

RIGIDITÉS DISCURSIVES ET FLOU SÉMANTIQUE

Carmen Pineira-Tresmontant

Université de Paris VII et INALF Saint-Cloud, CNRS

CPineiraT a ol com

Résumé de l'article / Abstract

RIGIDITÉS DISCURSIVES ET FLOU SÉMANTIQUE

A travers une description automatisée de textes syndicaux, on tente de cerner des unités complexes appelées « lexies ». Les problèmes rencontrés pour reconnaître et relever automatiquement de telles unités sont mis en évidence. En se limitant à la catégorie des lexies grammaticales, on évalue leur importance en proportion de la surface du texte. Une analyse détaillée d'une catégorie de lexies est proposée: la prédétermination complexe. La liaison entre les caractéristiques de cette prédétermination et le sémantisme du discours est étudié en rapport avec le concept de « langue de bois » qui met l'accent sur le flou sémantique.

Mots clés: lexie, langue de bois, textes syndicaux.

DISCOURSE RIGIDITY AND SEMANTIC VAGUENESS

Through a computerized description of trade-union texts, an attempt has been made to automatically isolate and record complex units called "lexies". Such an endeavour highlights the difficulties met when identifying and making a list of such units. By restricting our scope to the category of syntactical lexies, we evaluate their importance in relation with the size of the text. A detailed analysis of a category of lexies is proposed, i.e. complex predetermination. The link between the characteristics of predetermination and discourse semanticism is studied in relation with the concept of "langue de bois", which stresses semantic vagueness.

Key-words: "lexie", "langue de bois", trade-union texts.

LA NOTION DE « LEXIE »

Pour qui cherche à décrire et à analyser un texte, la lexicométrie propose, pour le traitement automatique du langage écrit, une série d'instruments actuellement performants. Ces divers instruments ont en commun d'être fondés sur la forme graphique minimale chaîne de caractères comprise entre deux séparateurs -, *i.e.* sur les « mots », et permettent l'étude statistique des formes - ou vocabulaire - d'un ou plusieurs textes. Les index obtenus résultent d'une segmentation du texte en unités graphiques, isolées ou en contexte ; les index de « segments répétés » (ISR) repèrent les groupes récurrents de deux ou plusieurs unités; les programmes de cooccurrences, quant à eux, offrent la possibilité de reconnaître automatiquement les « couples » répétés d'unités simples, que celles-ci soient ou non séparées par une ou plusieurs autres formes. Ces diverses représentations permettent d'obtenir une mesure quantitative rigoureuse des unités lexicales et de leur localisation dans un texte ; elles ne

prennent cependant pas en compte toutes les relations syntagmatiques, privilégiées ou non, qu'entretiennent entre elles certaines formes, puisqu'elles ne mettent en évidence que le seul rapport statistique.

L'observation montre que, loin d'être constitués de formes isolées, les textes sont partiellement composés d'unités syntagmatiques regroupées entre elles (voir Annexe). Bien que les ISR ne rendent compte que des « suites de formes graphiques non séparées par un caractère délimiteur de séquence qui apparaissent plus d'une fois dans un corpus de textes »¹, et éliminent par conséquent les séquences hapax, leurs résultats soulignent l'importance quantitative des groupes d'unités récurrents dans les textes: on se trouve bien devant un premier indice de l'existence d'un phénomène linguistique qui est mal ou rarement mesuré. Ces « blocs » sont des séquences d'éléments qui forment parfois une « unité » nouvelle au signifié propre dont le sens ne peut être reconstruit à partir des composantes isolées. Pourtant, bon nombre de ces unités complexes ne figurent pas en tant que telles dans les dictionnaires. Plus ou moins fortement lexicalisées et formant une unité syntagmatique, elles seront désignées ici sous l'appellation de « lexies ». Relever et analyser ces lexies, qu'elles soient répétées ou non dans un texte, offre une voie de recherches à part entière au linguiste.

Rappelons de façon cursive que l'importance prise par le phénomène linguistique de la lexie a été signalée et analysée depuis fort longtemps. Des études « fondatrices », telles que celles de C. Bally, H. Frei, E. Benveniste, A. Martinet, I. Fónagy, B. Pottier, ou plus récentes comme celles de M. Gross ou de N. Catach ou, dans une autre optique, les études de R. Barthes illustrent, par la diversité et même les divergences des critères retenus, la variété et la complexité des lexies².

Une tentative de reconnaissance automatique de ces lexies est menée au Laboratoire de Saint-Cloud depuis 1985³ sur des Résolutions syndicales: la Résolution générale du congrès de la CGT de 1978 et les Résolutions du congrès de Force ouvrière de 1984. Dans une première étape, le travail a porté sur les lexies d'ordre fonctionnel (prépositionnel, adverbial, conjonctif) et sur les prédéterminants nominaux complexes⁴, la reconnaissance de telles unités complexes devrait faciliter le repérage des autres lexies (nominales et verbales) et permettre d'aborder l'organisation syntagmatique du texte.

Précisons immédiatement que le repérage automatisé de toutes les lexies d'un texte est difficile car elles peuvent comporter en surface une part de variation plus ou moins importante. Leur rigidité en surface et leur régularité ne peuvent pas toujours constituer des critères pertinents ou suffisants pour une reconnaissance exhaustive. Aussi convient-il d'élaborer des procédures spécifiques de repérage et d'analyse de ces séquences.

Nous rejoignons là les propos de B. Pottier⁵ : « Ce sont les mécanismes linguistiques de la construction de l'énoncé qui sont envisagés ici. La construction d'un énoncé est une progression par tranches syntagmatiques, lesquelles peuvent être amplifiées intérieurement à volonté ». « Tout énoncé complexe est donc simplement l'exploitation des combinaisons fondamentales en nombre très réduit. La langue répète sa structure fondamentale dans les procédés d'élargissement. »

L'objet du présent article est d'exposer quelques résultats obtenus grâce aux instruments déjà mis au point et de proposer des hypothèses d'interprétation.

¹ A. Salem, *Pratique des segments répétés. Essai de statistique textuelle*, Paris, Klincksieck, 1987, p. 21.

² Outre les travaux des auteurs mentionnés ci-dessus, on pourra se reporter au résumé proposé par G. Misri : « Approches du figement linguistique : critères et tendances », *La Linguistique* 23 (2), 1987, p. 71-85. D'autre part, P. Fiala, B. Hahert soulignent certains éléments de choix retenus dans ces travaux : « Un modèle évolutif de la phraséologie en discours », septembre 1987, à paraître.

³ Au sein de l'équipe GAUS (Groupe d'analyse des unités syntagmatiques) — URL 3. Lexicométrie et textes politiques, EN9~Saint-Cloud-Fontenay — CNRS.

⁴ P. Fiala, W. Habert, P. Lafon, C. Pineira, « Des mots aux syntagmes. Figements et variations dans la Résolution générale de la CGT de 1978 ». *Mots*, 14, mars 1987, p. 47-88.

⁵ B. Pottier, *Introduction à l'étude des structures grammaticales fondamentales*, Nancy, Université de Nancy, 1969, § 7.2.

Variations et rigidités dans les lexies

L'équipe GAUS a tenté une première expérience qui a consisté à recenser à partir des Résolutions syndicales CGT et FO certaines constructions grammaticales relativement rigides, puis à élaborer un dictionnaire de locutions grammaticales, qui prenne en compte leurs variations potentielles, attestées dans ces textes ou dans d'autres ; un tel dictionnaire n'est pas une simple liste fermée de locutions grammaticales, il offre la possibilité de procéder à des repérages automatiques évolutifs.

Les relevés effectués à partir des concordanciers et des ISR nous ont montré la complexité du problème: en effet, si certaines locutions grammaticales sont facilement repérables (*en effet, en bref, à défaut, etc.*) parce qu'elles constituent des séquences invariables, d'autres le sont beaucoup moins parce qu'elles sont semi-figées (*dans ce domaine, dans ces domaines, dans tous les domaines, en ce domaine, etc.*). Une unité complexe peut en effet présenter des variations de surface, dont certaines peuvent même entraîner un glissement catégoriel. Il en est ainsi, par exemple, des expressions suivantes :

<i>au niveau de</i>	locution prépositionnelle
<i>au niveau + adjectif</i>	syntagme adverbial

Alors que le repérage des mots composés, tels que *socio-économique, antisociale*, ne pose pas de problème car la liaison est de type morphologique (présence d'une apostrophe ou d'un trait d'union, ou d'une seule unité graphique), les relations d'ordre syntaxique sont au contraire difficiles à cerner avec rigueur du fait de l'interposition possible de certains éléments, ou des différentes sortes de flexions possibles, comme le montrent les exemples suivants :

avant que, avant même que (présence d'un élément optionnel)
à l'égard de, à son égard (transformation par possessivation)

On constate en outre que certaines de ces lexies peuvent commuter entre elles sur l'axe paradigmatique (substitution), et se regrouper ainsi en classes d'unités définies par un certain environnement grammatical et/ou syntaxique, obéissant à un schéma, ou « patron », syntaxique. On obtient par exemple pour le « patron » syntaxique:

Prép + Dét + N + Adj

au niveau régional
au niveau politique
au niveau national
au niveau international
au niveau européen

c'est-à-dire *au niveau* suivi d'un groupe adjectival variable, ou

sur le plan régional
au plan régional
au niveau régional
dans le domaine régional

c'est-à-dire un syntagme prépositionnel variable suivi d'un adjectif fixe, *régional*.

L'hétérogénéité de ces expressions complexes souligne la difficulté de réunir des critères satisfaisants de définition d'une lexie. Pour certaines, la commutation porte sur un seul élément libre, pour d'autres, elle s'effectue sur plusieurs éléments. L'utilisation des ouvrages usuels (Grévisse, etc.) apparaît insuffisante dans cette optique car les listes proposées sont incomplètes et nécessitent un élargissement aux variations potentielles qui sont très fréquentes.

C'est cette richesse potentielle de la combinatoire qui a justifié le choix d'un repérage contextuel et évolutif, *i.e.* spécifique au texte, et le refus d'une liste délimitée a priori. La reconnaissance des lexies en est rendue plus difficile; pourtant, on est conduit à mettre en évidence l'existence de nouvelles séquences complexes en se fondant sur une observation à la fois syntaxique, sémantique et distributionnelle, dans laquelle l'intuition et la compétence du linguiste jouent un rôle important : les lexies se comportent en définitive comme des unités simples et dans un contexte déterminé, en partageant le même fonctionnement syntaxique. Les propositions suivantes illustrent la complexité des critères à retenir dans la définition de ces unités et la mesure de leur degré de figement⁶.

(1) *Face* à l'agression gouvernementale et patronale contre...

(2) Nous devons donc *faire face* à cette offensive...

Dans le premier exemple, la séquence graphique *face* à forme une locution de type prépositionnel, alors que dans le second elle s'intègre dans une construction plus étendue pour former une locution verbale.

Cette approche des textes par la prise en compte et le repérage de certaines de ses lexies nous permet de questionner la grammaire traditionnelle, fondée surtout sur la forme graphique minimale et sur une catégorisation grammaticale en parties du discours. En effet, si l'on veut attribuer une catégorie grammaticale à chaque unité syntagmatique, on pose le problème de sa délimitation et de sa définition, on est bien obligé par conséquent de se référer à une théorie des parties du discours qui renvoie au lexique (description des unités simples). Or les différentes grammaires divergent sur les listes des parties du discours : on est donc en présence d'une catégorisation hétérogène⁷ ; en outre, une telle optique ne permet de prendre en compte ni les unités syntagmatiques ni leurs variations internes.

Cette critique adressée à certaines grammaires traditionnelles est renforcée par le fait que chaque élément constitutif de ces lexies perd son identité grammaticale propre pour ne plus être qu'un rouage d'une composition. Il y a donc modification des éléments de base qui, dans les exemples suivants, ne constituent plus un groupe nominal (*les domaines, l'ensemble*) mais, respectivement, une locution adverbiale et un prédéterminant nominal complexe:

<i>dans tous les domaines</i>	locution adverbiale
<i>l'ensemble de</i>	prédéterminant nominal

L'insertion de ces éléments dans des lexies modifie par conséquent certaines de leurs caractéristiques : le plus souvent, ils perdent l'identité grammaticale et sémantique que leur conférait leur emploi libre, au profit d'une nouvelle catégorie grammaticale et d'une nouvelle signification propre à la lexie. Ce cas de figure est bien illustré par les locutions grammaticales citées plus haut dans lesquelles chaque élément nominal est plus ou moins vidé de son contenu sémantique initial et par conséquent de sa fonction syntaxique, et dont le rôle est surtout d'établir certains rapports à l'intérieur de la phrase. Une telle prise en compte nécessite une théorie des fonctions énonçant les règles de combinaison de ces unités dans le cadre de la phrase: celle-ci n'est plus une simple juxtaposition de lexèmes mais une structure dans laquelle les fonctions se définissent entre des groupes de mots, ou syntagmes.

La démarche consistant à concevoir des procédures appropriées à la construction automatisée de formes complexes repose sur un raisonnement peu habituel. Pourtant, l'évolution du lexique s'opère largement par composition et dérivation de formes préexistantes pour répondre au besoin sans cesse renouvelé de désigner le monde, par ailleurs, chaque texte, chaque discours présente sa propre combinatoire lexicale. Le repérage des lexies dans les corpus et dans les textes respecte le principe dynamique du langage dans la mesure où il reste

⁶ D. Gaatone « Locutions prépositives et groupes prépositionnels: observations sur la syntaxe de certains groupes prépositionnels ». *Linguistics*, 167, 1976 ; « La locution ou le poids de la diachronie dans la synchronie », *Le moyen français*, 14-15, 1986.

⁷ M.-N. Garyrieur, *De la grammaire à la linguistique. L'étude de la phrase*. Paris, A. Colin, 1985.

sensible aux figements/défigements, le langage étant pris dans sa spécificité textuelle ; il s'agit donc d'une démarche inductive et non a priori.

Prendre les textes constitutifs d'un corpus comme référence exclut aussi l'élaboration d'un dictionnaire général qui prétendrait recenser l'ensemble des lexies ; selon le texte envisagé, il sera possible d'utiliser des dictionnaires spécifiques (médical, technique, informatique, économique, syndical ...) tout en connaissant les limites de tels instruments : l'exhaustivité est un idéal jamais atteint. Par exemple, si l'outillage grammatical paraît être d'une utilisation plus générale, les lexies nominales et verbales semblent, elles, relever d'un genre discursif, voire du texte.

IMPORTANCE DES LEXIES DANS LES RÉOLUTIONS SYNDICALES

Quelle est l'importance des lexies dans nos textes ? On s'attachera ici à décrire certains éléments constitutifs du langage syndical en montrant l'importance relative des lexies relevées dans les Résolutions du corpus. Il est intéressant d'autre part d'observer la répartition et la fréquence de ces lexies dans les diverses catégories grammaticales.

Cette étape essentiellement descriptive peut conduire à une réflexion d'ordre linguistique : à partir de la comparaison des deux textes, peut-on esquisser les bases d'une typologie des discours syndicaux ? Existe-t-il, par exemple, des préférences des Confédérations pour telle ou telle forme de lexies ? Quels sont les schémas de construction préférentiels ? Comment telle Confédération détermine-t-elle les actants sociaux dans ses textes ?

Dans cette optique, le regroupement des unités lexisées sous forme d'index a été fait automatiquement après catégorisation des unités minimales de nos textes et construction des locutions grammaticales à partir d'un dictionnaire⁸. Ont été considérées les catégories grammaticales suivantes: locutions adverbiales (Loc-adv), conjonctives (Loc-conj), prépositionnelles (Loc-prép) et les prédéterminants complexes (Det).

Le tableau 1 ci-dessous présente les données brutes des nouveaux index intégrant ces quatre types de lexies:

Tableau 1. Données brutes (occurrences)

Catégories	Loc-conj	Loc-prép	Loc-adv	Det	Total
CGT					
Occ T = 22714					
Occ	236	1840	923	3530	6529
Occ S	173	1605	551	3283	5612
Occ X	63	235	372	247	917
FO					
Occ T = 20141					
Occ	314	1167	674	2906	5511
Occ S	231	1372	439	2778	4820
Occ X	83	245	235	128	691

⁸ La catégorisation automatique des unités minimales a été effectuée en utilisant le module CBSF (Catégorisation de base syntaxique du français) du logiciel DEREDEC (P. Plante — Montréal) Cf. *Mots*, 6, mars 1983, p. 101-134.). La lexisation est opérée par un programme écrit en IQ-LISP Cf. *Mots*, 14, mars 1987, p. 51-60.

Sur le tableau 1, on observe le nombre d'occurrences par catégorie grammaticale de toutes les formes simples (Occ S) ou complexes (Occ X) concernées, pour chaque Confédération [CGT, Confédération Générale des Travailleurs et FO, Force Ouvrière]. Il s'agit bien d'une représentation des unités et locutions grammaticales et non des unités graphiques.

Pour comparer l'importance respective des catégories grammaticales dans chaque résolution en tenant compte de la distinction forme simple/forme complexe, on utilise le rapport suivant:

$$R_C = \frac{\text{Nombre d'occurrences des lexies d'une catégorie}}{\text{Nombre total d'occurrences des formes simples du texte (Occ T)}} \times 100$$

soit

$$R_C = \frac{100 f(C)}{T}$$

où f désigne la fréquence, C une catégorie de lexies et T la longueur du texte.

Ainsi, pour la CGT, on obtient pour la catégorie Loc-conj Complexe: $63 / 22714 = 0,28\%$.

On compare donc des formes si ur des unités de statut différent et posent la question simples. Les comptages portent donc s de l'unité de référence. Ce rapport n'est alors qu'une approximation grossière car les lexies peuvent se composer d'un nombre variable de formes simples. Le résultat chiffré est donc trompeur sur le poids lexical que représentent en surface les unités complexes. Pour plus de rigueur, le rapport devrait être défini en calculant la « surface couverte » par ce type de phénomène⁹, comme suit:

Nombre d'occurrences de formes simples insérées dans une lexie, par catégorie grammaticale /
Nombre total d'occurrences des formes simples du texte

soit, pour le même exemple: $168 / 22\ 714 = 0,74\ %$ dans lequel le chiffre 168 correspond au nombre de formes simples insérées dans des lexies du type Loc-conj. Ainsi, pour la locution conjonctive *tant que* présente 5 fois, la surface recouverte est de 10 occurrences.

Le tableau 2 suivant est construit en prenant en compte cette critique: on y distingue les occurrences simples, i.e. les occurrences des formes employées isolément, des occurrences des formes simples insérées dans des lexies. Ainsi, numérateur et dénominateur confrontent des données de même espèce. Le tableau 3 fournit ces résultats en pourcentages.

Tableau 2. Données reconstruites (exclusivement sur occurrences de formes simples)

Catégories	Loc-conj	Loc-prép	Loc-adv	Det	Total
CGT					
Occ T = 22714					
Occ	341	2339	1660	3903	8243
Occ S	173	1605	551	3283	5612
Occ X	168	734	1109	620	2631
FO					
Occ T = 20141					
Occ	445	2151	1075	3125	6796

⁹ A. Salem, *Pratique des segments répétés*, op. cit., p. 116.

Occ S	231	1372	439	2778	4820
Occ X	214	779	636	347	1976

Tableau 3. Données reconstruites (pourcentages)

Catégories	Loc-conj	Loc-prép	Loc-adv	Det	Total
CGT					
Occ T = 100 %					
Occ	1,5	10,3	7,31	17,18	36,29
Occ S	0,76	7,07	2,43	14,45	24,71
Occ X	0,74	3,23	4,88	2,73	11,58
FO					
Occ T = 100 %					
Occ	2,21	10,68	5,34	15,52	33,74
Occ S	1,15	6,81	2,18	13,80	23,93
Occ X	1,06	3,87	3,16	1,72	9,81

Le poids de l'outillage grammatical total, figuré dans ce tableau, est supérieur à 36 % des occurrences de la CGT, et atteint près des 34 % des occurrences de FO. Ces occurrences représentent soit des formes grammaticales simples, soit des combinaisons de formes simples constituant des lexies. Ces chiffres ne prennent pas en compte la coordination, simple ou complexe, qui n'a pas encore été intégrée dans notre dictionnaire; cela signifie que l'outillage grammatical total intègre une part plus importante encore des occurrences de ces deux textes.

Ces données permettent de donner une idée du poids relatif des lexies grammaticales dans les deux textes observés. Pour ces mêmes catégories, l'importance des lexies par rapport aux formes simples est très grande puisque les lexies grammaticales relevées mobilisent 12 % des occurrences de la CGT et 10 % des occurrences de FO. Ce poids peut être comparé à celui de la part d'outillage grammatical simple recensé ici, soit respectivement 25 % et 24 %.

Rappelons que la représentation proposée dépend des unités présentes dans le dictionnaire utilisé: l'importance attribuée au phénomène de lexisation est donc liée au contenu de ce dictionnaire. Cependant, quand les lexies fonctionnelles mobilisent environ 10 % des occurrences d'un texte, il est clair qu'on ne peut ignorer leur poids spécifique dans les analyses automatisées. Cette remarque paraît d'autant plus forte que le poids des locutions considérées ci-dessus devrait être accru par celui des locutions grammaticales manquantes, des locutions nominales et des locutions verbales pour disposer d'une mesure quantitative du phénomène de la locutionnalité dans les Résolutions.

On peut alors classer ce poids des lexies par catégorie grammaticale selon un ordre hiérarchique:

Tableau 4. Classement hiérarchique par catégorie grammaticale

CGT			FO		
Forme simple	Forme complexe	Total	Forme simple	Forme complexe	Total
Det	Loc-adv	Det	Det	Loc-prép	
Loc-prép	Loc-prép	Loc-prép	Loc-prép	Loc-adv	Loc-prép
Loc-adv	Det	Loc-adv	Loc-adv	Det	Loc-adv
Loc-conj	Loc-conj	Loc-conj	Loc-conj	Loc-conj	Loc-conj

A la lecture de ce tableau 4, on peut remarquer la grande stabilité de l'appareillage grammatical dans les deux textes, on constate une unique différence parmi les locutions: la catégorie Loc-adv représente près de 5 % du total à la CGT contre 3 % chez FO; la catégorie Loc-prép est d'importance équivalente dans les résolutions

des deux Confédérations (3,2 % et 3,8 % respectivement). En l'absence de mesure des lexies nominales, il est difficile de proposer une interprétation globale de cette comparaison quantitative ; on peut en revanche proposer une description plus fine par catégorie grammaticale: sur un plan analytique en effet, on ne peut ignorer l'importance fonctionnelle propre de ces locutions grammaticales qui introduisent les constituants de l'énoncé et définissent leurs interrelations. La prédétermination nominale complexe analysée ci-après en constitue une illustration.

UN EXEMPLE: LA PRÉDÉTERMINATION COMPLEXE

La prédétermination joue en effet un rôle fondamental dans la production du sens parce qu'elle actualise le substantif, autonome sémantiquement mais non syntaxiquement. En outre, cette catégorie grammaticale comprend des unités qui peuvent être complexes.

Par exemple, les formes *ensemble* et *tout, tous, toute, toutes* apparaissent dans les index avec les fréquences les plus élevées parmi les quantificateurs complexes recensés: F = 10 et 58 pour FO, et F = 48 et 90 pour CGT respectivement; si l'on prend en compte leurs différentes compositions, on observe la constitution des lexies suivantes :

Lexie Déterminant + ensemble + préposition

<i>CGT</i>	<i>fréquence 48</i>	<i>FO</i>	<i>fréquence 10</i>
<i>l'ensemble de ces pays</i>	1	<i>l'ensemble de la revendication salariale</i>	1
<i>l'ensemble de cette situation</i>	1	<i>l'ensemble du territoire national</i>	1
<i>l'ensemble de l'activité syndicale</i>	1	<i>l'ensemble des hiérarchies ouvrières</i>	1
<i>l'ensemble de la société</i>	1	<i>l'ensemble des minima conventionnels</i>	1
<i>l'ensemble de leur activité</i>	1	<i>l'ensemble des salariés</i>	2
<i>l'ensemble de nos publications</i>	1	<i>l'ensemble des secteurs concernés</i>	1
<i>l'ensemble de notre mouvement syndical</i>	1	<i>l'ensemble des secteurs d'activité</i>	1
<i>l'ensemble de ses organisations</i>	1	<i>l'ensemble des travailleurs</i>	1
<i>l'ensemble des catégories</i>	1	<i>un ensemble de mesures</i>	1
<i>l'ensemble des forces + Adj.</i>	2		
<i>l'ensemble des mass-media</i>	1		
<i>l'ensemble des militants de la CGT</i>	2		
<i>l'ensemble des organisations</i>	3		
<i>l'ensemble des pays</i>	2		
<i>l'ensemble des rémunérations</i>	2		
<i>l'ensemble des responsabilités de la CGT</i>	1		
<i>l'ensemble des salariés</i>	3		
<i>l'ensemble des travailleurs</i>	12		
<i>l'ensemble du mouvement</i>	2		
<i>l'ensemble du secteur</i>	1		
<i>l'ensemble du système</i>	2		
<i>un ensemble d'initiatives</i>	1		
<i>un ensemble de conceptions communes</i>	1		
<i>un ensemble de formules figées</i>	1		
<i>cet ensemble d'obstacles</i>	1		
<i>cet ensemble de réalités</i>	1		
<i>aucun ensemble de travailleurs</i>	1		

Lexie: TOUT (tout, tous, toute, toutes) + Déterminant + N

CGT	fréquence 58	FO	fréquence 90
toutes ses + N	2	toutes ses + N	1
toutes les + N	21	toutes les + N	29
toutes ces + N	1	toutes les autres + N	2
toute leur + N	1	toutes ces + N	2
tous ses + N	1	toute une + N	1
tous leurs + N	1	toute son + N	2
tous les + N	31	toute sa + N	1
		toute leur + N	1
		toute la + N	8
		tout leur + N	1
		tout le + N	1
		tous ses + N	1
		tous les + N	35
		tous les autres + N	1
		tous ces + N	2
		toutes sortes de + N	1

Le poids de cette classe est considérable, tant pour les unités simples que complexes. Constaté l'importance quantitative de la prédétermination nominale ne suffit pas pour proposer une interprétation du phénomène, le constat doit être complété à différents titres.

En effet, la prédétermination complexe apparaît comme une catégorie remarquable puisqu'un substantif ne peut exister sans prédéterminant; il est vrai que de nombreuses constructions montrent des substantifs ou des groupes nominaux non précédés, en surface, d'un déterminant essentiel: on sait qu'il convient alors de considérer l'existence du prédéterminant zéro, le substantif nu renvoyant au concept¹⁰.

D'autre part, la prédétermination constitue un instrument privilégié d'un genre discursif particulier, elle permet de savoir comment un locuteur tel que la CGT ou FO va spécifier les substantifs-clés de son discours. Enfin, cette catégorie introduit à une analyse du contexte de la phrase, et même du texte tout entier.

L'analyse syntaxique des prédéterminants complexes montre que ces derniers peuvent être confondus avec un groupe nominal lorsqu'il y a identité graphique: ainsi en va-t-il des expressions *l'ensemble de + N*, *la masse de + N*, etc., qui sont, suivant le contexte, prédéterminant ou groupe nominal. Cette ambivalence est évidemment de première importance pour la constitution d'un analyseur automatique puisqu'elle interdit le recours exclusif à la reconnaissance graphique. La construction des prédéterminants complexes, en fonction du type de discours, peut permettre au contraire de lever ces ambiguïtés au cas par cas. Ainsi, dans les Résolutions syndicales étudiées, *l'ensemble de* et *la masse de* constituent toujours des prédéterminants complexes.

Ces résultats rejoignent les propos de J.-P. Sueur¹¹: «Les études statistiques portant sur le vocabulaire des discours doivent nécessairement être complétés par l'étude des structures syntaxiques et des cadres énonciatifs dans lesquels les formes lexicales et grammaticales apparaissent ».

L'élaboration d'un analyseur automatique nécessite donc la reconnaissance des différentes classes syntaxiques qui ordonnent la structure de la phrase. Or la définition de ces classes ne repose pas sur les propriétés sémantiques de leurs éléments constitutifs, mais semble plutôt répondre à une combinatoire repérable. Par exemple, pour les prédéterminants complexes, la combinatoire peut s'organiser autour de la présence d'un

¹⁰ J.-C. Anscombre, « L'article zéro en français: un imparfait du substantif? », *Langue française*, 72, décembre 1986.

¹¹ J.-P. Sueur, «Syntaxe, lexique et énonciation dans quelques Résolutions syndicales,» *Actes du 2^e colloque de Lexicologie politique*, Paris, Klincksieck, 1982, p. 473-492.

déterminant essentiel, de la façon suivante:

<i>Det4</i>	<i>Det3</i>	<i>Det2</i>	<i>Det1</i>	<i>N</i>
	le		dernier	N
	ces		derniers	N
	ces	trente	dernières	N
	des		différentes	N
	un		même	N
	le		même	N
	les		autres	N
	une		seule	N
	un		seul	N
	la		seule	N
	un		tel	N
toutes	ces			N
	les			N
	une		certaine	N
<i>extensifs</i>	<i>articles</i> <i>démonstratifs</i> <i>possesifs</i>	<i>numéraux</i>	<i>adjectif</i>	<i>Nom</i>

On note dans ces tableaux la présence obligatoire d'un déterminant essentiel (Det 3 dans le tableau 5, Det 7 dans le tableau 6). Ce type de déterminant peut se combiner avec les autres constituants qui, eux, sont facultatifs. Ainsi, dans le cas du Det du type Det + N, les différentes séquences possibles sont:

Det4 + Det3 + Det2 + Det1 + N

Det3 + Det2 + Det1 + N

Det4 + Det3 + Det1 + N, etc.

Tableau 6. Déterminants complexes avec déterminant essentiel rentrant dans la construction d'un SN du type Det + de N

Extension du SN par complémentation du substantif

<i>Det7</i>	<i>Det6</i>	<i>Det5</i>	<i>Det4</i>	<i>Det3</i>	<i>Det2</i>	<i>Det1</i>	<i>de N</i>
un			ensemble				de N
le			maximum				de N
des			milliers				de N
un		grand	nombre				de N
un		certain	nombre				de N
un	trop	grand	nombre				de N
un	plus	grand	nombre				de N
un			nombre	toujours	plus	grand	de N
la			plupart				de N
une			sorte				de N

La combinatoire schématisée ci-dessus montre la richesse des variations possibles autour de quelques modèles et, par là, la difficulté de clore des listes de prédéterminants complexes. C'est ainsi que les Résolutions syndicales comportent des expressions renvoyant au modèle Det + de N telles que *un certain nombre de*, *un plus grand nombre de* mais non *une kyrielle de*, *une myriade de*, *une foule de*. Il est à noter que ce modèle ne peut être utilisé de façon automatique sans procéder préalablement à une série de tests permettant de distinguer

les déterminants complexes des syntagmes nominaux. En effet, ce modèle offre la possibilité de repérer les syntagmes nominaux du type *le monde du travail*, *la volonté de changement*, etc.; la distinction de tels syntagmes avec les prédéterminants complexes peut être opérée par des règles de transformation¹².

On remarque que le substantif intégré dans les prédéterminants complexes a valeur collective: *nombre*, *millier*, *maximum*, *ensemble*, etc. Or ces prédéterminants, associés à des substantifs comme *travailleurs*, *salariés*, *militants*, etc.¹³ soulignent leur valeur collective. Il apparaît un effet de renforcement de l'aspect collectif, ici en ce qui concerne les actants sociaux. En contrepartie, le syntagme nominal ainsi constitué a un référent flou: les travailleurs et les salariés sont présentés comme une entité, sans spécification, sans diversité, sans différence. Paradoxalement, la prédétermination ne semble pas déterminer dans ces cas de figure: nous avons affaire ici à un procédé de quantification floue.

A l'opposé, des actants tels que *CGT*, *congrès*, etc., sont présentés dans leur spécificité, sans prédétermination complexe. La prédétermination produit un effet de sens en conférant sa valeur à la construction d'ensemble Det + N: les *travailleurs* n'existent plus en tant que tels, c'est de leur *masse* ou de leur *ensemble* dont il est question.

La comparaison menée ci-après de la prédétermination dans les deux corpus montre l'opposition des procédés rhétoriques utilisés par chaque Confédération.

Prédéterminants complexes spécifiques de la CGT

les dernières
la dernière
des dernières
des vingt dernières
ces dernières
aux différentes
l'ensemble de cette
l'ensemble de ces
cet ensemble de
cet ensemble d
aucun ensemble de
la majorité des
la masse des
du maximum de
centaines de milliers de
de milliers d
des millions de
la minorité des
bon nombre d
nombre d
de nombreux
de très nombreux
la plupart des
une autre
un autre
des autres
d'autre

Prédéterminants complexes spécifiques de FO

le dernier
ces derniers
ces trente dernières
les différents
des différentes
un minimum d
un même
moins de
peu d
un seul
une sorte de
de tels
de telles
toute autre

¹² M. Gross, *Grammaire transformationnelle du français*, 2, *Syntaxe du nom*, Paris, Cantilène, 1986.

¹³ Voir *Mots*, 14, mars 1987, p. 75.

une quantité de
une quelconque
quelque N que ce soit
les seuls
toutes sortes de
telle ou telle
tel ou tel
une telle
toutes les autres
toute une
toute son
toute sa
toute la
tout leur
tout le
tous les autres
tous ces
tous leurs
l'une des
beaucoup de
chacune des
chacun des
chacune de
chacun de

Prédéterminants complexes communs aux deux Confédérations

<i>les différentes</i>	<i>d'autres</i>
<i>des différents</i>	<i>une seule</i>
<i>un ensemble de</i>	<i>la seule</i>
<i>l'ensemble du</i>	<i>un tel</i>
<i>l'ensemble des</i>	<i>toutes ses</i>
<i>l'ensemble de</i>	<i>toutes les</i>
<i>le maximum de</i>	<i>toutes ces</i>
<i>le maximum d</i>	<i>toute leur</i>
<i>des milliers de</i>	<i>tous ses</i>
<i>Cardinal + millions de</i>	<i>tous les</i>
<i>le même</i>	<i>une des</i>
<i>la plupart des</i>	<i>un des</i>
<i>plus de</i>	<i>l'un des les autres une certaine</i>
	<i>nombre de</i>
	<i>de nombreuses</i>

Il apparaît à la lecture de ces listes comparatives que la quantification semble être de type globalisant, voire universel en ce qui concerne la CGT, et plus distinctive, voire discriminante, dans les textes de FO. Nous rejoignons là les résultats présentés dans *La parole syndicale*¹⁴ pour ce qui concernait les déterminants simples. Quoiqu'elle soit significative, cette différence de procédés entre les deux confédérations ne peut conduire à une interprétation trop rapide relative aux stratégies soit collectiviste, soit corporatiste, dont on serait tenté d'affubler ces deux syndicats.

¹⁴ A. Bergounioux, M.-F. Launay, R. Mouriaux, J.-P. Sueur, M. Tournier, *La parole syndicale*, Paris, PUF, 1982, p. 235.

QUELQUES RÉFLEXIONS SUR LANGUE ET DISCOURS DE BOIS

Le caractère partiellement politique des textes syndicaux nous incite à faire le parallèle avec le discours politique: la nature des phénomènes décrits précédemment renvoie-t-elle à la notion de «langue de bois»?

Les études portant sur la «langue de bois» sont déjà anciennes et concernent le discours soviétique pour la plupart.

Françoise Thom, dans un ouvrage récent intitulé *La langue de bois*¹⁵ souligne certaines caractéristiques de cette «langue de bois» à propos du discours soviétique. On peut sommairement en rappeler quelques-unes ici: la lourdeur syntaxique produite par l'abondance des tournures nominales, des groupes verbo-nominaux où le verbe n'assume plus qu'une fonction d'auxiliaire, l'absence d'embrayeurs, la présence massive du «nous », l'importance des tournures passives et impersonnelles, l'emploi excessif des comparatifs, etc. Selon l'auteur, ces procédés syntaxiques créent l'imprécision et maintiennent le flou et l'abstraction dans le discours. L'utilisation fréquente de figures rhétoriques (allégories, métonymies, etc.) transfigure la réalité.

Elle observe que chacun des procédés caractéristiques de la «langue de bois» repose à la fois sur une construction lexicale et un fonctionnement syntaxique spécifique: « Le discours communiste se trahit par sa syntaxe (syntaxe « de bois ») autant peut-être que par sa terminologie »¹⁶. Elle conclut¹⁷: «La langue de bois est essentiellement faite d'éléments empruntés à la langue naturelle; elle lui doit tout, les mots et la grammaire. Et pourtant, elle a subi une mutation tératologique par rapport à la langue naturelle. Elle renvoie à un univers sémantique extrêmement rudimentaire, bicolore en quelque sorte, mais elle ne se contente pas de priver les mots de leur substance, elle vise la capacité même du langage de représenter ».

Dans un travail antérieur¹⁸, L. Martinez rejoint cette conception de la « langue de bois » : « L'expression “ langue de bois ” a le mérite, éclairant mais incommode, de poser son objet comme une entité paradoxale. La qualification adjectivale (“ de bois ”), nécessairement secondaire par rapport à un concept premier de langue, limite celui-ci non par spécification, mais par négation implicite et globale. La “ langue de bois ” se propose métaphoriquement comme une non-langue en se définissant par la négation des attributs ordinairement appliqués à toute langue »¹⁹.

Soulignons l'ambiguïté de la notion de «langue de bois », telle qu'elle est illustrée par ces deux études: les auteurs partent d'une description du discours communiste et aboutissent à une critique des fonctions idéologiques de ce discours; la «langue de bois» communiste constituerait un ensemble de procédés visant à déconnecter le langage de la réalité; la «langue de bois» produirait donc un discours coupé du «réel», qui tournerait « à vide » pour satisfaire des objectifs politiques. Le formalisme, la répétition, l'uniformité, la rigidité contreviendraient à cette propriété essentielle d'une langue de permettre une communication dynamique, adaptative, créatrice entre énonciateurs. En ce sens, la « langue de bois » serait un langage vidé de ses messages. Une telle vision, fréquente en Occident parmi les politologues, conduit à décrire la « Novlangue » soviétique comme celle d'une « surréalité » destinée à assurer un pouvoir: A. Besançon par exemple parle de « logocratie ».

¹⁵ Paris, Julliard, 1987.

¹⁶ F. Thom, *La langue de bois*, op. cit., p. 22.

¹⁷ *Ibid.*, p. 102.

¹⁸ L. Martinez, «La “langue de bois” soviétique », *Commentaire*, 16, 1981-1982, p. 506-515.

¹⁹ *Ibid.*, p. 506.

Le linguiste n'est pas concerné par ce type de critique idéologique, et il convient de distinguer l'apport spécifiquement linguistique que recouvre la notion de « langue de bois », c'est-à-dire le repérage de certains mécanismes discursifs, de la visée idéologique que ces auteurs lui adjoignent.

En outre, cette expression renvoie trop souvent à un même type de discours: le discours communiste ou apparenté; or on sait que d'autres types de discours sont désignés comme des formes spécifiques de « langue de bois », par exemple le discours juridique, administratif, voire politique en général, et le discours syndical dont nous venons de voir quelques caractères; souvent, le destinataire non initié de tels discours n'en possède pas les termes clés (lexique technique) ni l'habitude de ces tournures syntaxiques, et n'est pas au fait des nuances internes à la « spécialité ». Il appréhende ces énoncés comme des descriptions ésotériques de la réalité.

On est alors conduit à penser que la « coupure » de la langue et de la réalité, si elle est effectivement ressentie par le lecteur, n'est pas toujours liée à des fins politiques ou doctrinales comme le prétendent les analystes de la « langue de bois » soviétique, mais à une « manière de parler », indépendante du lexique employé. Un tel discours s'apparente à une langue codée, un jargon dont il faut mettre en évidence les spécificités. Décrire de tels procédés rhétoriques revient donc à dégager la notion de « langue de bois » de son étiquette politique, polémique et ambiguë, pour lui conférer une légitimité scientifique.

Patrick Sériot dans ses analyse du discours soviétique et polonais²⁰ refuse la notion de « langue de bois » ossifiée, rigidifiée, mensongère, surréelle par opposition à une langue naturelle qui parlerait vrai. Il fait justement remarquer le caractère symétrique d'une telle conception: selon les Soviétiques, la langue occidentale, « bourgeoise », décrit, elle aussi, une pseudo-réalité, et donc est mensongère; elle vise à imposer une vision fautive à des esprits aliénés, elle est un instrument de pouvoir. Pour cet auteur, il convient de parler d'un « discours de bois » qui désigne l'expression langagière de l'univers public soviétique, expression distincte en tous points des langages pratiqués dans la société civile. Cette distinction langue/discours chez P. Sériot recouvre une discussion importante parmi les linguistes et dans « l'analyse de discours » en particulier; elle permet de comprendre comment les Soviétiques sont capables d'utiliser le « discours de bois », ce qui constitue une forme d'appropriation de cette rhétorique. Les descriptions des discours soviétiques effectuées par P. Sériot visent donc l'analyse des procédés linguistiques et non leurs fonctions idéologiques.

C'est ce dernier point de vue que nous avons retenu pour étudier le corpus syndical CGT et F0. Nous croyons qu'il est possible de dégager des éléments propres à élaborer une typologie du discours syndical fondée sur des traits proprement linguistiques, tels que la prédétermination nominale, et non, pas sur des catégories idéologiques. La métaphore « de bois » désigne dans ce cas un discours construit à partir d'un ensemble de procédés rhétoriques repérables.

A notre sens et malgré certaines interprétations, cette métaphore ne doit pas faire croire que l'on a affaire à un langage définitivement « figé », « mort », immobile: en effet, on caractérise souvent la « langue de bois » ou le « discours de bois » comme un langage composé d'expressions invariables et répétées. On peut par exemple entendre l'expression de « langue bureaucratique » ou, comme disent les Tchèques, de « langue morte ». Toutes ces expressions font référence à ces deux caractéristiques que sont le figement et la répétition. Or l'étude de Jean Breuillard²¹ montre que les discours de plusieurs chefs d'Etat soviétiques sont dans l'ensemble fort différents les uns des autres malgré des similarités ponctuelles: « On retiendra que si l'uniformité est incontestable, mais non absolue, dans certains corpus (éditoriaux de la *Pravda*), elle ne se vérifie pas en revanche, dans d'autres corpus: non seulement V. Lénine et I. Staline, mais aussi L. Brejnev ne “chargent” pas leurs phrases comme le font A. Kossiguine ou N. Tikhonov.... Langue de bois peut-être, mais — au moins au niveau syntagmatique superficiel

²⁰ P. Sériot. « Langue de bois et discours de vent », in *Essais sur le discours soviétique, sémiologie, linguistique, analyse discursive*, 1982, p. 5-39; *Analyse du discours politique soviétique*, Paris, Institut d'études slaves, 1985; « La langue de bois et son double », *Langage et société*, 35. mars 1986, p. 7-32; « Langue et langue de bois en Pologne », *Mots*, 13, octobre 1986, p. 181-189; Maurice Tournier, « Critique de la critique: langue de bois et parler vrai », *Mots*, 13, octobre 1986, p. 191-195.

²¹ J. Breuillard, « D'une seule voix ? », in *Essais sur le discours soviétique, sémiologie, linguistique, analyse discursive, op. cit.*, p. 49-61.

— bois dont le grain peut varier! ».

Le recours plus ou moins important aux nominalisations et à la subordination est très variable d'un milieu social et d'une région à l'autre en France même: C. Hagège y voit de simples indices sociolectaux qui « reflètent le milieu professionnel, familial, culturel, régional »²². La répétition de certaines tournures doit donc être interprétée avec une certaine prudence.

La comparaison de corpus de nature différente conduit P. Fiala²³ à affirmer dans le même sens: « Si la phraséologie paraît varier d'un type de discours à l'autre, elle est présente dans tous, et apparemment en proportion équivalente. Cela signifie qu'il n'y a pas des discours plus ou moins " phraséologiques ". On relève des unités complexes en nombre à peu près égal dans tout texte. Les différences proviennent de la nature et du fonctionnement des unités phraséologiques ».

Cette revue sommaire de quelques études centrées sur la « langue de bois » dans lesquelles notre travail sur le discours syndical prendra sa place montre clairement, par delà les divergences fortes sur la notion même, la place essentielle qu'il convient de réserver à l'analyse des lexies rigides et de leur rôle dans le discours.

On notera à cet égard que, si tous les auteurs précités soulignent le caractère massif du phénomène de lexisation, la plupart des études ne quantifient pas l'importance du recours aux différents types de lexies. Cette quantification différentielle a posé, comme nous l'avons indiqué, un problème d'interprétation. A cet égard, l'analyse du fonctionnement de ces lexies, par exemple celle des prédéterminants complexes, nous a paru éclairante en vue de l'organisation syntagmatique du texte. Cette nécessité d'analyse spécifique, au-delà de la quantification, est mise en évidence a contrario en remarquant que certaines variations de surface, internes à ces lexies, peuvent être considérées comme différentes attestations d'une même lexie dans un texte, et venir renforcer l'impression de répétition, alors qu'il n'y a pas d'uniformité ni de figement de surface statistiquement significatifs.

C'est dire que l'expression « langue de bois » recouvre des problèmes linguistiques réellement importants, dont certains aspects sont analysés depuis longtemps: leur réduction à l'espace politique constitue bien une dérive idéologique; l'étude de tels problèmes apparaît une nécessité s'il est possible de la réaliser sur un plan strictement scientifique, et en bénéficiant d'une mesure quantitative rigoureuse.

La construction d'une typologie, si elle est fondée, conduit le linguiste à considérer le discours comme une catégorie d'analyse pertinente, récusant et la forme graphique minimale et la phrase comme bases d'analyse car l'une et l'autre occultent les relations sous-jacentes entre les éléments constitutifs d'un discours. Ce qui importe dans l'étude du discours, c'est, à notre sens, plus que la seule répétition de formes lexicales, simples ou complexes, l'existence de structures syntaxiques qui constituent des modèles à travers lesquels s'opère la combinaison lexicale, et la régularité de ces structures.

février 1988

© Carmen Pineira-Tresmontant. *Cercle de Linguistique Appliquée à la Communication/ Círculo de Lingüística Aplicada a la Comunicación* 4, novembre 2000. ISSN 1576-4737. Publié dans *Mots* 17, 1988, p. 145-169.

²² C. Hagège, *La structure des langues*, Paris, PUF, 1982, p. 113.

²³ P. Fiala, « Pour une approche discursive de la phraséologie. Remarques en vrac sur la locutionnalité et quelques points de vue qui s'y rapportent, sans doute », à paraître dans *Langage et société*.