

DE L'UTILITÉ DE LA SÉMANTIQUE TEXTUELLE COMME MEDIUM ENTRE CORPUS ET ANALYSE

Les jeunes de Chirac ; analyse sur corpus numérisé

Baptiste FOULQUIÉ
CPST, Université de Toulouse 2

SOMMAIRE

1. Préambule
2. Analyses
 - 2.1. Constitution des molécules sémiques, analyse micro et mésosémantique
 - 2.1.1. Les cooccurrences, nébuleuse de signifiants
 - 2.1.2. Les corrélats, réseaux de signifiés
 - 2.2. Analyse macrosémantique et textualisation des formes sémantiques
3. Conclusion
 - 3.1. Remarques sur l'unité du corpus et l'importance d'une répartition problématisée des textes dans un corpus
 - 3.2. Remarques sur l'analyse thématique et sur la lexicalisation des thèmes
 - 3.3. Évocation du concept de para synonymie thématique

***Résumé :** Le développement des logiciels dit d'analyse de texte s'accompagne de nombreuses analyses sur grands corpus et intéresse la sémantique.*

Nous tenterons de voir dans ce travail, comment la sémantique peut tirer avantage des outils que sont ces logiciels, sans pour autant leur laisser la place qui est la sienne dans le champ de l'analyse du contenu des textes. Nous tenterons de faire la part, concernant l'analyse thématique, entre ce qui relève de la compétence du manipulateur du logiciel, et ce qui relève de la compétence du sémanticien. Nous nous plaçons ici dans la perspective de F. Rastier, et tentons d'utiliser les concepts de la sémantique interprétative (isotopies et paratopies, thèmes spécifiques et molécules sémiques).

Prenant prétexte des travaux de Damon Mayaffre, à qui nous devons notre corpus, nous verrons l'importance que peuvent prendre les précautions méthodologiques, notamment en ce qui concerne la répartition raisonnée des textes lors de la création du corpus, sur l'interprétation des textes.

Nous reviendrons ensuite sur les concepts qui justifient l'analyse des cooccurrences, en nous concentrant sur la paratopie qui permet l'établissement des molécules sémiques représentant les thèmes.

1. Préambule

La numérisation de grands corpus ainsi que l'apparition de logiciels permettant leur traitement permet aujourd'hui l'émergence d'une pratique analytique nouvelle : l'analyse textuelle assistée par ordinateur. Notre propos sera ici prioritairement centré sur la pratique qui consiste à passer un corpus au crible d'un logiciel (ici Hyperbase). En effet, l'ergonomie logicielle est telle qu'elle n'incite pas toujours l'utilisateur à avoir un regard critique sur sa pratique, mais elle peut en revanche l'inciter à dégrader sa théorie en fonction des limitations et des objectifs de la pratique. Cette adaptation étant nécessaire, il n'en reste pas moins qu'elle doit être questionnée et si possible limitée.

Cette étude prend prétexte des travaux réalisés par Damon Mayaffre dans son ouvrage : *Paroles de président, Jacques Chirac (1995-2003) et le discours présidentiel sous la V^{ème} république*. Rappelons que le corpus de Mayaffre est composé de l'intégralité des discours des présidents de la V^{ème} république. Dans ce corpus, l'auteur compare chaque sous-corpus (constitué des textes d'un président) au corpus global, avec une attention particulière au sous-corpus Chirac.

Le point de départ de notre analyse est la constatation que fait Mayaffre à la page 140, à propos de la forme *jeunes*. Cette forme est statistiquement la troisième forme la plus discriminante du discours chiraquien lorsqu'on le contraste sur les discours des autres présidents de la V^{ème} république. Il titre « plaire : les jeunes ». Selon lui, l'utilisation du mot *jeunes* a « trois fonctions politiques évidentes » qui sont :

- « - d'abord le mot permet de dépasser les catégories sociales et partisans habituelles. Les « jeunes » sans distinction de classes et de parti se trouvent considérés en bloc. [...]
- ensuite, loin de la gratuité linguistique, la sur-présence des « jeunes » est le signe fort de la logomachie engagée par Chirac et Jospin. À partir de 1997, les jeunes sont devenus un enjeu de pouvoir politique patent entre le président et son premier ministre. [...]
- ultime hypothèse : aux antipodes de l'innocence linguistique, la sur-présence de « jeunes » dans le discours peut relever des manipulations de la nouvelle communication politique (discours émotionnel plutôt que rationnel pour attirer la sympathie, [...] touche l'affect.)¹»

C'est, en fait, la première remarque qui a motivé une étude plus approfondie de cette forme. La question qui se pose est de savoir si le signifiant *jeunes* lexicalise un thème comme pourrait le laisser penser la citation précédente (création d'une entité subsumant les différences entre les différents jeunes), s'il en lexicalise plusieurs en fonction des contextes (p.142 « "l'intégration des jeunes" apparaît comme un moyen lexical détourné pour traiter des catégories issues de l'immigration »), s'il ne lexicalise qu'une partie d'un thème, ou s'il n'en lexicalise aucun.

Remarque : rappelons à ce propos les mises en garde de Rastier² concernant les difficultés inhérentes au passage de l'analyse lexicale à l'analyse thématique :

« à la différence des lexèmes, les thèmes ne sont pas des signes, ni, corrélativement des unités du français : ils dépendent en effet d'autres normes que la langue. Si le lexème et le thème diffèrent aussi bien par le niveau que par le palier d'analyse, le premier étant un signe, et relevant de la morphologie et de la microsémantique, le second une unité du contenu au palier mésosémantique, il est clair que tout lexème n'est pas un thème.

Une analyse thématique qui en resterait au palier lexical compterait potentiellement autant de thèmes que de mots de la langue. [...] On objectera que les thèmes sont ordinairement dénommés par un lexème. Mais ce lexème est simplement une lexicalisation privilégiée du thème. Et l'on pourrait fort bien rencontrer des thèmes sans lexicalisation privilégiée.³ »

Rastier donne plusieurs exemples de ce phénomène, notamment celui de la molécule de l'ennui dans *Madame Bovary*, molécule dont seules certaines parties sont lexicalisées⁴. Un autre exemple est celui du nombril chez Flaubert qui est selon lui :

« une lexicalisation partielle d'un thème dans la correspondance, d'un motif dans deux romans, et d'un topos isolé dans la correspondance. »

Cette problématique de la lexicalisation des formes sémantiques doit être appréhendée dans la perspective d'une théorie des genres. On peut donner cette citation de Rastier qui permet de faire le lien entre ces deux problématiques, et qui nous permettra de passer à la présentation de notre corpus :

« Résumons, un lexème peut ne lexicaliser aucun thème, par exemple, le mot thème ne correspond à aucun thème dans le corpus romanesque que nous avons étudié, mais il peut aussi en lexicaliser plusieurs. Enfin, son lien avec le palier thématique est relatif à un discours (littéraire, médical, etc.), un genre, et un corpus. »

Nous revenons donc dans ce travail sur l'analyse du corpus Chirac. Il faut noter cependant, qu'à la différence de Mayaffre, nous avons réorganisé le corpus Chirac, non plus seulement en fonction de l'ordre chronologique, mais d'abord en fonction des situations et du mode de communication. Encore loin d'une typologie des genres constitutifs de la pratique présidentielle, nous avons distingué dans un premier temps trois sous corpus en fonction du public (international, national ou régional), eux-mêmes divisés en deux en fonction du mode de communication (interactif ou non) : interviews ou discours. Cette répartition des discours entraîne des résultats sensiblement

¹ Mayaffre, D., 2004, pp. 140-143.

² Rastier, F., « La sémantique des thèmes ou le voyage sentimental » www.revue-texto.net

³ Nous pouvons peut-être même aller plus loin en proposant qu'un thème spécifique n'a, par définition, pas de lexicalisation puisqu'il est abstrait de tout domaine. Lexicaliser un thème conduit à l'indexer sur un domaine.

⁴ Rastier, F., *Arts et sciences du texte*, pp. 200-201

différents de ceux de Mayaffre. Ce constat était prévisible, on se souvient en effet que Brunet a montré la dominance des contraintes liées au genre sur celles liées à l'auteur :

« Car de toutes les forces qui s'exercent sur un texte, le genre semble la plus pesante et la plus pressante. Nous gardons le souvenir décevant d'une expérimentation, réalisée avec Charles Muller, qui avait consisté à étudier les 60 mots français les plus fréquents dans une dizaine de textes (romanesques, dramatiques ou poétiques) de Hugo, Lamartine et Musset. Abusé par les méthodes statistiques, l'ordinateur avait reconnu des différences et invitait naïvement à conclure qu'il y avait trois auteurs différents : un romancier, un dramaturge et un poète.¹ »

Notre méthode consistera dans un premier temps à observer le réseau de collocations de la forme *jeunes*, à l'aide de la fonction « contexte » du logiciel Hyperbase. On appliquera ensuite la fonction thème aux résultats obtenus, afin d'obtenir la liste des termes les plus fréquemment associés à cette forme. Les collocations relevées devront faire l'objet d'une analyse afin de préciser les relations qu'elles entretiennent avec le mot pôle et de déterminer leur degré de corrélation. Les résultats de cette phase nous permettront d'avoir une première idée sur le ou les thèmes que lexicalise en tout ou partie la forme *jeunes*. Nous pourrons alors étendre les requêtes aux corrélats les plus intéressants.

Cette étude a donc deux objectifs, le premier concerne l'importance d'une répartition problématisée du corpus en sous-corpus. Le second consiste, à travers l'étude détaillée des contenus lexicalisés par une forme, à évaluer les concepts nécessaires à l'analyse et ceux qui peuvent disparaître dans une dégradation raisonnée de la théorie.

2. Analyses

2.1. Constitution des molécules sémiques, analyse micro et mésosémantique

2.1.1. Les cooccurrences, nébuleuse de signifiants

Dans un premier temps, nous observons les collocations de la forme *jeunes* dans le corpus global, puis dans chacun des sous-corpus. Nous effectuons donc la requête contexte appliquée à la forme « jeunes », puis nous appliquons la fonction thème aux résultats obtenus. Les tableaux résultant de ce processus sont les suivants (voir tableaux page suivante)².

Les résultats proposés pour les sous-corpus, sans contredire la première hypothèse de Mayaffre vont cependant la relativiser. En effet, même si le terme « jeunes » permet d'envisager un groupe « en bloc » et de faire l'économie de la distinction entre des populations fort différentes et parfois clivées, il ne recouvre pas la même réalité en fonction des situations d'énonciation.

¹ Brunet, E., « Un texte sacré peut-il changer ? Variations sur l'Évangile. »

² Nous n'avons pas jugé utile ici de travailler sur une base lemmatisée qui aurait permis de ne retenir que les substantifs et aurait délaissé des expressions comme : « les jeunes diplômés ». L'inconvénient de notre choix est en revanche de retenir des expressions comme : « les jeunes démocraties », qui ne sont pas pertinentes et risquent de fausser les résultats. Une fois encore, l'analyse ou du moins le contrôle manuel reste nécessaire. Ce relevé des collocations n'est donc qu'une étape informelle, il convient ensuite d'aller vérifier à la main les sauts qualitatifs qui permettent d'établir les collocations.

De l'utilité de la sémantique textuelle comme médium entre corpus et analyse

Jeunes corpus total	Jeunes TI	Jeunes EI
151.78 1069 1069 JEUNES	62.59 64 64 JEUNES	46.12 23 23 JEUNES
24.16 517 137 FORMATION	13.94 5 4 APPRENTISSAG	17.16 5 4 OFFICIERS
22.41 64 40 QUALIFICATIO	12.96 22 8 MILLIERS	11.04 3 2 RELIGION
21.69 42 31 FILLES	11.34 35 9 JEUNESSE	11.04 3 2 INTELLECTUEL
18.53 193 62 INSERTION	10.73 56 11 FORMATION	11.04 3 2 HUMANISTE
18.20 923 156 EMPLOI	10.38 5 3 UNIVERSITAIR	11.04 3 2 EXPORTER
17.46 3845 397 LEUR	10.38 5 3 ÉLITES	11.04 3 2 AUBE
15.96 14 13 DIPLÔMÉS	9.77 15 5 ÂGE	10.02 8 3 OPPOSITION
15.93 59 28 ALTERNANCE	9.65 10 4 PROFESSIONNE	9.51 4 2 MALADIES
15.90 98 37 APPRENTISSAG	8.68 7 3 MOBILITÉ	8.46 5 2 PAYSANS
14.71 24 16 GARÇONS	8.68 7 3 DITES	8.46 5 2 MULTIPLIER
13.19 262 56 PROFESSIONNE	8.07 8 3 RESPECTIVES	8.46 5 2 ÉCOLES
12.94 45 20 QUALIFIÉS	8.07 8 3 FORMER	8.46 5 2 ATTENTES
12.79 59 23 CONTRATS	8.02 14 4 CENTAINES	7.71 13 3 MESSAGE
12.21 423 71 EMPLOIS	7.81 53 8 SAINT	7.68 6 2 SAGE
12.16 59 22 RETRAITÉS	7.70 4 2 RAPPELONS	7.68 6 2 DESTIN
12.11 353 63 ÉCOLE	7.70 4 2 ENFANT	7.07 7 2 SIDA
11.86 620 89 EUX	7.44 16 4 ÉTUDIANTS	7.07 7 2 PORTÉE
11.63 15 10 SORTENT	6.83 5 2 RÉCIPROQUES	7.07 7 2 ALLÉ
11.54 157 37 JEUNESSE	6.83 5 2 INDIFFÉRENCE	7.07 7 2 AFGHANS
10.99 69 22 SCOLAIRE	6.83 5 2 DIEU	6.87 16 3 CULTURELLE
10.79 60 20 FORMATIONS	6.83 5 2 ARTISTIQUES	6.58 8 2 ÉTUDE
10.39 58 19 MOBILITÉ	6.83 5 2 ADULTES	6.58 8 2 COMMENCE
10.37 165 35 MÉTIERS	6.77 11 3 TRAVAILLEURS	6.58 8 2 CIVILES
10.36 43 16 INSTALLATION	6.73 40 6 PÈRE	5.82 10 2 CULTURES
10.20 2577 223 ILS	6.36 107 10 JAMAIS	5.73 22 3 CONTACTS
10.11 195 38 JEUNE	6.18 22 4 LANGUES	5.52 11 2 REMARQUER
10.04 164 34 MÉTIER	6.15 13 3 APPELÉ	5.25 12 2 MODERNISATIO
9.72 28422 1616 LES	5.91 703 31 AUX	5.19 26 3 ALLEMANDS
9.69 180 35 ENSEIGNEMENT	5.90 14 3 OFFRIR	5.07 27 3 MAIN
9.65 77 21 PROFESSIONNE	5.72 25 4 BIENVENUE	
9.42 137 29 ÂGE	5.67 7 2 PORTÉS	
9.32 18 9 INSÉRER	5.67 7 2 MANIFESTATIO	
9.31 26 11 BOULANGERS	5.67 7 2 JOURNÉES	
9.19 357 52 CHÔMAGE	5.67 7 2 ÉTABLISSEMEN	
9.18 126 27 QUARTIERS	5.67 7 2 BUENOS	
9.02 59 17 ÉDUCATIF	5.67 7 2 BRÉSILIENS	
8.51 321 46 EXPÉRIENCE	5.67 7 2 AIRES	
8.36 26 10 FERMÉS	5.45 16 3 ARRIÈRE	
8.27 227 36 AIDER	5.39 41 5 ENTIER	
8.25 48 14 UNIVERSITAIR	5.39 41 5 ÉDUCATION	
8.09 184 31 GÉNÉRATIONS	5.26 17 3 UNIVERSITÉS	
7.86 58 15 FORMER	5.26 8 2 JEUNESSES	
7.84 34 11 CLASSE	5.26 8 2 ENTHOUSIASME	
7.81 87 19 ÉLÈVES	5.00 31 4 CONNAISSANCE	
7.73 17348 990 QUI		
7.66 41 12 AÎNÉS		
7.51 55 14 ADULTES		
7.49 4780 322 SONT		
7.44 1592 133 NOTAMMENT		
7.42 93 19 ÉPANOUISSEME		
7.34 50 13 ACQUÉRIR		
7.30 350 44 TROUVER		
7.08 125 22 UNIVERSITÉ		
7.03 117 21 ÉTUDIANTS		
6.95 557 59 ENFANTS		
6.85 23962 1296 DES		
6.83 565 59 ENTREPRISE		
6.82 7950 483 PLUS		
6.76 56 13 ACTIFS		
6.60 31 9 APPELÉS		
6.59 241 32 CHANCES		
6.52 4488 292 AUX		
6.46 45 11 SCOLAIRES		
6.39 1570 123 VIE		
6.35 227 30 CHANCE		
6.28 1359 109 BEAUCOUP		
6.24 691 65 NOMBRE		
6.23 266 33 ÉDUCATION		
6.19 34 9 ISSUS		

Jeunes TL

100.27 615 615 JEUNES
 19.22 29 26 FILLES
 18.10 42 30 QUALIFICATIO
 17.63 293 88 FORMATION
 14.29 1999 267 LEUR
 12.67 470 92 EMPLOI
 12.23 15 12 GARÇONS
 10.81 121 35 INSERTION
 10.73 11 9 DIPLÔMÉS
 10.64 32 16 ALTERNANCE
 9.45 11 8 SORTENT
 9.30 305 56 EUX
 9.19 9 7 PÂTISSIERS
 8.99 37 15 MOBILITÉ
 8.75 26 12 BOULANGERS
 8.62 44 16 FORMATIONS
 8.57 113 28 MÉTIER
 8.54 27 12 CONTRATS
 8.37 67 20 APPRENTISSAG
 8.27 17 9 ENCOURAGE
 8.25 176 36 PROFESSIONNE
 8.17 14 8 FERMÉS
 8.00 137 30 MÉTIERS
 8.00 109 26 JEUNE
 7.79 31 12 LAURÉATS
 7.73 46 15 SCOLAIRE
 7.53 162 32 CHÔMAGE
 7.24 161 31 TROUVER
 7.23 1151 125 ILS
 7.19 30 11 ACQUÉRIR
 7.17 239 40 EMPLOIS
 7.00 22 9 QUALIFIÉS
 6.89 37 12 ÉDUCATIF
 6.83 188 33 ÉCOLE
 6.55 12609 911 LES
 6.51 12 6 COLLÈGES
 6.26 7136 542 QUI
 6.06 115 22 AIDER
 6.04 556 66 NOTAMMENT
 5.88 52 13 TROUVENT
 5.78 29 9 INSTALLATION
 5.71 114 21 CHEZ
 5.50 31 9 ADULTES
 5.46 173 27 SALUER
 5.38 214 31 NOMBREUX
 5.31 66 14 ÉPANOUISSEME
 5.29 150 24 EXPÉRIENCE
 5.24 33 9 FORMER
 5.19 22 7 ISSUS
 5.06 41 10 CARRIÈRE
 5.00 130 21 PROBLÈME

Jeunes EL

19.94 63 63 JEUNES
 6.51 20 13 EMPLOIS
 5.50 2 3 CROISADE
 5.49 34 16 EMPLOI
 5.08 15 9 QUARTIERS

Jeunes EN

49.29 141 141 JEUNES
 9.28 5 5 INSÉRER
 8.32 78 21 EMPLOI
 8.30 4 4 OPTIMISTES
 7.66 7 5 ANPE
 7.32 14 7 JEUNESSE
 7.19 3 3 MINITEL
 7.19 3 3 KILOMÈTRES
 6.72 16 7 INSERTION
 6.70 332 46 LEUR
 6.60 9 5 ALTERNANCE
 6.58 6 4 APPRENNENT
 6.44 13 6 ACTIFS
 6.00 7 4 VOLONTARIAT
 6.00 7 4 APPRENTISSAG
 5.53 8 4 ÉTABLISSEMENT
 5.51 12 5 TRAITER
 5.43 61 13 EUX
 5.35 5 3 TRADITIONS
 5.35 5 3 MINORITÉ
 5.35 5 3 DÉLINQUANTS
 5.35 5 3 COMMENÇANT
 5.22 49 11 FORMATION
 5.13 9 4 RETRAITÉS
 5.13 9 4 CONTRATS
 5.13 9 4 CITÉ

Comme nous le disions en introduction, la segmentation du corpus en fonction des situations d'énonciation et du mode d'interaction entraîne des différences de résultats dans chacun des sous-corpus. Les nébuleuses lexicales entourant le mot pôle nous orientent vers des réseaux différents. Quelques premières remarques peuvent être effectuées : le mot *emploi*, omniprésent dans les sous-corpus national et local est absent du sous-corpus international. On observe une certaine unité des sous-corpus National et Local face au corpus International.

On peut facilement imaginer que les jeunes dont parle le président dans ses discours et entretiens internationaux ne correspondent qu'à une certaine partie de la jeunesse. La jeunesse française sur laquelle le président communique à l'étranger étant mise en parallèle avec celle des autres pays (Brésiliens, Allemands, Japon) doit être une jeunesse brillante.

Outre les différences, on constate aussi des points communs. La forme *formation* est présente que ce soit au niveau international ou aux niveaux national et régional.

Une fois de plus, une grande prudence est de mise en ce qui concerne les signifiés de ces formes. Une relance des fonctions « contexte » et « thème » sur *formation* nous a permis de remarquer

deux emplois différents. La formation aux niveaux national et local est une formation concrète, appliquée et professionnalisante, au niveau international, elle est surtout universitaire. Se pose ici la question de l'unité du corpus Chirac. Une unité présupposée entraînerait la suspicion d'une duplicité de la part du président, alors qu'on peut considérer qu'il entre dans des pratiques différentes en fonction des situations de communication, et que ces pratiques ont leurs doxa propres. Il en va sans doute de même pour la forme *jeunes*. (cf. aussi note [8])

2.1.2. Les corrélats : réseaux de signifiés

Une fois ce premier traitement exploratoire réalisé, nous devons retourner vers le texte et essayer de définir quelles sont les isotopies ou paratopies¹ qui permettent d'établir les corrélations.

Rappel : « l'hypothèse qui fonde la transformation de la cooccurrence en corrélation est celle-ci : le contexte proche est structuré par des isotopies qui marquent l'appartenance à un même fond sémantique, ou des paratopies – qui marquent l'appartenance à la même forme sémantique²»

Une première remarque concerne le sémème de *jeunes*. Mayaffre obtient des résultats différents en termes de listes, mais qui vont dans le même sens que les nôtres :

« La consultation des contextes d'utilisation de « jeunes » montre que le terme apparaît presque exclusivement dans un environnement lexical favorable. »

Nous nous rangeons aussi dans une certaine mesure à sa dernière remarque :

« Il s'agit dans le corpus du Chirac d'un thème et d'un terme positifs d'un discours quasi subliminal. »

Ce terme paraît en effet ne lexicaliser qu'une valorisation /méliorative/, et une thymie euphorique. Nous ne sommes pas pour autant sûr qu'il lexicalise un thème. Si on relance la requête « contexte » sur la forme *élites* dans le discours international, toutes les occurrences sauf une se rapportent aux *jeunes*.

Observons maintenant ce que peuvent être les sèmes communs de ces cooccurrents. On a vu que les évaluations thymiques et axiologiques sont présentes dans les sémèmes de « *jeunes* » et « *élites* », ils le sont aussi dans ceux de « *apprentissage* », « *formation* », « *universitaire* » et « *mobilité* ». Autre sème commun à la plupart de ces lexèmes, le sème aspectuel /inchoatif/ que l'on retrouve dans « *jeunes* », « *apprentissage* », « *formation* », « *emploi* », « *insertion* », « *optimistes* », et enfin le sème mélioratif dans « *formation* » « *apprentissage* », « *insertion* ».

On observe donc un faisceau d'isotopies spécifiques qui constituent une partie du fond du discours (les isotopies /euphorie/, /inchoatif/ et /mélioratif/). Une fois relevées, ces isotopies demandent à être articulées, et on peut donc tenter d'établir une molécule sémique. Cependant, le problème qui se pose est celui des limites de cette molécule et de sa lexicalisation. Ici, comme nous venons de le voir, plusieurs termes peuvent prétendre au statut de lexicalisation synthétique de la molécule. Nous proposons donc une représentation structurée des parties de leurs sémèmes qui sont lexicalisables par les autres.

¹ Définitions : L'isotopie est l'effet de la récurrence syntagmatique d'un même sème dans différents sémèmes. L'identité des sèmes entraîne l'équivalence des sémèmes. La paratopie est constituée par la récurrence syntagmatique de sèmes différents, mais appartenant à la même molécule sémique.

² Rastier, F., 1996, « La sémantique des thèmes ou le voyage sentimental » 2.3.b.

Commençons par le sémème de « jeunes ». Il est ici considéré comme un processus inchoatif, valorisé de façon positive. On aurait donc une première ébauche de la molécule de la forme suivante :

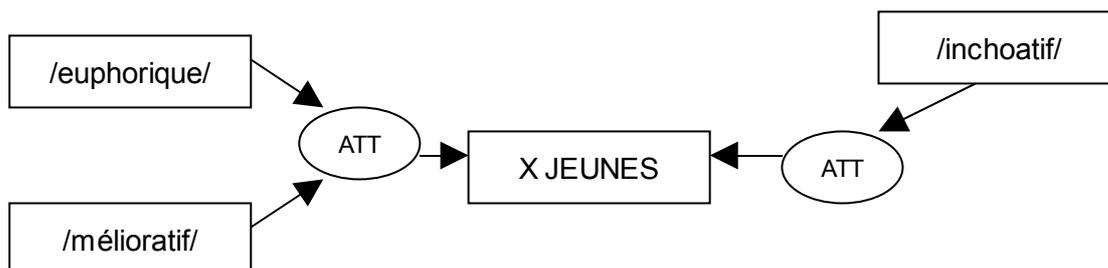


Figure 1 : Molécule sémique de "jeunes"

Une fois cette molécule établie, nous constatons que certains traits identifiés dans les autres termes ne sont pas présents. Nous devons donc les intégrer à cette représentation, quitte à la modifier. Le sème /mélioratif/ présent dans « formation » ou « insertion » doit par exemple être intégré. On peut faire évoluer cette molécule en remplaçant JEUNES par un sème /processus/ qui permet d'évacuer les sèmes génériques de JEUNES et d'abstraire la molécule d'un domaine particulier. Nous proposons la molécule suivante :

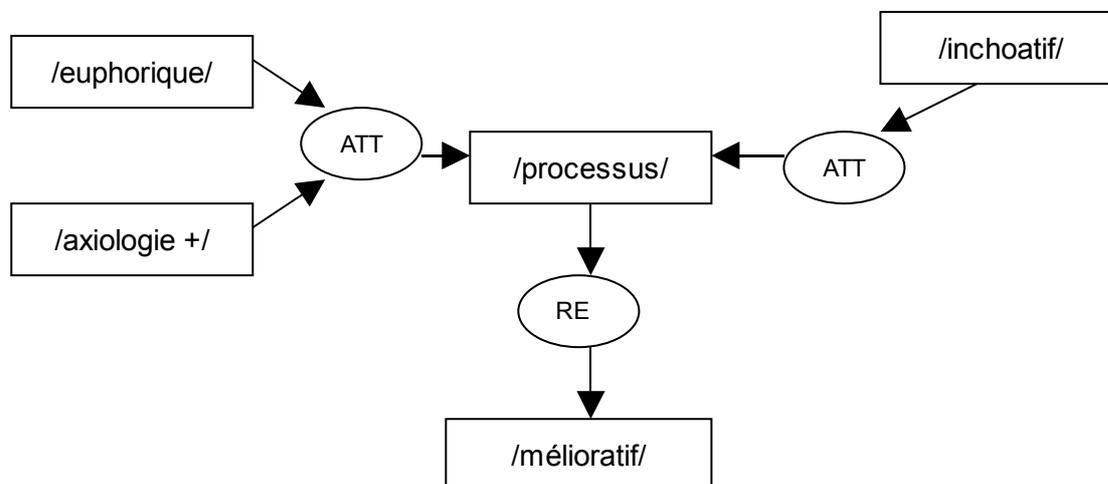


Figure 2 : Molécule globale sans lexicalisation

Cette représentation nous permet de renchérir sur la remarque de Mayaffre concernant le « terme positif d'un discours quasiment subliminal ». Le procédé employé est assez semblable à celui de l'image subliminale : la forme est bien présente mais quasiment jamais totalement visible (forme bruitée). Nous revenons en revanche sur la première partie de son affirmation « un thème ». Il semble qu'avec les outils de la sémantique interprétative, on puisse affirmer que la forme jeunes ne lexicalise pas un thème tel que nous venons de le décrire, mais simplement une partie de thème, plutôt axée sur la valorisation et l'aspect inchoatif. La forme élites quant à elle ne lexicalise que la partie évaluative de la molécule, la forme insertion semble être un bon candidat pour une lexicalisation synthétique.

Cette molécule semble lexicaliser un topos¹ quasi constitutif du discours politique et qui est celui qu'on pourrait nommer : « les lendemains qui chantent », présent dans presque tous les slogans de campagne (un nouvel élan...). Nous pouvons observer maintenant la lexicalisation de cette molécule dans les différents domaines du texte politique². Toujours en partant du postulat que le

¹ Définitions : (glossaire de *Arts et Sciences du Texte*)

- thème (spécifique) : molécule sémique relevant du palier mésosémantique.

- topos :

interne : enchaînement récurrent d'au moins deux molécules sémiques ou thèmes.

externe : axiome normatif sous-tendant une afférence socialisée.

- motif : structure textuelle complexe de rang macrosémantique. [...] Ainsi le motif est un syntagme narratif stéréotypé, partiellement instancié par les topos.

² Remarquons au passage l'utilité que pourrait avoir une base dans laquelle les sèmes seraient déjà associés aux formes, il suffirait alors de relancer une recherche sur la collocation de ces sèmes pour voir

contexte proche est structuré par les relations sémantiques, on peut étendre la recherche avec d'autres formes qui lexicalisent des parties de cette molécule. Cette étude permettrait en outre d'observer comment une même forme sémantique prend corps dans différents domaines et les différences en fonction des sous-corpus. En effet, même si nous avons défini une molécule suffisamment générale (donc spécifique) qui peut rendre compte de ce que lexicalise la forme jeunes dans le corpus, il convient maintenant de voir quels sont les actants qui l'investissent dans les sous-corpus et comment elle s'articule avec les autres formes.

2.2. Analyse macrosémantique et textualisation des formes sémantiques

La macrosémantique relevant de l'ordre du texte et de l'intertexte, est le palier auquel on met en relation les différentes formes sémantiques. Ce palier est donc le lieu de l'articulation dialectique des formes sémantiques, celui de la description des motifs. Nous revenons ici vers les textes des différents sous-corpus pour commencer à envisager les différents investissements que peut prendre une même lexicalisation d'une même forme, dans des contextes différents. Le cas de notre étude paraît très illustratif. La figure 1 présentait la molécule sémique correspondant à la forme *jeunes*. Cette molécule est lexicalisée dans tous les sous-corpus par la forme *jeunes*, et pourtant elle renvoie à des réalités différentes car elle est contextualisée de façon différente et parce qu'elle s'inscrit dans des schémas narratifs différents.

Notons simplement que dans les sous-corpus local et national, l'état s'engage à assurer une formation professionnalisante qui garantira l'insertion des jeunes (en difficulté ou au chômage) :

« [...]ouvrir le plus possible l'école et l'université sur l'entreprise. Il faut encourager, familiariser les garçons et les filles qui se forment avec le monde du travail. Ne pas se trouver ensuite avec une réponse faite à un jeune qui se présente et à qui l'on répond : " Mais quelle est votre expérience ? - Vous n'en avez pas, alors allez en acquérir et ensuite on vous donnera un emploi ¹". »

alors qu'au niveau international, les jeunes sont déjà formés dans une grande mesure, l'apprentissage ne concerne pas ici l'alternance en entreprise, mais celui des langues pratiquées dans l'union². Les jeunes du sous-corpus international correspondent à une population différente. Il s'agit là de l'avènement des élites formées dans les grandes capitales européennes selon les vieilles traditions du moyen-âge et de la renaissance.

« Il faut que nos jeunes, quel que soit leur parcours, deviennent européens en acquérant leur formation dans plusieurs villes universitaires de l'Union, renouant ainsi avec une tradition qui remonte au Moyen-âge.³ »

3. Conclusion

3.1. Remarques sur l'unité du corpus et l'importance d'une répartition problématisée des textes dans un corpus

On mesure dans cette étude l'importance de la répartition / organisation des textes dans le corpus. Les corpus, comme ensembles de textes, ne permettent pas toujours toutes les observations. Il semble aller de soi que l'étude des caractéristiques d'un discours ne peut faire l'économie des outils que propose la sémantique interprétative, notamment en ce qui concerne l'écologie des textes. Rappelons que d'après Rastier :

comment la molécule prend chair dans les différents domaines. Faute d'une telle base, nous pouvons toujours relancer les requêtes sur les termes polysémiques de notre liste, ou sur certains synonymes susceptibles d'être présents dans le corpus.

¹ Intervention de Monsieur Jacques CHIRAC Président de la République lors de la rencontre avec des jeunes des travaux publics à Villepinte VILLEPINTE - MARDI 22 AVRIL.

² Une rapide observation révèle que tous les emplois de *apprentissage* concernent, au niveau international, l'apprentissage des langues, alors qu'ils concernent quasi exclusivement, au niveau local, l'apprentissage pratique d'un métier, l'acquisition d'un savoir-faire, d'une expérience dans le monde du travail.

³ Discours de Monsieur Jacques CHIRAC président de la République à l'occasion de la réception des Ambassadeurs (Palais de l'Elysée) Palais de l'Elysée - Mercredi 26 août 1998.

« les textes sont configurés par les situations concrètes auxquelles ils participent ; en outre, par la médiation des genres et les discours, ils s'articulent aux pratiques sociales dont les situations d'énonciation et d'interprétation sont des occurrences.⁴ »

Mayaffre, dans son ouvrage, s'applique à étudier le discours présidentiel (comme l'indique le sous-titre : *Jacques Chirac (1995-2003) et le discours présidentiel sous la V^{ème} république*). Il contraste les textes de chacun des présidents sur ceux de tous les autres et pense ainsi pouvoir caractériser le discours de chaque auteur. Nous constatons cependant que le corpus Chirac ne présente pas une unité thématique, en ce qui concerne les référents des lexicalisations de formes sémantiques. Là où une interprétation sociologique ou politique pourrait tirer des conclusions sur les stratégies de communication d'un auteur, une analyse sémantique doit observer plus précisément les critères écologiques qui lient le texte à son entour, son genre, son discours, la pratique à laquelle il correspond, le champ dans lequel elle s'insère. Il semble donc que l'auteur soit un facteur faible de caractérisation des textes sans doute nécessaire, mais en rien suffisant, le style ou l'unité thématique ne devant être envisagés qu'au sein des genres. Une typologie des structures de corpus adaptée à une typologie des analyses reste sans doute un des chantiers de la sémantique voire de la linguistique de corpus.

3.2. Remarques sur l'analyse thématique et sur la lexicalisation des thèmes

Après avoir évoqué quelques recommandations concernant l'importance des conditions d'écologie fondamentales au sein de l'herméneutique, nous évoquons maintenant quelques remarques sur la construction des formes sémantiques. Nous reprenons ici les recommandations de Missire concernant les conditions herméneutiques nécessaires à la perception d'un thème latent. Cette perception est selon lui conditionnée par : « une lexicalisation synthétique dans l'entour proche où l'on identifie une lexicalisation analytique² ».

Dans nos textes, nous avons constaté qu'il y avait plusieurs lexicalisations synthétiques de la forme évoquée. Ce qui permet de repérer facilement le thème. Il n'en reste pas moins que le poids statistique d'une unité (en valeur absolue, ou relativement à d'autres corpus) ne garantit pas le statut de lexicalisation synthétique d'un thème. Comme nous avons pu le voir, la forme *jeunes* ne lexicalise qu'une partie de la forme.

3.3. Évocation du concept de para synonymie thématique

Revenons une dernière fois sur les propositions de Régis Missire dans son chapitre consacré à la morphosémantique textuelle. Il fait, à propos des liens des molécules sémiques, le constat suivant :

« Si, en tant que facteurs structurants des molécules sémiques, les liens différencient les éléments qu'ils connectent, leur disparition doit pouvoir se comprendre comme une indifférenciation des valeurs instanciant ces nœuds. Cette indifférenciation est nécessaire pour permettre à chacune des parties de valoir pour la forme intégrale. »

L'application de cette remarque conduit au concept présenté en titre. Les différentes lexicalisations analytiques se trouvent présenter des parties de formes évoquant la forme complète. Même si leurs sémèmes sont très différents, ils convoquent, au niveau thématique, la même molécule. Ces remarques restent encore trop générales, et de plus amples analyses sont nécessaires. Donnons à titre illustratif les deux exemples suivants :

- dans le sous-corpus National+Local, les termes *emploi, insertion, jeunes, formation, diplômés, sortent...* traduisent une même réalité ou un même topos qui est fondamental (pour ne pas dire fondateur) dans le discours politique et qui est celui de l'amélioration, de l'espoir (lexicalisable par exemple par : ça ira mieux demain, ou les lendemains qui chantent).
- dans le discours international, ce même topos est repris, la seule différence se situe au niveau du point de départ de l'amélioration. Là où le discours local propose l'amélioration d'une situation préoccupante, le discours international part d'une situation déjà excellente.

⁴ Rastier, F. « Eléments de théorie des genres » (Texte diffusé sur la liste fermée Sémantique des textes, 2001) www.revue-texto.net

² Missire, R., 2005, *Sémantique des textes et modèle morphosémantique de l'interprétation*, Thèse de doctorat, Université de Toulouse 2, p.180.

Ici, les termes para synonymes sont les suivants : *jeunes, apprentissage, formation, universitaire, élites, professionnel, mobilité, étudiants travailleurs...*

On touche ici à une idéologie sous-jacente au texte. On a pu observer ici que cette idée, même si elle était constante dans les différents sous-corpus, ne se traduisait pas au moyen des mêmes actants. Nous voudrions pour finir, nuancer la première remarque de Mayaffre concernant l'utilisation de la forme *jeunes* (permet de considérer une classe en bloc et d'éviter d'aborder certains clivages). La diversité des réalités recouvertes par cette dénomination générale dépend peut-être d'une volonté de dissimulation de certains clivages, d'un œcuménisme facile, mais elle dépend aussi sans doute des genres dans lesquels cette dénomination prend place.

Face à l'avènement d'une herméneutique numérique dans laquelle le texte interroge l'interprète, nous voudrions simplement rappeler que l'herméneutique est avant tout une question d'écologie des textes et que la tâche de l'interprète commence bien avant celle de la machine ne serait-ce que par le choix des textes à analyser. Rechercher l'unité d'un discours en ne différenciant pas les différents genres qu'il investit paraît illusoire – l'unité du discours ne pouvant s'appréhender qu'après avoir étudié les différents genres et fait la part de ce qui leur était propre.

Notre travail propose ici quelques recommandations méthodologiques préalables à l'étude des textes, il reste à une herméneutique critique (véritable épistémologie des sciences sociales) la tâche d'en faire le relevé complet et hiérarchisé.

BIBLIOGRAPHIE

- BEAUDOUIN, V. 2000. *Statistique textuelle : une approche empirique du sens à base d'analyse distributionnelle*, www.revue-texto.net
- BOMMIER-PINCEMIN, B. 2002. Sémantique interprétative et analyse de textes : que deviennent les sèmes ?, *Sémiotiques*, 17, pp. 71- 120.
- BRUNET, E. *Un texte sacré peut-il changer ? Variations sur l'Évangile*, <http://magyar-irodalom.elte.hu/colloquia/000601/brunet2.htm>
- MAYAFFRE, D. 2004. *Paroles de président. Jacques Chirac (1995-2003) et le discours présidentiel sous la Vème république*, Paris, Honoré Champion.
- MISSIRE, R. 2005. *Sémantique des textes et modèle morphosémantique de l'interprétation*, Thèse de doctorat, Université de Toulouse 2, en ligne sur : www.revue-texto.net
- RASTIER, F. 1987. *Sémantique interprétative*, Paris, PUF (édition de 1996).
- RASTIER F., 1996, La sémantique des thèmes ou le voyage sentimental, *Texto !* www.revue-texto.net
- RASTIER, F. 2001. *Arts et sciences du texte*, Paris, PUF.
- RASTIER, F. 2003. Formes sémantiques et textualité, in *Unité(s) du texte*, Cahiers du Crisco 12, pp. 99-114.
- RASTIER, F. 2005. Mésosémantique et syntaxe, *Texto !* www.revue-texto.net