Méta-Étiqueteur Adaptatif :
vers une utilisation pragmatique des ressources linguistiques

Gabriel ILLOUZ

LIMSI, CNRS
B.P.133, 91403 ORSAY Cedex
gabrieli@limsi.fr

Résumé

Le traitement automatique du langage requiert des corpus textuels de plus en plus volumineux, entre autres pour les étiqueteurs morpho-syntaxiques. Ces processus de traitement ne sont pas exempts d'erreurs. Dans l'optique d'améliorer cet étiquetage de corpus hétérogènes (composés de textes tout-venant), une approche adaptative au type de texte utilisant les ressources produites par une campagne d'évaluation sera proposée. Les résultats d'une première validation seront présentés sur les données MULTITAG. Les faits suivants sont constatés : les textes ne sont pas homogènes en terme de distribution de parties du discours, les classifications a priori ne fournissent pas une homogénéité en terme de performance et un même texte peut produire des variations positives pour un système et négatives pour un autre. De plus, il existe une relation entre la typologie de textes obtenue de façon non supervisée sur le jeu de caractères et les variations de performance.

1. Introduction

Le traitement automatique du langage requiert des corpus textuels de plus en plus volumineux pour l'acquisition de connaissances, l'extraction d'information, ainsi que la mise au point et l'évaluation de systèmes. Les traitements ou chaînes de traitements (étiqueteur morpho-syntaxique, analyseur syntaxique, désambiguïsation sémantique) ne sont pas exempts d'erreurs (ou s'ils le sont, c'est toujours par rapport à un corpus). Devant le nombre de choix technologiques existants pour diminuer ce nombre d'erreurs, il devient nécessaire d'évaluer ces différents choix pour identifier le plus performant. De ce fait, le nombre de campagnes d'évaluation va croissant : MUC (HIRSHMAN 98), TREC (HARMAN 98), GRACE (ADDA *et al.* 97), SENSEVAL (KILGARRIFF 98). En retour, ces protocoles d'évaluation permettent de rendre objectives les améliorations apportées à un traitement. Dans l'optique d'améliorer l'étiquetage morpho-syntaxique de corpus hétérogènes (composés de textes tout-venant), une approche adaptative au type de texte utilisant les ressources produites par une campagne d'évaluation sera proposée.

Tout d'abord, nous examinerons les méthodes existantes pour rendre les étiqueteurs plus performants (section 2). Ensuite, la méthode retenue pour s'adapter à des situations (des types de textes) hétérogènes sera exposée (section 3) et les résultats d'une première validation (section 4) sera effectuée sur les données MULTITAG (projet du CNRS de valorisation des ressources produites par la campagne d'évaluation GRACE). Enfin, nous présenterons le travail qui reste à effectuer pour valider complètement notre approche ainsi que les perspectives de développements futurs de la méthode.

2. État de l'art

Les raisons d'étudier une approche adaptative au type de texte à l'aide des étiqueteurs morpho-syntaxiques sont les suivantes : Tout d'abord, ils sont bien souvent un maillon de départ pour nombre de chaînes de traitements. Toute amélioration, même faible, est donc pertinente. Ensuite, leur protocole d'évaluation est presque unanimement accepté, contrairement à d'autres traitements pour lesquels arriver à un consensus sur le protocole d'évaluation relève du défi (par exemple, l'évaluation de résumés automatiques ou de dialogues Homme-Machine). Enfin, nombre de types d'étiqueteurs existent reposant sur des technologies variées (stochastique, à base de règles de transformation, à base de règles linguistiques) et entraînés sur des types de textes différents (extraits de journaux, de romans, de documentations techniques).

Une approche possible pour améliorer les performances d'un étiqueteur est de l'entraîner et/ou de lui ajouter de nouvelles ressources. Cependant, cette approche est coûteuse en temps et en travail. De plus, il est peu probable qu'un seul étiqueteur, avec une technologie et des ressources données soit adapté pour traiter tout type de texte.

Une autre approche existante est la **méthode des votants**. Dans (MÀRQUEZ *et al.* 98), partant d'un petit corpus d'entraînement, les auteurs utilisent deux étiqueteurs pour l'annoter. Ensuite, ils considèrent les résultats des étiqueteurs sur un corpus plus grand. Celui-ci leur sert à nouveau de corpus d'entraînement en donnant un poids plus fort aux étiquettes sur lesquelles les deux systèmes s'accordaient. L'opération est alors répétée en boucle jusqu'à stabilisation des performances. Cette méthode peut aussi être utilisée avec un plus grand nombre d'étiqueteurs. L'étiquette ayant le nombre de suffrages le plus important est retenue. Cette hypothèse est implicitement celles des projets d'étiquetage multiple de corpus (MULTITAG, AMALGAM). Cependant, cette approche reposant sur un méta-étiqueteur[i] soulève un problème qui nous semble crucial. En effet, un processus très bien entraîné pour une situation donnée peut avoir des performances médiocres pour d'autres. Pour illustration, soit trois locuteurs respectivement Français, Espagnol, et Anglais. Si le premier et le deuxième sont du même avis sur la façon d'étiqueter un mot en anglais, à la différence du troisième, leur choix devrait-il prévaloir ?

Ce qui nous amène a nous pencher sur des **méthodes adaptatives**, basées sur "le plus compétent l'emporte". Dans l'article *How One might Automatically Identify and Adapt to a Sublanguage : An Initial Exploration* (SLOCUM 86), l'auteur montre l'existence de règles syntaxiques différentes à utiliser selon deux types de textes, en allemand. Ces deux types de textes sont composés de deux manuels écrits par des ingénieurs et de deux brochures écrites par des commerciaux. Il propose aussi un moyen de caractériser le type "manuel" (impératif, acronyme, suppression de déterminants) par rapport au type "brochure" (phrases longues, utilisation des pronoms, syntaxe plus riche). Ce type d'approche nous paraît tout particulièrement intéressant. En effet, il permet de mieux appréhender la notion de domaine de compétence d'un processus de traitement automatique.

Dans (BIBER 93), l'auteur montre que le type de textes peut influer sur les résultats des étiqueteurs probabilistes. Pour ce faire, il sélectionne, à partir du corpus LOB, deux types de textes (textes explicatifs et romans). Il met en évidence des différences de probabilités des enchaînements de deux parties du discours d'un type de texte à l'autre. S'inscrivant dans ce paradigme, cet article propose une méthode pour choisir selon les données l'étiqueteur le plus adapté.

---

[i] Nous emploierons le terme **méta-étiqueteur** pour désigner un processus utilisant les sorties de plusieurs étiqueteurs pour améliorer les performances globales

3. Modélisation proposée

Le but de notre approche est non seulement de distinguer des situations, mais aussi et surtout de reconnaître des situations proches sur lesquelles on appliquera des traitements identiques.

3.1 Distinguer des situations (type de textes)

Les méthode de classification peuvent se résumer ainsi : à chaque objet (pour nous, un texte du corpus), on associe un vecteur le représentant. Puis, on utilise ces vecteurs pour obtenir une classification des types de textes présents. Le choix des traits[ii] est capital, et nombre de méthodes sont pertinentes pour représenter un texte. Une partie du vocabulaire présent dans le texte peut être utilisée (BENZÉCRI *et al*. 81). Cependant, cette méthode se place dans une perspective de classification dans laquelle le choix du vocabulaire ne peut être fixé une fois pour toutes, sinon cela nécessiterait des vecteurs d'une taille considérable. Les traits résultant d'un étiquetage (exemple : nombre d'adjectifs) peuvent aussi être utilisés (BIBER 93). Cette méthode permet une finesse de résultats appréciable. Néanmoins, elle nécessite d'avoir soit un corpus déjà étiqueté, soit un étiqueteur. Utiliser un étiqueteur réalisant un faible taux d'erreurs pourrait sembler peu problématique étant donné qu'un vecteur représente statistiquement un texte. Malheureusement, l'existence d'un étiqueteur robuste (faisant la même proportion d'erreurs) quel que soit le type de textes ne peut être garantie.

Par conséquent, tant que l'existence d'un tel étiqueteur ne peut être prouvée, la classification se fera sur des traits surfaciques. Le vecteur choisi pour représenter un texte sera composé des fréquences relatives des caractères. Cette information peut sembler bien pauvre. Néanmoins, elle a tout d'abord l'avantage de ne nécessiter aucun pré-traitement linguistique. Ensuite, elle est représentative d'un certain nombre de caractéristiques de la forme du texte. En effet, la fréquence du point est corrélée avec la longueur des phrases (ou plus précisément, la résultante de la longueur des phrases et la présence d'abréviations), celle de l'espace avec la longueur des mots, la proportion de majuscules avec celle de la proportion de noms propres. La présence des caractères spéciaux : + * / **: - ' ( ) , ? ! $** , etc., peut elle aussi se révéler significative (par exemple pour distinguer un texte scientifique).

Les vecteurs constitués permettent de classer les textes de façon non supervisée (par similarité des vecteurs), ou supervisée (ayant pour but de correspondre à une classification *a priori*). Dans les deux cas on obtient un certain nombre de classes (de types de texte). La classification non-supervisée sera préférée, car on ne sait pas si la classification optimale (corrélée à la classification des meilleurs étiqueteurs) correspond à une connaissance a priori. L'exploration est donc privilégiée dans un premier temps. Un texte nouveau peut alors être associé à la classe dont il est le plus proche. Si sa distance aux classes existantes est trop grande, il sera le premier individu d'une nouvelle classe.

Partant du corpus d'entraînement T composé de l'ensemble de textes $\{T_1, T_2, ... , T_n\}$, et d'une fonction de caractérisation f définie par : $f(T_i)= V_i$, où V_i est le vecteur caractéristique du texte T_i, une méthode de classification automatique fournit une partition D de l'ensemble T, $D=\{D1, D2, ... , Dk\}$, avec $k<n$. Elle fournit aussi une fonction d'appartenance h : $h(Vi) = Dm$. Cette classification permet alors de construire le méta-étiqueteur.

[ii] En classification, un **trait** est une coordonnée du vecteur représentant les valeurs possibles d'une caractéristique jugée pertinente dans la prise de décision.

3.2 Construction du méta-étiqueteur adaptatif (MEA)

L'évaluation des différents étiqueteurs fournit pour chaque étiqueteur un score pour chaque texte. Pour chaque type de texte, l'étiqueteur présentant la meilleure performance, en moyenne, sur ce type est retenu. D'où, soit E l'ensemble des étiqueteurs $\{e_1, e_2, ... , e_j\}$, on définit une fonction $ME(D_i) = max_j(perf(e_j, D_i))$

où $perf(e_j,D_i)$ donne la performance de e_j sur le sous-corpus D_i.

Une fois le méta-étiqueteur créé, il est possible de s'en servir sur un texte nouveau T_x. Pour ce faire, on suit l'algorithme suivant:

- $D_x = h[f(T_x)]$
- si $D_x \in D$, l'étiqueteur $ME(D_x)$ est sélectionné
- sinon,
    - soit on demande un nouvel étiqueteur pour ce nouveau type de texte,
    - soit on prend $ME(D_i)$ avec $min_i (D_x, D_i)$

3.3. Protocole d'évaluation

Le but est de mesurer l'amélioration obtenue par le modèle. Pour ce faire, la performance moyenne du MEA sur des textes nouveaux doit être comparée avec :

- celle des meilleurs étiqueteurs utilisés par le MEA;
- celle du meilleur MEA possible (en choisissant manuellement le meilleur traitement pour chaque texte);
- celle d'autres méta-étiqueteurs (par exemple, basé sur la méthode des votants).

Cette méthode d'évaluation permet de vérifier que la combinaison choisie d'étiqueteurs apporte un gain, que la méthode de classification est optimale, que cette technologie est performante.

4. Exploration des données.

La situation idéale pour tester notre modèle serait:

- un corpus étiqueté manuellement et composé de types de textes hétérogènes;
- l'homogénéité et un volume suffisant à l'intérieur de chaque type de textes;
- des étiqueteurs de maturité équivalente (ayant des performances moyennes comparables) entraînés sur des types de textes différents.

Or, les données disponibles pour le français (cf. partie suivante) ne peuvent être considérées comme suffisantes pour prouver la validité de notre modèle. Néanmoins, ces données nous permettent une première exploration ayant pour but de justifier la direction choisie précédemment, et de répondre aux questions :

1) Y a-t-il variation de performance selon la situation ?
2) Est-il possible de distinguer des type de textes à partir des caractères ?
3) Est-il alors possible de les classer de manière automatique ?

4.1 Données utilisées.

Le corpus MULTITAG, en cours de construction, représente un volume total de un million de formes lexicales. 650 000 sont utilisées dans la phase de test de GRACE. Il est composé de seize extraits du journal *Le Monde*(texte 11 à 26) et de dix textes issus de FRANTEXT. Pour FRANTEXT, il s'agit de six romans (textes : 2, 3, 5, 7, 8, 9), de deux essais (textes 4 et 10[iii]), et des mémoires du Maréchal Foch et de B. Constant (textes 1 et 6).

[iii] respectivement de Physiologie et de Psychologie

Les résultats d'évaluation présentés sont dérivés de ceux de GRACE (ADDA *et al*. 97). Nous considérons onze systèmes ayant étiqueté le corpus précédent. Leur performance en terme de précision et de décision[iv] est alors évaluée par rapport à la référence (étiquetée à la main). Cette référence n'existe que pour un sous-corpus. Celui-ci contient 100 000 formes provenant de deux extraits du *Monde* et d'échantillons contenant 5 000 formes consécutives tirées au hasard pour chaque texte issu de FRANTEXT.

Nous ne connaissons pas la technologie des systèmes. Nous pouvons néanmoins donner les technologies des deux systèmes minimaux (*baseline*) utilisés. Le premier système, M1, est lexical. Il repose sur un simple accès au dictionnaire MULTEXT-GRACE. Si la forme n'y est pas trouvée, elle reçoit l'étiquette des mots inconnus. Le second, M2, raffine l'approche lexicale en appliquant les heuristiques[v] suggérées par J. Vergne (GREYC) dans le cadre de l'action GRACE.

Ces données, quoique ressources appréciables pour le Français, ne sont pas encore en nombre suffisant pour permettre toutes les expériences nécessaires à la validation du modèle proposé. Les outils d'évaluation seront disponibles au printemps prochain (Projet européen ELSE), permettant entre autres d'évaluer une portion de texte quelconque par rapport à la référence. Ceci nous empêche, pour l'instant, d'évaluer plus finement les résultats, voire de nouveaux étiqueteurs.

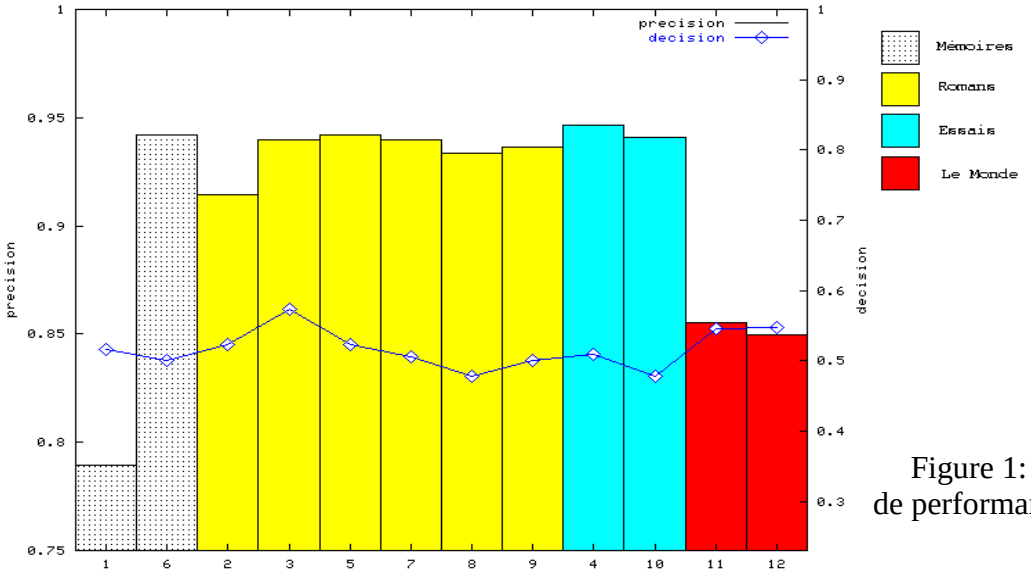

Figure 1: Variation de performances de M1

4.2 Observation des variations de performances selon situations.

Étant donné le faible nombre de textes (12), les observations possibles relèveront de la statistique descriptive simple. Les résultats des différents systèmes ont été observées, texte par texte. Une synthèse est exposée ci-dessous.

[iv] La **précision** est mesurée par le pourcentage de réponse correcte, la **décision** par le pourcentage de cas où le système renvoie un seul choix (et non plusieurs possibles).

[v] Ces heuristiques sont par ordre de priorités:

1. en cas d'ambiguïté Nom/Verbe, choisir Nom;
2. en cas d'ambiguïté Adjectif/Verbe, choisir Adjectif;
3. en cas d'ambiguïté Adverbe/Nom, choisir Adverbe;
4. en cas d'ambiguïté Nom/Adjectif, choisir Nom;
5. en cas d'ambiguïté Déterminant/Pronom, choisir Déterminant;
6. autrement choisir au hasard une étiquette parmi celles proposées.

A titre d'exemple, les résultats frappants du système M1 sont présentés en figure 1. Sur celle-ci, les abscisses représentent les textes (FRANTEXT de 1 à 10, réordonnées par genres, *Le Monde* de 11 à 12). En ordonnées, sont représentées la précision à l'aide de barres selon l'échelle à gauche et la décision à l'aide de points reliés selon l'échelle à droite. L'effet des textes sur les performances est très marqué : les textes 1, 11, et 12 subissent d'importantes pertes de précision.

Il s'agit alors de savoir si les autres systèmes ont des variations aussi prononcées (en précision). La mesure de l'écart type (σ) de la précision pour chaque système définit une mesure de l'influence des textes sur les performances. Cette mesure représente la robustesse d'un système aux textes présents. Ces mesures ont été calculées pour chaque système et sont données dans la tableau 1. Les variations d'un système à un autre sont donc importantes : allant de moins de 1% à plus de 6%. A quels textes sont dues ces variations ? Pour répondre à cette question, pour chaque système ont été définies une zone de variations non significatives (la moyenne plus ou moins 1σ), une zone peu significative (entre 1 σ et 1,5 σ de part et d'autre de la moyenne) et significatives (au-delà de 1,5 σ). Dans le tableau 2, pour chaque système, les textes situés dans la zone peu significative sont indiqués par un "+" (au dessus de la zone significative) ou par un "-" (en-dessous). De la même façon, les textes situés dans la zone significative sont indiqués par "++" ou "--".

Le texte 1 est donc celui qui crée le plus de variations : négatives pour la plupart, positives pour deux, nulle pour un. L'autre texte de "mémoire" (10) n'a pas d'effets aussi marqués. Parmi les romans, les textes 2, 5, 7, 9 produisent peu de variations, alors que les textes 8 et 3 sont cause de baisse de performances. Toutefois, il est à noter que la classe "romans" est peu homogène. Par exemple, il y a une nette différence de style entre Gustave Flaubert (texte 8) et Gaston Leroux (texte 2). Les essais produisent peu de variations. Les deux textes du *Monde* ont quant à eux un profil très comparable.

S1	M1	M2	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
6.47	4.88	2.66	2.57	2.10	1.81	1.51	1.33	1.24	1.06	0.99	0.94	0.88

Tableau 1 : Sensibilité aux textes (du plus au moins sensible en pour-cent)

		S1	M1	M2	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
Mémoires	t1	--	--	--	--	--	++	--	--	--	++	--		-
	t6											+		++
Romans	t2													-
	t3			++				--						
	t5						-	-						
	t7							+						
	t8						--				--	-		--
	t9							+						
Essais	t4				+									
	t10							+		+				
Le Monde	t11	-	-										--	
	t12	-	-										--	

Tableau 2 : influence de chaque texte pour chaque système.

Pour expliquer ces variations, au moins dans un contexte probabiliste, la méthode de Biber (BIBER 93) est reprise. Les probabilités les plus fortes des parties du discours suivant un nom sont présentées pour différents groupes de textes dans le tableau 3. Les parties du discours ont été extraites d'étiquettes plus complexes. Il est à noter dans ce tableau les ressemblances

existant entre le texte T1, et *Le Monde*, le groupe de textes 2,5,7,9 (romans) et les textes 6,4, et 10. Les distribution particulières des textes 8, et 3 rendent plus difficile à les rapprocher d'autres textes. Ces distributions sont à mettre en relation avec le tableau précédent. Mais des études plus approfondies de ces résultats sont nécessaires, car d'autres distributions (pour chaque partie du discours) sont à observer avant de conclure en terme de ressemblance.

		Textes 2,5,7,9	Le Monde	T1	T8	T3	T6	T4	T10
Nom	Ponctuation	36.2	26.2	26.5	43.1	39.8	31.9	34.3	31.0
	Préposition	21.6	27.6	31.5	15.0	22.7	28.9	22.7	22.9
	Adjectifs	10.2	13.3	8.7	7.6	12.4	5.8	10.5	13.9
	Pronom	7.5	2.6	3.7	10.0	4.7	9.6	12.2	8.5
	Verbe	5.7	6.2	6.0	13.8	5.8	2.7	8.3	8.0
	Conjonction	7.1	4.0	4.3	6.11	4.2	7.2	3.9	7.1
	Adverbe	4.2	2.0	1.9	2.4	3.8	5.2	5.1	5.4
	Nom	5.3	16.8	15.7	0.5	5.3	7.6	1.1	2.0

Tableau 3 : Fréquences de suite de parties du discours

Finalement pour répondre à la question initiale, des effets plus ou moins importants des textes sur les différents étiqueteurs ont pu être observés. Les faits qu'il nous semble important de retenir sont 1) que les textes ne sont pas homogènes en terme de distribution de parties du discours, 2) que les classifications a priori ne fournissent pas une homogénéité en termes de performance et 3) qu'un même texte peut produire des variations positives pour un système et négatives pour un autre.

4.3 Étude de la fonction de caractérisation

Les données utilisées dans cette sous partie sont celles de la phase de test (650 000 formes), d'où un plus grand volume de données et plus d'extraits du *Monde*. Pour savoir si l'emploi des variations en fréquence des caractères est significatif, le point et la virgule seront pris comme exemple. Observons au préalable ces distributions pour les différents textes, le point en figure 2, la virgule en figure 3. La fréquence du point est remarquablement stable pour les numéros du journal *Le Monde*. Pour la partie FRANTEXT, on ne distingue pas de comportement aussi significatif. La virgule globalement moins employée dans *Le Monde* que dans FRANTEXT, ne peut être retenue comme distinctive d'une catégorie à l'autre. Nous allons voir ci-dessous si ses caractéristiques mènent à une classification automatique intéressante.

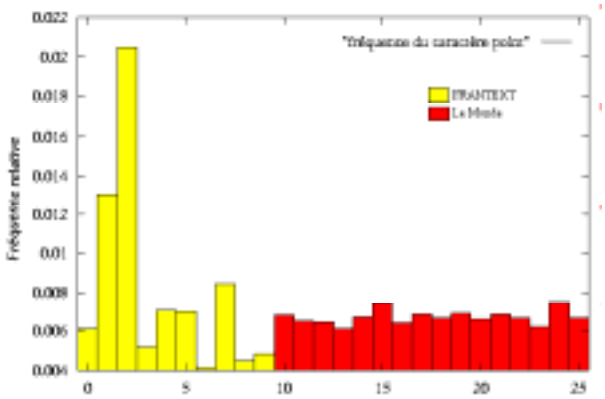

Figure 2 : fréquence du point

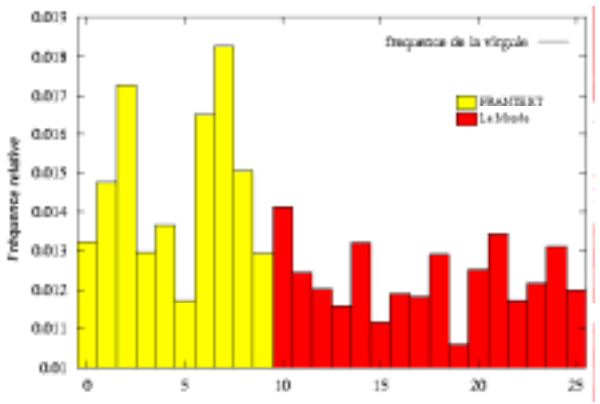

Figure 3 : fréquence de la virgule

4.4 Possibilité de classification

Notre but n'est pas de classer les données mais simplement d'observer si ces données se prêtent à une classification. Les vecteurs caractéristiques sont explorés à l'aide de la méthode de Sammon (SAMMON 69) qui, partant d'une dimension n, projette les données dans un espace de dimension k (k<n), avec la propriété de conserver au mieux les distances[vi] existant dans l'espace de départ. De plus, cette méthode peut être utilisée comme préalable à une méthode de classification automatique simplifiant ainsi les données et rendant la classification plus performante (LERNER 98).

La figure 4 est obtenue par cette méthode sur l'espace défini par le jeu de caractères. Les points de cet espace à n dimensions[vii] projeté en deux dimensions représentent des textes. Le corpus utilisé est celui de test.

Sur cet espace projeté, les textes du *Monde* sont proches les uns des autres. Les romans (2, 3, 5, 7, 8, et 9) bien que moins regroupés que les textes du *Monde* constituent toutefois un ensemble identifiable. Pour les essais (4 et 10), le regroupement est moins direct. Néanmoins, le texte 10 a pour plus proche voisin le texte 4. Les mémoires (1 et 6) se situent à l'opposé dans le plan de projection. Le texte 1 a pour voisins les textes du *Monde* tandis que le texte 6 jouxte les romans.

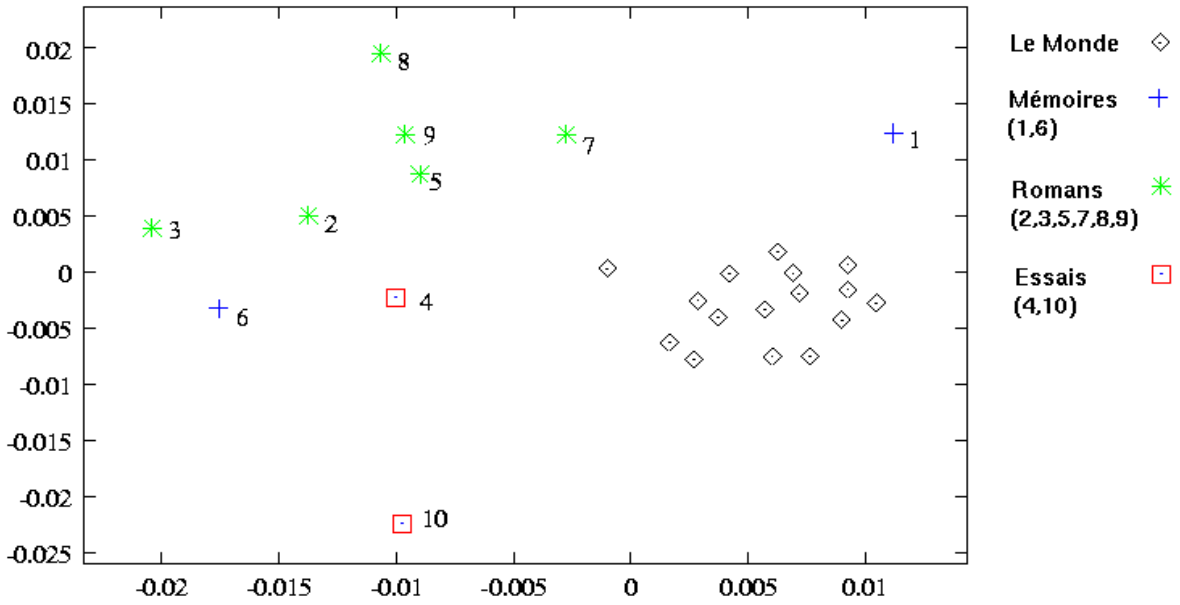

Figure 4 : Projection de Sammon des textes.

A l'exception du type "Mémoire", la classification FRANTEXT, définie a priori, et le type textes du *Monde* peuvent être retrouvés dans l'espace de projection. Cette exception est à mettre en relation avec les variations de performance (cf. section 4.2). En effet, les systèmes P1 et M2 obtiennent des performances faibles pour le texte 1 et sur les textes du *Monde*. La plupart des systèmes obtiennent peu de variations de performance pour le texte 6, ce qui le différencie fortement du texte 1.

[vi] Dans l'expérience qui suit, la mesure de distance utilisée est euclidienne.

[vii] L'espace a été enlevé de la représentation, car dans FRANTEXT, l'apostrophe résultante d'une élision ajoute un espace. par exemple "`l'apostrophe`" devient "`l'  apostrophe`"

Cette description montre qu'il existe une relation entre la typologie de textes obtenue de façon non supervisée sur le jeu de caractères et les variations de performance. A ce stade de la recherche, il est prématuré de développer le MEA. En effet, trop peu de résultats d'évaluation sont disponibles, à ce jour pour le français, pour valider une typologie en corrélation avec les performances.

5. Conclusion et Perspectives.

Dans le présent article, la modélisation d'un méta-étiqueteur adaptatif au type de texte a été exposée. Son utilité a été justifiée par les explorations des données MULTITAG : un même texte peut produire des variations positives pour un système et négatives pour un autre; ces variations peuvent s'expliquer par rapport à une classification automatique.

Pour développer et valider notre approche sur le français, une nouvelle campagne d'évaluation est nécessaire sur des données plus appropriées, c'est-à-dire plus proches de la situation idéale (un corpus étiqueté et hétérogène, chaque type de textes homogène et en volume suffisant, et des étiqueteurs de maturité équivalente entraînés sur des types de textes différents).

Dans les perspectives d'évaluation de cette méthode, une évaluation similaire sur un corpus plus proche de la situation idéale (tel le BNC, British National Corpus) et des étiqueteurs entraînés sur des domaines différents est envisagé. Dans les perspectives d'extension de la construction d'un méta-étiqueteur adaptatif utilisant la méthode des votant est aussi prévue, avec les suffrages des étiqueteurs pondérés selon le type de textes. Enfin, la classification des types de textes, nécessite elle aussi des approfondissements. Des méthodes basées sur des traits plus nombreux, tant surfaciques que linguistiques, restent à étudier.

Remerciements

Je tiens à remercier toutes les personnes qui m'ont aidé dans ce travail. Tout particulièrement C. Jacquemin, B. Habert, V. Prince pour leurs conseils, et leurs soutiens. Mais aussi et surtout P. Paroubek sans qui cela ne serait pas.

Références

ADDA *et al.* (1997), G. Adda, J. Lecomte, J. Mariani, P. Paroubek, M. Rajman, *Les procédures de mesure automatique de l'action GRACE pour l'évaluation des assignateurs de Parties du Discours pour le Français,* Actes des 1 ères Journées Scientifiques et Techniques du Réseau Francophone de l'Ingénierie de la Langue de l'Aupelf-Uref, Avignon.

BIBER D. (1993), *Using Register-Diversified Corpora for General Language Studies*, Computational Linguistics, Vol 19, 2.

BENZÉCRI *et al*. (1981), Benzécri, J.-P. et Collaborateurs, *Pratique de l'analyse des données*, Dunod.

HARMAN D. (1998)**,** *The Text REtrieval Conference (TRECs) and the Cross-Language Track* , in Proceedings of the First International Conference on Language Resources and Evaluation (LREC), Granada.

HIRSHMAN L. (1998), *Language Understanding Evaluations: Lessons Learned from MUC and ATIS*, in Proceedings LREC, Granada.

KILGARRIFF A. (1998), *SENSEVAL: An Excercise in Evaluating Word Sense Disambiguation Programs*, in Proceedings of LREC, Granada, 1998.

LERNER *et al.* (1999), B. Lerner, H. Guterman, M. Aladjem and I. Dinstein, *A Comparative Study of Neural Network Based Feature Extraction Paradigms*, Pattern Recognition Letters, vol. 20(1), pp. 7-14.

MÀRQUEZ *et al.* (1998), L. Màrquez, L. Padró & H. Rodríguez, *Improving Tagging Accuracy by Using Voting Taggers Proceedings* NLP+IA/TAL+AI 98. Moncton, New Brunswick. Canada.

SAMMON J. (1969), *A nonlinear mapping for data structure analysis*. IEEETC 18, 5, pp. 401-409.

SLOCUM J. (1986), *How One might Automatically Identify and Adapt to a Sublanguage : An Initial Exploration*, in Analyzing Language in Restricted Domains : Sublanguage Description and Processing, Grishman, R., Kittredge, R. (eds.), Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 195-210.