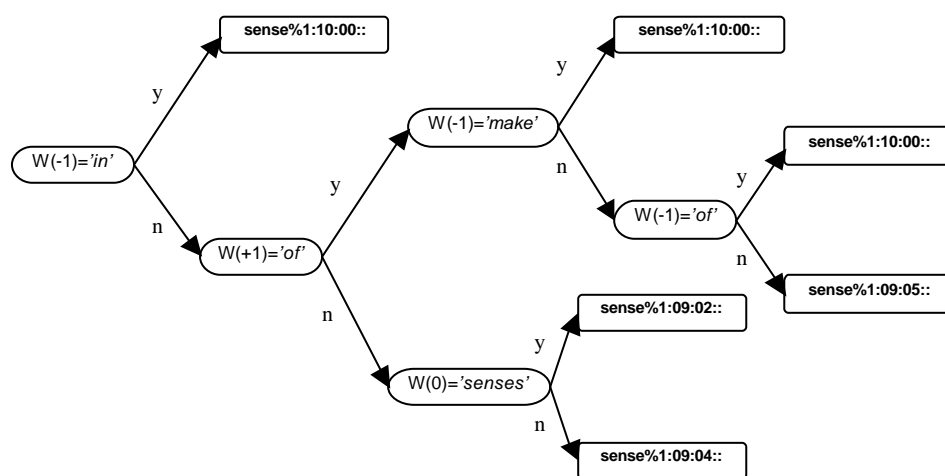


Figure 2: *SCT construit pour le nom 'sense'*

Pour ne pas souffrir du manque de données, nous avons proposé l'utilisation d'éléments d'information extra-contextuels. Pour chaque lemme en contexte, toutes ses Classes Sémantiques (CS), telles qu'elles sont définies dans *Wordnet*, ont été ajoutées comme questions potentielles et cela indépendamment de toute catégorie grammaticale. Cela a pour avantage d'accroître considérablement le champ de couverture des questions car elles concernent dorénavant non seulement les lemmes, mais aussi leurs classes sémantiques (pour plus d'information se référer à Crestan *et al.*, 2001b). Cela ouvre la voie à un nouveau format de question :

'Est-ce que l'élément à la position  $X$  appartient à la Classe Sémantique  $CS$  ?'

### 3 Optimisation de paramètres : *Leave-one-Out*

#### 3.1 Principe

Le réglage des paramètres n'est pas une chose évidente dans le cadre de systèmes supervisés. Pour cela, la technique bien connue du *leave-one-out* (Lachenbruch et Michéy, 1968) a été employée pour optimiser les paramètres entrant en jeu dans la construction des arbres. Le principe de cette technique consiste à effectuer une évaluation circulaire sur le corpus d'apprentissage, en excluant à chaque tour un

exemple du corpus d'apprentissage qui sera utilisé par la suite comme élément de test. Le corpus d'apprentissage de SENSEVAL-2 a été utilisé pour effectuer ces réglages. Les résultats présentés dans la prochaine section

Cette technique rend possible plusieurs types d'optimisations :

- Optimisation de la taille de la fenêtre de contexte : des expériences menées précédemment sur les données SENSEVAL-1 ont montré que la taille de fenêtre optimale varie d'un mot à l'autre (cf. 4.1) ;
- Utilisation des informations additionnelles : bien que l'utilisation des CS apporte une amélioration sensible des performances en précision moyenne, ce surcroît d'information pénalise, malgré tout, certains mots;
- Optimisation de profondeur d'arbre : la profondeur optimale de chaque arbre est également variable selon les arbres.

Dans la section suivante, seul le critère d'optimisation de la longueur de contexte sera étudié.

### 3.2 Application à la détection de taille optimale de contexte

Le but de cette expérience est de vérifier l'hypothèse selon laquelle il serait possible d'appliquer la technique du *leave-one-out* afin de détecter, de manière automatique, la taille de fenêtre optimale (dans le cadre des SCT).

Lors de récents travaux (Crestan et El-Bèze, 2001a) nous avons observé qu'il n'y a pas de taille de fenêtre unique adaptée à tous les noms (corpus SENSEVAL-1), mais que celle-ci varie d'un mot à l'autre. D'après ces mêmes travaux, il a été montré qu'un gain en précision globale de 1% serait possible si l'on pouvait déterminer *a priori* la longueur de fenêtre optimale à utiliser lors de l'apprentissage. Des tests similaires, réalisés sur les données SENSEVAL-2, ont conduit aux mêmes conclusions.

	Résultats sur corpus d'apprentissage (leave-one-out)			Résultats sur corpus de test		
	Nb Test	SCT $k=3$	SCT $k=5$	SCT $k=7$	Avec $\hat{k}$	SCT $k=3$
<i>Moyenne</i>	98	57.0	55.3	54.0	63.6	65.3

Tableau 1: Optimisation de taille de fenêtre par *leave-one-out* sur les NOMS

Partant des observations faites précédemment, la technique du *leave-one-out* a été appliquée au corpus d'apprentissage des 29 NOMS fournis lors de l'évaluation SENSEVAL-2 (*art, authority, bar, bum, chair, channel, child, church, circuit, day, detention, dyke, facility, fatigue, feeling, grip, hearth, holiday, lady, material, mouth, nation, nature, post, restraint, sense, spade, stress* et *yew*). Ces résultats sont présentés dans le Tableau 1. Pour chaque longueur de fenêtre considérée ( $k=3, 5$  et  $7$ ), les arbres ont été construits séquentiellement sur tous les exemples sauf un, puis testés sur celui-ci. Les colonnes SCT  $k=3, 5$  et  $7$  indiquent la précision moyenne obtenue pour chacun des mots (en prenant successivement chaque exemple comme test et tous les autres comme entraînement). La fenêtre qui donne globalement la meilleure précision pour

un mot  $m$  est considérée par la suite comme la longueur de fenêtre optimale ( $\hat{k}$ ) lors du test de  $m$ . De par la finesse des sens de *Wordnet*, plusieurs exemples du corpus d'apprentissage n'ont pas été désambiguïsés par les juges. Pour respecter le caractère ambigu de ces exemples, nous les avons dupliqués selon autant d'exemple qu'il y avait de sens.

### 3.3 Validation de l'approche

Les tests menés sur le corpus d'évaluation de SENSEVAL-2 ne sont malheureusement pas corrélés avec les observations faites sur le corpus d'apprentissage. En effet, les résultats obtenus en utilisant les tailles de fenêtre "optimales" déterminées par la méthode du *leave-one-out* ( $\hat{k}$ ) sont en deçà des résultats obtenus avec une fenêtre fixe  $k=3$ . La précision moyenne est de 1.7% plus mauvaise. Sur les 29 noms de cette expérience, la colonne ( $\hat{k}$ ) présente seulement 2 fois des résultats supérieurs à la colonne (SCT  $k=3$ ), alors que le cas contraire se vérifie 9 fois, le reste des résultats étant identique.

Plusieurs raisons peuvent être avancées pour expliquer cet échec. Tout d'abord, le nombre d'exemples pour chaque sens est très faible. Qui plus est, pour certains sens seuls un ou deux exemples sont présents dans le corpus d'apprentissage, ce qui provoque forcément des erreurs sur ces exemples lors de l'utilisation du *leave-one-out*. Ainsi, 1 seul exemple est fourni pour 3 sens différents du nom *art* et seulement 2 pour 5 autres sens de ce mot. Le processus de construction des arbres ne pouvant pas caractériser ces sens par manque de données, les résultats sont d'autant plus mauvais. La présence de plusieurs sens possibles pour un même exemple est également source de problèmes. Il n'est pas possible de trouver une question pour séparer ces exemples dupliqués car les contextes sont identiques. D'autres expériences devront être menées dans cette direction pour déterminer quelle est la meilleure stratégie à appliquer avec les exemples "ambigus" (par exemple ne conserver qu'un seul des sens). La méthode du *leave-one-out* ne semble donc pas être appropriée pour détecter automatiquement la taille optimale de fenêtre de contexte, dans le cadre de notre approche par SCT.

## 4 Sélection automatique de fenêtre optimale

Différents travaux ont été menés par le passé sur l'optimisation de la taille de fenêtre de contexte. Il convient de citer les travaux de Yarowsky (1992) qui avance que deux types d'ambiguïté existent : l'ambiguïté relevant du domaine, qui nécessite une fenêtre d'étude de 20 à 50 mots et l'ambiguïté locale pour laquelle une fenêtre de 2 à 3 mots semble suffisante. Néanmoins, il ne fournit pas de solution à ce problème de sélection de taille de fenêtre. Les expériences présentées par la suite permettent d'y répondre en partie.

### 4.1 Principe

Dans la section précédente, nous avons montré que la technique du *leave-one-out* n'est pas adaptée à l'optimisation de la taille de fenêtre de contexte. Cette section est centrée sur une approche novatrice, faisant intervenir un système décisionnel, ce qui permet de s'affranchir de la tâche d'optimisation de taille de fenêtre de contexte. L'idée principale consiste à trouver des informations utiles à la désambiguïsation sémantique dans un contexte plus étendu, dans le but de renforcer les observations

effectuées en contexte court. Ce système à l'avantage d'être dynamique, dans le sens où il sélectionne la fenêtre optimale pour chaque phrase de test et non pas une fois pour toute pour un mot donné comme c'était le cas en section 3. Le principe en est le suivant : pour chaque mot à désambiguïser  $W$ , trois SCT sont entraînés sur le corpus d'apprentissage en utilisant pour chacun des contextes de longueur différente (*respectivement*,  $SCT^{k=3}(W)$ ,  $SCT^{k=5}(W)$  et  $SCT^{k=7}(W)$ ). Ensuite, lors de la phase de désambiguïisation, pour chaque test  $t$ , les arbres sont parcourus successivement, ce qui a pour but de générer trois propositions de sens ( $S^{k=3}(t)$ ,  $S^{k=5}(t)$ ,  $S^{k=7}(t)$ ). Le système décisionnel prend alors le relais et détermine, parmi ces sens, lequel semble le plus conforme aux indices récoltés sur un contexte plus large.

Le point crucial de cette approche reste la fonction de décision. L'approche décisionnelle la plus simple consiste à utiliser un système probabiliste basé sur un modèle unisem, mais les expériences que nous avons menées dans cette voie ont conduit à des résultats décevants.

En 1993, Gale *et al.* ont appliqué une approche d'extraction d'information au domaine de la désambiguïisation sémantique. Ils ont utilisé une taille de fenêtre de contexte arbitraire de 50 mots à gauche, ainsi qu'à droite du mot à désambiguïser. Ensuite, ils ont utilisé un système de recherche documentaire probabiliste pour comparer le contexte des mots à désambiguïser, avec le contexte de ces mêmes mots contenus dans le corpus d'apprentissage. Leurs conclusions furent que les résultats sont significativement améliorés lorsque l'on utilise un contexte plus vaste. Le système décisionnel utilisé dans cette expérience est inspiré de ces travaux, il met en œuvre une mesure de distance entre une phrase de test  $t$  et un pseudo-document créé à partir du corpus d'apprentissage. Pour un mot donné  $W$ , un pseudo-document  $D_{S_i(W)}$  est un document construit par la concaténation de toutes les phrases présentes dans le corpus d'entraînement pour un même sens de ce mot  $S_i(W)$ . La mesure de similarité classique du *Cosinus* définie par Salton et McGill (1986), est alors utilisée pour calculer la distance entre les phrases de test et les pseudo-documents.

## 4.2 Validation de l'approche

Pour vérifier la validité de l'approche présentée dans la section précédente, des tests ont été menés sur les données d'évaluation SENSEVAL-1. Le système de détection automatique de taille de contexte a été appliqué aux 12 noms, ainsi qu'aux 13 verbes de l'évaluation. Les résultats sont présentés dans le Tableau 2 :

	SCT $k=3$	SCT $k=5$	SCT $k=7$	MaxStat	SCT + Cos
NOMS	83,3	82,1	82,1	84,3	<b>85,7</b>
VERBES	71,1	68,6	68,0	71,6	<b>72,8</b>

**Tableau 2:** Évaluation sur les données SENSEVAL-1 (en %)

La colonne intitulée *MaxStat* correspond à la précision moyenne qu'aurait pu atteindre un système opérant une sélection de taille de fenêtre de contexte appropriée à chaque mot. Il faut toutefois noter que cela ne garantit pas l'optimalité au niveau de chaque test, mais seulement au niveau du mot. Cela correspond donc à la borne supérieure pour une *adaptation statique*. La dernière colonne (*SCT+Cos*) contient les précisions moyennes obtenues pour les noms et les verbes en utilisant l'*adaptation dynamique*

avec le système basé sur la mesure de similarité du *Cosinus*. Ces résultats sont tout à fait satisfaisants car, en plus d'être significativement supérieurs aux résultats obtenus par les SCT seuls, ils sont également supérieurs aux meilleurs résultats que nous aurions pu obtenir par *adaptation statique*. Au vu de ces résultats, il semble clair qu'il n'existe pas de taille fixe pour un mot donné. En élargissant le contexte, il est possible de choisir entre trois tailles de fenêtre au niveau du test.

NOMS	Nb Test	SCT k=3	SCT k=5	SCT k=7	MaxStat	SCT +Cos
art	98	63,3	<b>64,3</b>	62,2	<b>64,3</b>	63,3
authority	92	<b>72,8</b>	67,4	66,3	<b>72,8</b>	<b>72,8</b>
bar	151	<b>61,6</b>	53,6	56,3	<b>61,6</b>	58,9
bum	45	75,6	73,3	68,9	75,6	<b>80,0</b>
chair	69	79,7	<b>81,2</b>	76,8	<b>81,2</b>	<b>81,2</b>
channel	73	54,8	54,8	53,4	54,8	<b>56,2</b>
child	64	54,7	57,8	60,9	60,9	<b>62,5</b>
church	64	54,7	46,9	42,2	54,7	<b>56,3</b>
circuit	85	<b>57,6</b>	48,2	<b>57,6</b>	<b>57,6</b>	<b>57,6</b>
day	145	64,8	61,4	60,0	64,8	<b>65,5</b>
detention	32	84,4	84,4	84,4	84,4	<b>87,5</b>
dyke	28	<b>78,6</b>	71,4	67,9	<b>78,6</b>	<b>78,6</b>
facility	58	65,5	<b>67,2</b>	58,6	<b>67,2</b>	<b>67,2</b>
fatigue	43	<b>86,0</b>	76,7	76,7	<b>86,0</b>	<b>86,0</b>
feeling	51	<b>66,7</b>	64,7	64,7	<b>66,7</b>	<b>66,7</b>
grip	51	54,9	64,7	<b>66,7</b>	<b>66,7</b>	62,7
hearth	32	75,0	78,1	71,9	78,1	<b>81,3</b>
holiday	31	<b>83,9</b>	80,6	83,9	<b>83,9</b>	<b>83,9</b>
lady	53	62,3	58,5	56,6	62,3	<b>66,0</b>
material	69	52,2	52,2	46,4	52,2	<b>58,0</b>
mouth	60	63,3	60,0	55,0	63,3	<b>65,0</b>
nation	37	73,0	64,9	62,2	73,0	<b>78,4</b>
nature	46	56,5	54,3	52,2	56,5	<b>63,0</b>
post	79	67,1	64,6	60,8	67,1	<b>72,2</b>
restraint	45	60,0	60,0	55,6	60,0	<b>62,2</b>
sense	53	73,6	<b>75,5</b>	73,6	<b>75,5</b>	<b>75,5</b>
spade	33	<b>72,7</b>	<b>72,7</b>	66,7	<b>72,7</b>	<b>72,7</b>
stress	39	<b>51,3</b>	51,3	48,7	<b>51,3</b>	48,7
yew	28	<b>78,6</b>	<b>78,6</b>	<b>78,6</b>	<b>78,6</b>	<b>78,6</b>
Moyenne	98	65,3	62,9	61,5	66,1	<b>67,1</b>

Tableau 3 : Influence du système décisionnel sur fenêtre de contexte – données SENSEVAL-2 (en %)

### 4.3 Évaluation sur SENSEVAL-2

Pour vérifier les résultats obtenus sur le corpus SENSEVAL-1, les données d'évaluation de la tâche *lexical-sample* de SENSEVAL-2 ont été employées. Les résultats obtenus pour les 29 noms de cette évaluation sont présentés dans le Tableau 3. Il faut noter que les résultats obtenus dans ce tableau ne sont pas directement comparables avec les résultats obtenus sur l'évaluation précédente car le référentiel sémantique n'est pas le même. Comme pour les tests précédents, la précision moyenne décroît lorsque la taille de la fenêtre de contexte utilisée pour construire les arbres croît ( $k=3$ , 5 et 7). Par contre, cela n'est pas vrai pour chaque mot pris de manière individuelle. C'est le cas pour les noms *child*, *grip*, *sense*, qui voient leur précision moyenne améliorée de manière significative avec une fenêtre de contexte plus étendue. L'exemple de *sense* est éloquent puisque le gain est de presque 2% lorsque l'on passe d'une fenêtre  $k=3$  à  $k=5$ . Une analyse rapide des données montre que le gain est principalement dû à la possibilité de distinguer les acceptions « *sense of humour* », grâce à la présence du



## Peut-on Trouver la Taille de Contexte Optimale en Désambiguïsation Sémantique?

mot *humour* en position  $p=+2$  par rapport au mot à désambiguïser. Bien que l'élargissement de la taille de fenêtre de contexte soit bénéfique dans certain cas, cela reste toutefois une erreur dans la plupart des cas. Par exemple, le nom *nation* voit sa précision moyenne décroître de manière très importante avec l'élargissement de la taille de fenêtre.

De la même manière que dans la section 4.2, les colonnes *MaxStat* et *SCT+Cos* ont été calculées. Les observations sont les mêmes que sur le jeu de données précédent : un système décisionnel utilisé en seconde phase aide à l'optimisation automatique de la taille de fenêtre contextuelle. Il faut noter que le nom *nature* obtient un gain de 6,5% par rapport au meilleur score obtenu par les SCT seuls ( $k=3$ ). Une étude approfondie pour ce nom, montre que dans 41% des cas, les 3 sens proposés ( $k=3, 5$  et  $7$ ) sont identiques. Donc, le gain de 6,5% n'est obtenu que sur les 59% des tests restants, ce qui correspond à 27 tests. Cette approche ne permet cependant pas à tous les coups de sélectionner la meilleure fenêtre car, dans le cas des noms, la bonne réponse ne se trouve que dans 72,9% des cas dans les 3 sens proposés.

### 4.4 Discussion

Pour démontrer la capacité d'un tel système à utiliser des informations dans un contexte étendu, l'exemple suivant apporte des éléments de réponse intéressants :

*'furthermore , nothing have yet be say about all the research that do not depend on the collection of datum by the sociologist ( primary datum ) but instead make use of secondary datum - the wealth of material already available from other source , such as government statistics , personal diary , newspaper , and other kind of information .'*

Le nom *material* est présenté dans son contexte après pré-traitement, avec la partie grisée correspondant à la fenêtre maximale prise en compte par les SCT. Deux sens sont proposés par les SCT : **material%1:27:00::** ('*tangible substance that goes in the makeup of a physical object*') pour  $k=3$  et  $7$ , et **material%1:10:00::** ('*information that can be reworked into a finished form*') pour  $k=5$ . L'utilisation du contexte dans sa globalité, permet de faire le bon choix entre ces deux sens, notamment grâce aux mots *research, collection, newspaper* et *datum* qui décrivent une thématique précise (domaine documentaire). Il semblerait donc qu'un système décisionnel utilisant un contexte plus étendu soit capable de tirer profit d'indices à un niveau plus thématique.

Les tests conduits sur les verbes et les adjectifs ont montré des comportements similaires (voir Tableau 4). Il y a toutefois une nuance à apporter en ce concernant les verbes. Le score obtenu avec *SCT+Cos*, bien que supérieur de presque 3% aux *SCT* ( $k=3$ ), reste inférieur au score avec sélection optimale de taille de fenêtre par mot (*MaxStat*). Cela peut s'expliquer par le caractère très ambigu des verbes de la tâche SENSEVAL-2.

	SCT $k=3$	SCT $k=5$	SCT $k=7$	MaxStat	SCT + Cos
Noms	65,3	62,9	61,5	66,1	<b>67,1</b>
Verbes	50,7	50,4	49,9	<b>54,2</b>	53,5
Adjectifs	64,6	59,1	57,4	65,9	<b>66,4</b>
Total	59,1	57,0	55,9	61,1	<b>61,3</b>

**Tableau 4** : Améliorations apportées par un système décisionnel (en %)

Ces résultats nous ont permis de nous positionner dans les 3 premiers systèmes (dans les 5 premiers après re-soumission des tests) à moins de 3% du participant « vainqueur » de l'évaluation.

## 5 Conclusion

Nous avons montré, par nos expériences, qu'il est possible d'accroître les performances d'un système de désambiguïsation sémantique au-delà des performances optimales obtenues de manière statique. L'utilisation d'un système à adaptation dynamique pour la désambiguïsation sémantique a montré des améliorations substantielles. Bien que les verbes ne semblent pas profiter autant que les noms et les adjectifs de cette approche, le gain moyen reste tout de même de 2,3 %, toutes catégories grammaticales confondues. De futures expériences devront être menées en utilisant un contexte étendu, pour déterminer dans quelles mesures la désambiguïsation d'un mot se joue également dans un contexte plus large. D'autres pistes sont également à explorer, notamment concernant l'utilisation d'une grammaire fonctionnelle pour augmenter les champs d'application des questions des SCT. Cela pourrait profiter grandement aux verbes, principalement grâce aux structures : *sujet+verbe+objet*. Enfin, il serait intéressant d'utiliser un autre référentiel sémantique que *Wordnet* pour la généralisation par classes sémantiques. La présence de sens trop fins au niveau des mots en contexte, génère une multitude de CS possible pour ceux-ci. Par exemple le nom *dog* est également associé à la classe *humain* par l'intermédiaire de son sens familier qui est rarement utilisé.

## Références

- Breiman L., Friedman J. H., Olshen R. A., et Stone C. J. (1984). *Classification and Regression Trees*. Wadsworth International, Belmont, CA.
- Crestan E. et El-Bèze M. (2001a). *Improving Supervised WSD by Including Rough Semantic Features in a Multi-Level View of the Context*. SEMPRO Workshop, Edinburgh.
- Crestan E., El-Bèze M. et Loupy C. de (2001b). *Improving WSD with Multi-Level View of Context Monitored by Similarity Measure*. In Proc. of SENSEVAL-2 Workshop, Toulouse, pp. 67-70.
- Gale W., Church K. W., et Yarowsky D. (1993). *A Method for Disambiguating Word Senses in a Large Corpus*. Computers and the Humanities, 26 : pp. 415-39.
- Kaplan A. (1950) *An experimental study of ambiguity and context*. Mechanical Translation, (2:2), 39-46 (issue appeared in 1955).
- Kilgarriff A. et Rosenzweig J. (2000). *English SENSEVAL : Report and Results*. In Proc. LREC, Athens, Greece, 3 : pp. 1239-44. <http://www.itri.ac.uk/events/senseval>
- Kuhn R. et De Mori R. (1995). *The Application of Semantic Classification Trees to Natural Language Understanding*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(5) : pp. 449-60.
- Lachenbruch P. A. et Mickey M. R. (1968). *Estimation of Error Rate in Discriminant Analysis*, Technometrics, 10, no.1 pp. 1-10.
- Loupy C. de et El-Bèze M. (2000), *Using Few Clues can compensate the small amount of resources available for Word Sense Disambiguation*. LREC, Athens, Greece, 1 : pp. 219-23.
- Loupy C. de, El-Bèze M. et Marteau P. F. (2000), *Using Semantic Classification Trees for WSD*. Computer and the Humanities, Kluwer Academic Publishers, 34: pp. 187-92.
- Ney H., Martin S. et Wessel F. (1997). "Statistical Language Modeling Using Leaving-One-Out", in S. Young & G. Bloothoof (eds.), *Corpus-Based Methods in Language and Speech Processing*, Kluwer Academic Publishers, pp. 174-207.
- Salton G. et McGill M. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill, NY.
- Yarowsky D. (1992). *Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora*. In Proceedings of COLING-92, Nantes, France, pp 454-460.
- Yarowsky D. (1993). *One sense per collocation*. In Proceedings of the ARPA Workshop on Human Language Technology, Princeton, pp. 266-71.