

Pour une grammaire du discours

Elaboration d'une méthode; exemples d'application

Ce texte présente un certain nombre de propositions d'ordre méthodologique pour une grammaire du discours à base statistique intégrant des faits lexicaux syntaxiques et énonciatifs. Ces propositions sont suivies d'exemples d'application à divers aspects du texte d'une résolution de congrès syndical¹.

PRINCIPES GÉNÉRAUX

Position par rapport à d'autres analyses

L'analyse des discours (ou, plus concrètement, des textes politiques, puisque de nombreuses études se sont attachées à ce type de corpus) s'est largement développée en France durant les quinze dernières années dans deux directions principales : la lexicométrie ou

1. Les travaux présentés ici ont largement utilisé les données et les résultats de l'étude sur le vocabulaire syndical qui a été menée par l'équipe Mouvement ouvrier de l'Unité de recherche Lexicologie et textes politiques de l'Institut de la langue française — étude aujourd'hui parue sous le titre : *La parole syndicale, analyse du vocabulaire des confédérations syndicales ouvrières*, par A. Bergognoux, M. Launay, R. Mouriaux, J.-P. Sueur, M. Tournier, Paris, PUF, 1982. La partie statistique et informatique de ce travail a été effectuée avec la collaboration de Monique Pontier, de l'Université d'Orléans. Quelle en soit vivement remerciée.

analyse statistique des formes lexicales d'un texte d'une part, et, d'autre part, les études inspirées des principes à la fois distributionnels et transformationnels exposés par Harris dans un texte intitulé *Discourse Analysis*, paru en 1952, et présentant presque toujours divers types de modification ou d'extension des principes posés par Harris — le terme d'«analyse de discours» désignant le plus souvent cette seconde tendance. De nombreux travaux ont été menés dans ces deux directions qui permettent de mieux mesurer l'apport, mais aussi les limites de chacune d'entre elles².

Deux limites sont communes à ces deux orientations ; elles ont été rapidement notées :

La première tient au fait que, ni la lexicométrie, ni les analyses harrisiennes, considérées stricto sensu, ne prennent en compte l'étude des faits d'énonciation, dont les textes pionniers de Jakobson (1963)³ et Benveniste (1965) avaient montré l'importance. Très vite, dans un cas comme dans l'autre, la description de ces faits a été intégrée, ou juxtaposée, ou présentée comme l'un des postulats ou des résultats de l'analyse⁴ — la question de savoir comment devait précisément s'opérer cette «intégration» aboutissant souvent à des remises en cause assez profondes des modèles de départ.

La seconde est liée au fait que si ces deux méthodes permettent la mise à jour de récurrences globales dans un corpus (lexicales dans un cas, syntactico-lexicales dans l'autre), ou encore de similitudes ou de différences dans la répartition de telles récurrences dans deux ou plusieurs corpus distincts, elles ne permettent pas de décrire les règles qui relient une phrase et la suivante et, de proche en proche, la succession totale des phrases dans un corpus. Autrement dit, si la cohérence d'un texte est appréhendée dans sa globalité (par la mesure ou la mise en évidence de régularités lexicales ou syntaxiques), elle ne l'est pas par rapport au déroulement linéaire du texte. Les travaux de Bellert (1970), de Van Dijk (1972), et plus généralement du courant de la «grammaire de texte» — qui s'est particulièrement développé en Allemagne — ont montré qu'il était possible de poser les bases d'une grammaire transphrastique — même si le risque existe (mais il existe dans toutes les formes d'«analyse de discours») de voir la rigueur de la démarche se dissoudre aux confins de la

2. Pour un bilan sur l'ensemble de ces approches, voir notamment : Robin (1973), Guespin (1975), Maingueneau (1976), Guilhaumou (1977).

3. Ces références renvoient à la bibliographie, en fin d'article.

4. Dans le cadre de l'analyse de discours, voir notamment Maldié (1971 a et b), Marcellesi (1971 et 1976). Pour une analyse de faits d'énonciation à partir de données lexicométriques, voir notamment Sueur (1980).

rhétorique, de la «pragmatique», et d'une «logique naturelle» qui apparaît, dans de nombreux travaux linguistiques des dix dernières années, comme la dénomination d'une stratégie linguistique — voire d'une morale — universelle, stratégie et morale qui n'existent comme telles — comme «naturelles» — que dans les constructions idéologiques qui fondent cette prétendue logique.

A ces premières critiques s'en ajoutent d'autres qui tiennent aux fondements même de chacune de ces démarches.

LA LEXICOMÉTRIE

Je reviendrai plus loin sur les critiques qui mettent en cause la légitimité de l'utilisation d'outils statistiques, et aussi sur les critiques liées aux diverses solutions qu'il est possible d'adopter pour résoudre le problème de la lemmatisation.

Je m'en tiendrai ici à une remarque sur l'objet de la description statistique ou lexicométrie : l'objet de la lexicométrie n'est pas le texte ou le discours ; il est le texte réduit à un dictionnaire de formes — c'est-à-dire n'existant plus comme texte. Qu'il soit utile de décrire les structures, les régularités et les spécificités de cette liste de formes, cela est clair. Mais l'étude du discours, en ce qu'il est configuration spécifique de structures syntaxiques, est une autre tâche, même si elle peut se nourrir de la précédente. Ainsi, l'élimination de la syntaxe amène à mettre sur le même plan des occurrences de formes dont le fonctionnement est très différent à l'intérieur du réseau des relations syntaxiques et des systèmes d'énonciation.

En dépit de leur grand intérêt pour la mise en évidence de nombreux stéréotypes, les programmes dits de «cooccurrence» (cf. Lafon et Tournier, 1978) ne portent que très partiellement remède à cet inconvénient, dans la mesure où ils n'étudient les rapports entre une forme et celles qui lui sont cooccurrentes que sur la base d'une proximité évaluée en termes strictement numériques. Ainsi, on examine les cinq formes qui précèdent et suivent une forme donnée. Mais, là encore, les rapports de dépendance syntaxique sont gommés. Toutes proportions gardées, c'est comme si on décrivait pour chaque pièce d'un moteur, la pièce située à une proximité de 5 cm., puis celle située à 10 cm., etc. Ces mesures n'auraient qu'un intérêt relatif : une pièce peut être en connexion directe avec une pièce éloignée, et fonctionner de manière totalement indépendante d'une pièce proche. Il en va de même dans

les discours où les relations de dépendance syntaxique (« profondes ») entre les formes ne se confondent pas avec leurs relations de proximité (« superficielles »).

LES ANALYSES « HARRISSIENNES »

Les analyses élaborées par Dubois, Marcellesi, Maldidier, etc.⁵, ou par Pêcheux et d'autres⁶, malgré d'importantes différences de traitement, posent toutes l'hypothèse que les rapports entre syntaxe et lexique dans un discours peuvent s'analyser à partir d'une déconstruction⁷ des phrases de ce discours en propositions de base, points de départ d'une analyse distributionnelle. Ces analyses, qui partent le plus souvent de mots-pivots sélectionnés en fonction de leur fréquence dans le texte, rééditent donc partiellement le texte sous la forme d'une liste de propositions élémentaires qui présentent un cadre adapté à la comparaison de l'occurrence des mots-pivots dans ces propositions. Les relations entre ces propositions élémentaires au sein de la phrase sont, soit laissées de côté, soit incorporées à l'analyse sous la forme d'un graphe dans l'Analyse automatique de discours.

Ces analyses donnent de précieuses indications pour l'étude des « relations profondes » existant entre les formes d'un texte. Mais leur principale limite vient de ce que le sens d'un discours et même d'une phrase ne se réduit pas à l'addition ou à la combinaison des lectures sémantiques des « propositions de base » ainsi définies. Le modèle sémantique qui régit ces descriptions reste celui de la grammaire générative classique (la « théorie standard ») en ce qu'il postule d'abord que les règles sémantiques procèdent par amalgame, en remontant de bas en haut les hiérarchies syntaxiques telles qu'elles sont représentées par les « arbres » (c'est le modèle des « règles de projection » de Fodor-Katz, 1963), et ensuite, et surtout, que « les transformations ne changent pas le sens » (c'est le modèle de Katz-Postal, 1964).

Or il est pour le moins paradoxal que les analyses de discours « harrissiennes » — fondées en fait sur ce modèle sémantique et syntaxique — se soient précisément développées en France au moment même où les limites de ce modèle apparaissaient, à mesure qu'on

5. Cf. les références de la note précédente, et aussi les bibliographies des ouvrages indiqués dans la note 2.

6. Sous le nom d'« Analyse automatique de discours » (cf. Pecheux, 1969 et Pecheux-Fuchs, 1975).

7. Processus qui renvoie à ce que Marcellesi appelle une « contre-grammaire » (cf. Marcellesi, 1976).

observait (cf. notamment Jackendoff, 1972) que les phénomènes sémantiques relevaient à la fois de la structure profonde et de la structure superficielle⁸.

Allant plus avant dans la même voie, on a pu soutenir que toutes les transformations (du moins toutes les transformations facultatives) modifiaient le sens d'une phrase. On observe dans tout discours qu'il n'est sans doute pas indifférent que tel nom apparaisse comme le sujet d'un verbe actif ou le complément d'agent du même verbe à la forme passive. Il n'est pas non plus indifférent qu'une proposition finale précède ou suive un verbe principal, que tel rapport «sujet-verbe» apparaisse sous cette forme ou sous celle d'un syntagme nominalisé. (Ainsi, les réductions en propositions élémentaires aboutissent souvent à d'importantes listes de propositions prédicatives avec *être* : mais, ce faisant, on met sur le même plan des phrases de définition, présentes comme telles dans le texte, et des relations d'adjectivisation ou de nominalisation où la définition (ou l'équivalence ou la qualification) ne sont qu'implicites — autrement dit, on abolit la différence entre une définition explicite et une définition implicite, ou présupposée, alors qu'il s'agit là de deux actes d'énonciation distincts : il est très éclairant d'étudier dans un texte les définitions ou les postulats qui sont posés comme tels et ceux qui «vont de soi» (font partie des présupposés)).

Le champ des opérateurs de modalité (déterminants, verbes modaux) n'est pas non plus indifférent. Et l'on pourrait multiplier les exemples à l'infini. A la limite, le sens est tout entier inscrit dans la «structure de surface», dans la mesure où elle porte, de toute manière, la trace (pour reprendre, dans un sens différent, mais dans le même esprit, la métaphore de Chomsky, 1980) de relations dites profondes. Mais le sens ne se limite pas à cette trace, et le discours ne se réduit pas aux propositions de base : l'interprétation sémantique est aussi étroitement liée à la structure linéaire du texte. Il en va de même pour la plupart des faits d'énonciation, et des faits qui constituent, dans sa linéarité, la cohérence d'un texte.

Grammaire de discours

J'appellerai «grammaire de discours» l'étude des régularités dans la distribution, les uns par rapport aux autres, des faits de lexique, de syntaxe et d'énonciation dans un corpus. Cette grammaire présuppose donc l'étude de régularités lexicales d'une part, syntaxiques et

8. Du moins, selon la terminologie des théories «interprétatives».

énonciatives d'autre part, dans un texte. Elle s'appuie sur ces deux types d'analyse pour décrire les régularités existant dans le texte dans l'occurrence de formes lexicales au sein de cadres syntaxiques et énonciatifs spécifiques, et à des places spécifiques à l'intérieur de ces cadres.

Autrement dit, le but de cette grammaire est de définir le plus précisément possible les caractéristiques qui permettent d'identifier spontanément un discours : nous reconnaissons le discours de tel individu, de tel groupe, de tel parti, de tel syndicat, etc. Or, si la spécificité du lexique joue un rôle dans ce processus de reconnaissance, d'autres facteurs interviennent : la syntaxe, les faits d'énonciation, mais surtout la connexion entre ces divers types de faits. Tel mot apparaît de manière privilégiée, mais, surtout, il apparaît de manière privilégiée à telle place et dans tel cadre. Tout cela fait partie de la « compétence » propre au locuteur — et vient se combiner avec les traits qui définissent les diverses formes d'énonciation (une résolution n'est pas un discours ni la réponse à une interview, etc.). La grammaire du discours est donc l'étude systématique de l'intersection entre faits de lexique et faits de syntaxe et d'énonciation. En ce sens, elle est une contribution à l'élaboration d'une typologie des discours, dans le sens que J. Simonin-Grumbach (1975) donnait à cette expression.

Cette grammaire de discours possède plusieurs autres caractéristiques : elle est probabiliste ; elle est à la fois grammaire de fréquences et grammaire de séquences ; enfin, elle implique certains choix dans la détermination des régularités envisagées et dans la manière de les traiter.

UNE GRAMMAIRE PROBABILISTE

L'une des conséquences de la prédominance des travaux s'inspirant des grammaires génératives dans la linguistique contemporaine a été un moindre intérêt apporté à l'étude des statistiques des faits linguistiques⁹. Les considérations statistiques n'ont, en principe, aucune validité en grammaire générative, puisqu'un exemple suffit à légitimer ou à contredire une analyse. Ceci va d'ailleurs de pair avec l'absence de référence à un corpus.

Mais on peut se demander si l'une des limites — ou des illusions — des grammaires génératives ne vient pas de là. Limite, car l'étude des corpus montre nombre de régularités dans l'application (ou la non-application) de règles (syntaxiques et autres) qu'il n'est pas

9. En dépit des travaux menés, notamment, par Muller (1968, 1977, 1979).

satisfaisant de reléguer dans la « performance » (concept qui désigne « en creux » ce que la théorie se propose de ne pas analyser, ou, ce qui revient au même, renvoie constamment à une seconde étape qui ne vient jamais). Illusion, car les jugements de grammaticalité ne sont peut-être que la construction d'un corpus, ou plutôt la projection de l'image qu'on se fait d'un corpus sur quelques exemples. Dire que le modèle (ou l'axiome) de base de la description linguistique, c'est la phrase canonique comportant un syntagme nominal et un syntagme verbal, c'est probablement faire une affirmation qui n'est fondée que sur des considérations de fréquence. C'est faire une affirmation qui n'a pas d'autre justification que l'intuition selon laquelle cette structure est la plus fréquente (même si on ne fait pas de mesure) dans le type de langue (un certain type de langue écrite appelée « langue standard ») qu'on s'est donné — ou construit — comme objet d'étude.

Poser une règle comme universelle (comme le font les grammaires génératives), c'est toujours se référer à une norme, et à la possibilité (ou à la réalité) d'un corpus qui illustrerait cette norme. Partir d'un corpus concret, c'est postuler au départ que les règles seront relatives à ce corpus, qu'elles pourront ne se vérifier que pour une partie du corpus, et c'est donc s'engager sur la voie d'une grammaire de type probabiliste.

Cette démarche s'apparente à celle mise en œuvre par Gross dans un autre domaine lorsque, critiquant Postal (1974), il écrit :

« De nombreux travaux semblent avoir pour but de déterminer si le cas de *s'imaginer* a, ou n'a pas, un statut de transformation (pris dans un sens technique très particulier). Apparemment tous les auteurs sont d'accord sur le fait que *sembler* subit bien une transformation. Aucun de ces auteurs cependant ne s'est soucié de dresser la liste des verbes du type *sembler*, ni celle du type *s'imaginer*. Nous l'avons fait pour le français : il existe moins de cinq verbes du type *sembler*, et plus de six cents du type *s'imaginer*. Il apparaît donc, de ce point de vue, que le cas de *s'imaginer* possède un caractère de généralité que n'a pas du tout le cas de *sembler*, et rien ne permet de prévoir cette situation. Il n'est donc guère justifié de donner plus de poids dans une grammaire au cas de *sembler* qu'à celui de *s'imaginer* » (1976, p. 9).

Cette position marque une rupture avec celle de la grammaire générative dans la mesure même où elle implique que la fréquence d'une propriété dans le lexique constitue un critère pour décider entre plusieurs descriptions alternatives d'un même phénomène. La seule question qu'on pourrait se poser à ce sujet est de savoir s'il est légitime de mesurer cette

fréquence relative par rapport à l'ensemble du lexique (au dictionnaire) ou par rapport à des énoncés concrets : une propriété peut s'appliquer à cinq verbes qui seraient globalement plus fréquents en discours que les six cents verbes qui possèderaient une propriété alternative. (Il a d'ailleurs été souvent montré que les mots les plus fréquents étaient ceux dont le fonctionnement était le plus irrégulier ; cf. pour les verbes, Dubois, 1967). La méthode de Gross renvoie à une conception statistique de la grammaire, puisqu'elle cherche à évaluer l'extension de chaque phénomène dans le lexique — mais elle est fondée sur le postulat selon lequel chaque terme a dans le lexique le même « poids » qu'un autre ; or une telle configuration n'existe que dans la fiction des dictionnaires.

La grammaire de discours sera probabiliste en un autre sens. Le système de règles qui constituera cette grammaire sera fondé sur la mise en évidence de diverses corrélations. Ce système vise donc à rendre compte d'une compétence référée à la situation énonciative et fondée sur la combinaison optimale des diverses règles. Le fait que l'application des règles souffre d'un nombre plus ou moins grand d'« exceptions » doit être considéré comme l'une des propriétés de la syntaxe étudiée. D'où la nécessité de la mesurer. Les règles ne s'appliquent donc qu'à un certain pourcentage des données. Cela paraît inéluctable dès lors que l'on fonde la grammaire sur l'étude d'un corpus. La seule alternative qui consisterait à ne prendre en compte que les règles à valeur absolue conduirait à ne poser que des règles tellement générales qu'elles laisseraient échapper les spécificités du corpus (pour une première application de ces principes à l'étude de données linguistiques, voir Sueur, 1978).

GRAMMAIRE DE FRÉQUENCES ET GRAMMAIRE DE SÉQUENCES

La « grammaire de fréquences » considère le texte comme un ensemble homogène et étudie les fréquences lexicales, syntaxiques, et les corrélations entre les unes et les autres d'une manière globale, sur l'ensemble du corpus. Le corpus ainsi homogénéisé est une abstraction par rapport au corpus réel, mais l'idée même d'une grammaire probabiliste conduit nécessairement à ce type d'abstraction. D'où l'intérêt d'une « grammaire des séquences », qui étudie la distribution des règles précédentes à l'intérieur de la séquence des phrases, et donc dans le déroulement linéaire du texte. Ce qui conduit à la mise à jour d'autres régularités dans la disposition des séquences, sur une base également probabiliste, régularités qui peuvent être représentées sous la forme de règles similaires à celles proposées,

notamment par Marcus (1978), pour la grammaire de texte. (Pour une première application, voir également Sueur 1978).

CRITÈRES DE REGROUPEMENT ET FORMES D'ANALYSE

a. La description exhaustive de la structure syntaxique des phrases sous la forme d'une «grammaire en chaînes» du type de celle de Salkoff (1973) conduisant à une diversité de structures pratiquement aussi grande que le nombre des phrases paraît exclue : elle ne permet pas, en elle-même, de mesurer les récurrences. D'où la nécessité de sélectionner des sous-chaînes qui soient comparables parmi les chaînes majeures (définies par la nature du verbe principal), les chaînes enchâssées (complétives, etc.) et les chaînes nominales (correspondant à des configurations effectives de SN), etc., et de mesurer dans un second temps les contraintes et les combinaisons existant entre ces différents types de chaînes. L'analyse met en lumière à la fois les différences et les ressemblances entre chaînes majeures et chaînes nominalisées dans la mesure où le codage permet d'indiquer dans les deux cas les propriétés du verbe qui constitue la «tête» du syntagme. On retrouve là les hypothèses sur la notation \bar{X} de Chomsky (1972), et le choix pour une théorie interprétative qui ne nie pas les relations «profondes» entre les termes mais traite toujours ces relations en référence aux configurations superficielles.

b. Le nombre relativement restreint des verbes principaux permet de classer les chaînes majeures à partir de propriétés syntaxiques et d'énonciation. Les classes et sous-classes correspondent à divers actes d'énonciation dont la combinaison constitue la résolution (puisque c'est le «type» dominant du corpus) comme forme de discours.

c. Les régularités lexicales peuvent être définies au niveau des mots les plus fréquents, ou des mots les plus spécifiques (cette seconde solution renvoyant nécessairement à d'autres corpus). Mais la fréquence n'est pas un critère en elle-même : elle intervient seulement dans la mesure où la possibilité d'établir une analyse factorielle des correspondances syntactico-lexicales, qui soit statistiquement pertinente (voir ci-dessous), conduit nécessairement à la détermination de seuils. Des regroupements lexicaux doivent ainsi être faits sur la base de classes sémantiques ou de rapports métonymiques. De tels regroupements (qui ne sont pas nécessaires si on travaille sur la base d'un nombre restreint de mots fréquents) reposent pour une part sur des choix relativement arbitraires, qui doivent alors être présentés comme tels.

d. Pour ce qui est de la *lemmatisation*, j'adopterai une position quelque peu différente de celle de Geffroy, Lafon et Tournier (1973). En effet, le principe selon lequel il faudrait appliquer à cet égard la même solution pour toutes les parties du discours ne me paraît pas fondé. S'il est vrai que «le déséquilibre entre singulier et pluriel paraît bien être la règle et non l'exception pour chaque substantif» — il suffit de penser à la forme *travailleurs*, forme la plus fréquente du corpus des résolutions syndicales, dont le comportement est clairement dissymétrique au singulier et au pluriel —, on ne voit pas pourquoi la non-lemmatisation, justifiée dans ce cas, devrait pour autant être étendue à l'ensemble du lexique. C'est pourquoi, je considérerai que chaque forme des substantifs (selon le genre et le nombre) doit être considérée isolément, mais qu'en revanche les formes des verbes et des adjectifs dont le nombre et, éventuellement, le genre ne sont que des conséquences mécaniques du nombre (et éventuellement du genre) du ou des substantifs correspondant peuvent — et même doivent — être regroupées. Et cela d'autant plus que les différences de temps et de forme (actif/passif) sont de toute manière mentionnées dans le codage.

e. Sur ces bases, s'opère une analyse factorielle de correspondances. Cette méthode, qui a déjà été appliquée à des données linguistiques par Prost (1974) et par Girard, Serman, Cadet et Gossez (1976)¹⁰, permet de mettre en évidence les régularités existant entre un grand nombre de facteurs — pour peu que l'on interprète correctement les contributions de chaque facteur dans la constitution de chacun des axes que cette analyse détermine. Mais cette analyse ne prétend nullement à l'«automaticité». Le très grand nombre des structures syntaxiques et énonciatives et des occurrences lexicales qui constituent un texte exclut qu'on puisse «saturer» une analyse de leurs relations. Il y a donc nécessairement choix, simplification et réduction dans la mise en évidence de certains types de croisement plutôt que d'autres. Il importe simplement que ces réductions, qui permettent de constituer une (autre ?) lecture, soient définies de la manière la plus explicite possible.

10. Voir aussi certains travaux de l'Unité de recherche Lexicologie et textes politiques de l'ILF (notamment le chap. 2 de l'ouvrage à paraître, *La parole syndicale*) et Sueur (1980).

EXEMPLES D'APPLICATION

Nous allons maintenant présenter quelques exemples très limités d'application de la méthode exposée ci-dessus. Ces exemples — qui ne sont que quelques jalons vers une grammaire de discours qui intégrerait également de nombreux faits qui ne sont pas abordés ici — sont tous extraits de l'analyse de la Résolution générale du congrès de 1976 de la CFDT¹¹. La première étape de l'analyse consiste en un codage du texte dont on possède déjà un index des fréquences¹² des formes, les formes verbales et adjectivales ayant été lemmatisées selon les principes exposés ci-dessus.

Le codage

Chaque phrase est divisée en une série de chaînes qui sont toutes codées sur une carte de 80 colonnes selon le découpage suivant :

LES COLONNES 1 à 9

Elles contiennent un certain nombre d'indications relatives à l'ensemble de la chaîne. Les colonnes 1 à 3 indiquent le numéro de la phrase ; la colonne 4 le numéro de la proposition à l'intérieur de la phrase dans le cas de propositions principales ou indépendantes coordonnées ; la colonne 5 indique s'il s'agit d'une principale, d'une complétive, d'une relative, ou d'une circonstancielle (et le type de la circonstancielle : dans le texte présent on a isolé les propositions finales, extrêmement fréquentes, qui sont d'ailleurs codées différemment

11. Cf. *CFDT : textes de base*, Paris, Montholon-Service, 1977, p. 5 à 56.

12. En réalité l'index des fréquences porte à la fois sur la Résolution générale et sur la Résolution sur l'action du même congrès. Mais le second texte étant beaucoup plus court que le premier, et celui-ci ayant été entièrement codé selon la méthode exposée, il n'y a pas d'inconvénient à travailler à partir de l'index correspondant aux deux résolutions.

selon qu'elles précèdent ou suivent le verbe principal)¹³; la colonne 6 marque le type énonciatif de la phrase : cinq types ont été distingués qui correspondent à la classe du verbe principal, telle que celle-ci est définie ci-dessous (on a le type *performatif* (verbe de la classe 0), *législatif* (verbe de la classe 1), *possibiliste* (verbe de la classe 2), *définitoire* (verbe de la classe 3), *analytique* (verbe d'autres classes; la phrase présente une description)); les colonnes 7 et 8 indiquent le type de la chaîne selon le codage suivant :

11. chaîne centrale (le verbe est intransitif ou suivi de compléments nominaux ou à l'infinitif);
12. chaîne complétive (le verbe est suivi d'une complétive, codée comme telle sur la colonne 5, de la carte suivante);
13. chaîne contenant un adjectif attribut;
14. chaîne contenant un adjectif attribut du complément d'objet;
15. chaîne introduite par *c'est*;
16. chaîne introduite par *il y a*;
17. chaîne impersonnelle;
21. chaîne de nominalisation SV (le rapport entre le complément et le verbe nominalisé est un rapport de sujet à verbe, cf. Ruwet, 1968);
22. chaîne de nominalisation OV (le rapport entre les deux termes est un rapport de complément d'objet à verbe; un sujet est postulé dans le contexte);
23. chaîne de nominalisation OV (même description que précédemment; aucun sujet n'est postulé dans le contexte);
31. chaîne nominale OS (le rapport entre le complément et le nom est un rapport de complément d'objet à sujet; cf. Ruwet, 1968);
32. chaîne nominale contenant un adjectif épithète;
33. chaîne nominale contenant une relative;
40. chaîne adjectivale.

On remarque que les chaînes 21 à 31 présentent sous d'autres formes des relations existant au sein des chaînes 11, mais que ces chaînes ne sont pas réduites en chaînes 11 : ce qui permet à la fois d'établir les équivalences nécessaires et de marquer les différences entre ces divers types de chaînes au sein des processus d'énonciation.

La colonne 9 est marquée «0» s'il n'y a pas coordination et «1» s'il y a coordination entre l'un quelconque des constituants mentionnés sur la carte au-delà de la colonne 9 et un

13. Toutes les chaînes dépendant du même verbe principal portent donc le même numéro dans les colonnes 1 à 4.

ou plusieurs autres constituants du même type. Dans ce cas, les constituants coordonnés sont codés sur la carte marquée «1» exactement à la place où sont codés les constituants auxquels ils sont coordonnés sur la carte correspondante marquée «0». Cela permet de rendre compte de presque tous les cas de coordination, d'étudier les corrélations entre éléments coordonnés et de tenir compte de l'ensemble des éléments coordonnés ou non apparaissant à une place définie dans un cadre syntaxique lorsque l'on étudie, par exemple, les corrélations entre la nature des éléments et les places de leurs occurrences à l'intérieur de ce cadre.

LES COLONNES 10 à 80

Elles sont ainsi réparties :

- col. 10-15: N1 (chaînes 11 à 14: groupe nominal sujet; chaînes 21 à 33: nom tête du SN).
- col. 10-13: ADJ (chaînes 40 où l'ADJ est la tête d'un syntagme).
- col. 16-24: V1 (verbe principal).
- col. 25-30, col. 31-36: N2 et N3 (groupes nominaux compléments du verbe dans les chaînes 11 à 17, du nom dans les chaînes 21 à 31 et de l'adjectif dans les chaînes 40).
- col. 25-27, col. 31-33: ADJ attribut dans les chaînes 13 et 14.
- col. 37: préposition introduisant V2.
- col. 38-46: V2: verbe à l'infinitif dépendant de V1 (chaînes 11 à 17) ou verbe à l'infinitif, au participe présent ou passé dépendant de N1 (chaînes 21 à 31) ou de l'ADJ (chaînes 40).
- col. 47-52, col. 43-58: N4 et N5 (groupes nominaux compléments de V2).
- col. 59: nature des circonstanciels (les circonstanciels de but sont, à nouveau, isolés).
- col. 60-68: V3 (circonstanciel à l'infinitif).
- col. 69-74, col. 75-80: N6 et N7 (groupes nominaux circonstanciels ou compléments de V3).

Remarque: le nom tête (N1) de toute chaîne nominale (21 à 33) apparaît préalablement dans l'une quelconque des positions N1 à N7 d'une chaîne 11 à 17. Il faut donc veiller dans les décomptes à ne pas le compter deux fois.

LE CODAGE DU GROUPE NOMINAL

Il consiste en un nombre de six chiffres, répartis sur six colonnes. La première colonne indique la préposition (1: absence; 2: *de*; 3: *à*; 4: *par*; 5: *pour*; 6: *dans*; 7: *avec*; 8: *sur*; 9: divers). Les listogrammes des colonnes 25, 31, 37, 47, 53, 59, 69, 75 représentent donc l'ensemble des cadres syntaxiques existant dans le corpus, compte tenu que chacune de ces

colonnes peut être marquée «0», ce qui correspond non pas à l'absence de préposition (situation marquée «1»), mais à l'absence du complément correspondant (ou du verbe à l'infinitif pour la col. 37, ou du circonstanciel pour la col. 59).

La seconde colonne indique la nature du déterminant et l'interprétation sémantique qu'il reçoit.

Cette interprétation est déterminée à l'aide du clivage traditionnel : défini/indéfini, mais aussi des clivages : spécifique/non spécifique et spécifique/générique tels qu'ils ont été définis notamment par Jackendoff (1972) et Sueur (1977 a). On a le codage suivant :

1. Défini spécifique.
2. Défini non spécifique ou générique.
3. Indéfini spécifique.
4. Indéfini non spécifique.
5. Générique (= tout, etc.).
6. Possessif.
7. Démonstratif.
8. Divers.
9. Pronom anaphorique (dans ce cas, les colonnes suivantes indiquent les caractéristiques du syntagme nominal auquel ce pronom anaphorique se réfère).

On voit que la combinaison des colonnes 11, 26 et 32 (mais aussi des colonnes correspondantes pour N4 à N7) donne les informations permettant de déterminer (en combinaison avec la nature du verbe V1, son mode, son temps, ou d'autres facteurs) ce que Jackendoff appelle la structure modale d'une proposition.

Les trois dernières colonnes indiquent le numéro du nom : un numéro a été donné aux 244 formes nominales apparaissant au moins cinq fois dans la résolution. Les numéros sont disposés par ordre de fréquence décroissante (voir annexe 1), ce qui permet, compte tenu du nombre des formes, de n'appliquer tel ou tel programme de corrélations qu'aux 10, 20, 30, 50 ou 100 formes les plus fréquentes.

Enfin, la quatrième colonne indique la classe du nom. Seuls quelques noms ont été classés. Ce classement comporte, comme cela a été indiqué ci-dessus, une part d'arbitraire. Celle-ci a été limitée dans la mesure où dans la plupart des classes le principe retenu a été celui de la métonymie (la classe 1 comprend les locuteurs collectifs (la CFDT, le congrès), la

classe 2 comprend les individus ou groupes qui sont inclus dans la classe 1; la classe 3 regroupe ceux au nom de qui s'expriment les locuteurs de la classe 1; la classe 3 inclut donc la classe 1 qui inclut la classe 2) (voir l'annexe 2).

LE CODAGE DES VERBES

Ce codage consiste en un ensemble de neuf colonnes.

- La première indique la forme positive (1) ou négative (2).
- La seconde indique le statut d'actif (1) de passif (2) ou de pseudo-passif (3) (c'est le cas des verbes précédés du pronom «se» sous sa forme «se moyen» (cf. Ruwet, 1972 et Dubois, 1967)).
- La troisième indique le statut du verbe dans le discours par rapport à la notion de contrôle telle qu'elle a été définie notamment par Fisher et Marshall (1969) et Sueur (1977 b). Un verbe est marqué [+ contrôlable] si le sujet superficiel peut contrôler, maîtriser l'action exprimée par le verbe (seuls les verbes ainsi marqués peuvent s'employer à l'impératif; et un verbe passif ne possède pas par définition cette marque). Le codage est le suivant: 1) sujet animé (+ contrôlable); 2) sujet animé (– contrôlable); 3) sujet inanimé (– contrôlable).
- La quatrième colonne indique le mode et/ou le temps du verbe.
- Les trois dernières colonnes indiquent le numéro du verbe, les verbes étant numérotés comme les noms par ordre de fréquence décroissante (voir l'annexe 3).
- La cinquième et la sixième colonne indiquent la classe du verbe. Ces classes ont été définies à partir de regroupements sémantiques et syntaxiques.

Classe 0: performatifs

Classe 1: verbes d'obligation et de nécessité

10. opérateurs de phrase (*devoir*)

11. prédicats à deux places (*exiger*)¹⁴

12-15. verbes d'implication

Classe 2: verbes de possibilité

20. opérateurs de phrase (*pouvoir*)

21. prédicats à deux places (*permettre*)

14. Sur les notions d'opérateur de phrase et de prédicat à deux places, voir Sueur (1977 a et b).

Classe 3: verbes définitoires (*être, avoir, comporter, etc.*)

Classe 4: verbes analytiques

41. susceptibles d'être (+ contrôlables) (*manifeste, traduire, etc.*)

42. (-contrôlables) (*s'avérer, apparaître, etc.*)

43. verbes de coordination (*lier, etc.*)

44. analytiques divers

45. verbes d'état

Classe 5: verbes marquant l'extension dans l'espace et le temps

Classe 6: verbes « psychologiques »

60. verbes marquant une tension positive (*vouloir, etc.*) et toujours suivis d'un infinitif V2

61-63. divers

Classe 7: verbes exprimant un *processus* (marqués (+ contrôlables))

71 à 75. processus positif (*créer, construire, etc.*)

76. processus négatif (*mettre fin, détruire, etc.*)

Classe 8: autres verbes marqués (+ contrôlable)

On constate que ces classes, qui sont sémantiquement (et souvent syntaxiquement) cohérentes, ne correspondent pas systématiquement aux diverses classes syntaxiques établies par Gross (1975). Ainsi, les verbes performatifs, qu'il est assez facile d'isoler comme tels, relèvent de classes différentes parce qu'ils se construisent avec les prépositions différentes et entrent dans des cadres différents. Par rapport aux convergences qui apparaissent dès lors qu'on envisage les faits d'énonciation, certaines de ces différences syntaxiques peuvent donc apparaître comme relativement « superficielles » (au sens de Chomsky) (voir l'annexe 4).

Analyses

CHAÎNES COPULATIVES

Un premier exemple simple d'analyse peut être fourni par les chaînes majeures copulatives. Ces chaînes établissent un rapport d'équivalence entre deux SN. Ce rapport peut être de divers types selon la nature des déterminants des noms (définis/indéfinis). On a :

— Les définitions exclusives du type + *Déf N1 être* + *Déf N2* qui se caractérisent par le fait que (a) N2 ne définit que N1, mais que (b) N1 peut être défini autrement que par N2.

— Les définitions non exclusives du type + *Déf N1 être* – *Déf N2* où (b) reste vrai, mais non (a).

— Les définitions contenant deux indéfinis du type – *Déf N1 être* – *Déf N2*.

— Les définitions exclusives bilatérales du type + *Déf N1 être INFINITIF* où (a) le syntagme à l'infinitif ne définit que N1 et (b) N1 n'est défini que par lui.

On constate, à partir de cette analyse :

a. que les séquences – *Déf N1 être* + *Déf N2* sont exclues (contrainte qui se situe probablement au niveau de la langue) ;

b. que le fait de faire entrer certains SN dans une définition exclusive est un acte énonciatif par lequel le locuteur collectif pose les postulats à partir desquels l'argumentation s'ordonnera (les SN en question sont dans la résolution considérée : *le syndicat, la transition vers le socialisme autogestionnaire, la prise du pouvoir, les luttes sociales, les sociétés multinationales, l'unité d'action, l'union des forces populaires*) ;

c. que les phrases copulatives d'autres types ressortissent d'actes d'énonciation différents (les définitions avec deux indéfinis posent la définition dans l'ordre du possible, de l'éventuel) ;

d. que les structures de SN entrant en jeu dans ces phrases ne sont pas indéterminées dès lors que l'on raisonne sur une base probabiliste : ainsi, il y a une corrélation entre :

— les définitions exclusives et un N2 de forme *N3 de SN* (où N3 définit une classe sémantique à laquelle N2 appartient) ;

— les définitions non exclusives et un SN de forme *N3 ADJ de SN4* où ADJ marque l'emphase (et où N3 est défini comme précédemment par rapport à N1) (il y a donc corrélation entre cette structure et l'emphase) ;

— les définitions exclusives bilatérales et l'emploi d'un infinitif en position de N2.

Autrement dit, l'écriture du texte se marque par l'emploi significatif de sous-structures limitées parmi l'ensemble de celles qui sont théoriquement possibles. Seule une grammaire probabiliste (qui s'oppose donc à la conception en termes binaires de la grammaticalité dans le cadre des grammaires génératives) peut décrire cette opération de limitation des sous-structures.

CROISEMENT D'UNE SÉRIE DE FACTEURS

Le codage qui a été présenté ci-dessus permet l'étude des corrélations entre un assez grand nombre de facteurs. L'exemple exposé ci-dessous consiste en l'analyse des corrélations entre 8 facteurs, correspondant au total à 49 modalités (les modalités représentent ici l'ensemble des réalisations des différents facteurs). Le programme utilisé est le programme MULTEM, qui provient de la bibliothèque du CESIA, présentée dans Lebart, Morineau et Tabard (1977). L'analyse ne porte que sur les chaînes marquées « 1 » dans la colonne 7 (= les chaînes majeures).

Analyse factorielle

Ces facteurs et ces modalités sont les suivants :

1. Ordre: ce facteur caractérise les différentes parties du texte:
 - OR1: chaînes 001 à 084: «la crise de la société capitaliste, contexte de notre action»;
 - OR2: chaînes 085 à 221: «La stratégie de la CFDT»;
 - OR3: chaînes 222 à 329: «Des politiques d'action appropriées»;
 - OR4: chaînes 330 à 424: «Renforcer l'organisation pour améliorer le rapport de forces».
2. Classe du nom N1 (sujet)
 - NNC: noms non classés;
 - NM 1 à 9: cf. les classes de noms répertoriées en annexe.
3. Positif/Négatif: forme du verbe V1
 - P+ : positif;
 - P- : négatif.
4. Actif/Passif: forme du verbe V1
 - ACT: actif;
 - PAS: passif;
 - SEM: «Se-moyen».
5. Contrôle (rapport N1/V1)
 - C ++ : verbe (+ contrôlable); sujet animé;
 - C +- : verbe (- contrôlable); sujet animé;
 - C - - : verbe (- contrôlable); sujet non animé.

6. Classe du verbe V1

- VNC: verbe non classé;
- PRF: performatif (classe 01 à 03);
- DEV: obligation (classe 10);
- IMP: implicatif (classes 13 à 15);
- POU: possibilité (opérateur de phrase) (classe 20);
- PRM: possibilité (prédicat à deux places) (classes 21);
- DF: définitoire (classes 30 à 34);
- ETA: analytique (classes 44 à 45);
- TPS: temps (classes 55-56);
- PS1: psychologique (tension positive) (classe 60);
- PS2: psychologique (divers) (classes 61 à 63);
- CTR: contrôlable (= verbes actifs divers) (classes 80-88).

7. Structure modale. On a :

- 11 D, 12 D, 13 D, 21 D, 22 D, 23 D, et 55 D [où les valeurs de 1, 2, 3 et 5 correspondent à la nature et à l'interprétation des déterminants de, respectivement, N1 (premier chiffre) et N2 (second chiffre)];
- 124 correspondant à la somme des occurrences de 14 et de 24 (définis comme précédemment);
- F4D correspondant à la somme des occurrences dans lesquelles le déterminant de N1 est codé 4;
- NND: autres cas.

8. Préposition introduisant N2. On a A, DE, PAR, et :

- ABS: absence de préposition;
- PR5: autres prépositions.

L'ensemble des croisements entre ces diverses données est représenté par le tableau 1.

Tableau 1

	NR1	NR2	NR3	NR4	NVC	NM1	NM2	NM3	NM4	NM5	NM6	NM8	PA	P	ACT	PAS	SEM	C++	C-	C
NR1	108.																			
NR2	0.	167.																		
NR3	0.	0.	138.																	
NR4	0.	0.	0.	140.																
NVC	61.	118.	73.	83.	331.															
NM1	10.	13.	30.	31.	0.	84.														
NM2	1.	8.	1.	17.	0.	0.	25.													
NM3	4.	6.	10.	1.	0.	0.	0.	21.												
NM4	6.	10.	15.	5.	0.	0.	0.	0.	36.											
NM5	3.	12.	4.	1.	0.	0.	0.	0.	0.	20.										
NM6	22.	4.	4.	0.	0.	0.	0.	0.	0.	0.	30.									
NM8	1.	2.	1.	2.	0.	0.	0.	0.	0.	0.	0.	6.								
PA	96.	143.	130.	132.	295.	84.	21.	18.	35.	17.	26.	5.	501.							
P	12.	24.	8.	8.	36.	0.	4.	3.	1.	3.	4.	1.	0.	52.						
ACT	50.	143.	124.	124.	268.	80.	23.	19.	33.	15.	27.	6.	429.	42.	471.					
PAS	14.	13.	6.	9.	34.	0.	1.	2.	2.	2.	1.	0.	35.	7.	0.	42.				
SEM	14.	11.	8.	7.	29.	0.	1.	0.	1.	3.	2.	0.	37.	3.	0.	40.				
C++	36.	33.	40.	59.	61.	75.	10.	11.	6.	0.	23.	2.	178.	10.	173.	3.	12.	188.		
C-	40.	7.	6.	4.	11.	3.	3.	6.	0.	0.	4.	0.	24.	3.	24.	4.	0.	0.	27.	
C	62.	127.	72.	77.	259.	6.	12.	4.	30.	20.	3.	4.	299.	39.	275.	35.	28.	0.	0.	338.
VNC	0.	1.	1.	9.	10.	0.	0.	1.	0.	0.	0.	0.	10.	1.	9.	1.	1.	0.	9.	
DEV	4.	21.	20.	31.	44.	6.	10.	3.	8.	3.	0.	2.	65.	11.	73.	1.	2.	8.	6.	62.
IMP	10.	27.	12.	12.	51.	0.	1.	0.	3.	6.	0.	0.	59.	2.	58.	3.	2.	0.	1.	85.
PBU	2.	11.	8.	5.	19.	0.	0.	3.	1.	1.	2.	0.	18.	8.	26.	0.	0.	3.	3.	18.
PRM	3.	9.	5.	3.	10.	1.	1.	0.	5.	2.	0.	1.	20.	0.	19.	1.	0.	8.	0.	12.
DF	8.	34.	18.	13.	52.	2.	1.	2.	5.	3.	4.	1.	66.	7.	73.	0.	0.	3.	5.	65.
ETA	22.	14.	17.	9.	48.	1.	3.	2.	4.	1.	3.	0.	59.	3.	68.	9.	7.	8.	9.	88.
TPS	10.	4.	8.	3.	11.	8.	9.	0.	1.	3.	3.	1.	24.	1.	19.	2.	4.	13.	0.	10.
PS1	4.	3.	8.	5.	9.	6.	0.	1.	0.	0.	4.	0.	17.	3.	19.	0.	1.	15.	0.	6.
PS2	4.	5.	3.	4.	9.	4.	1.	0.	0.	0.	2.	0.	13.	3.	13.	3.	0.	12.	2.	2.
CTR	22.	26.	16.	23.	45.	16.	6.	3.	4.	3.	10.	0.	78.	9.	83.	16.	18.	48.	1.	38.
ABS	43.	62.	47.	52.	109.	40.	10.	7.	14.	6.	16.	2.	192.	11.	202.	0.	2.	87.	3.	408.
DE	1.	2.	3.	0.	3.	0.	0.	0.	2.	0.	1.	0.	6.	0.	9.	0.	2.	2.	1.	3.
A	4.	10.	7.	7.	20.	0.	2.	0.	1.	2.	0.	0.	25.	1.	24.	5.	0.	10.	1.	16.
PAR	4.	3.	2.	3.	12.	0.	0.	2.	2.	1.	0.	0.	15.	2.	2.	12.	3.	2.	2.	13.
PR5	53.	93.	76.	78.	187.	40.	13.	12.	17.	11.	13.	4.	209.	38.	229.	25.	33.	87.	18.	195.
11D	17.	3.	20.	26.	32.	19.	4.	0.	0.	3.	0.	0.	64.	2.	58.	7.	1.	22.	5.	89.
12D	0.	2.	4.	1.	1.	0.	0.	1.	0.	1.	1.	1.	8.	1.	8.	1.	0.	0.	1.	1.
13D	8.	3.	7.	7.	19.	5.	2.	0.	2.	0.	2.	0.	24.	3.	23.	1.	1.	3.	2.	18.
21D	1.	10.	3.	0.	10.	0.	0.	0.	2.	1.	1.	0.	13.	1.	14.	1.	0.	3.	2.	9.
22D	1.	13.	1.	1.	10.	0.	2.	0.	2.	2.	0.	0.	15.	1.	15.	1.	0.	1.	0.	15.
23D	0.	5.	4.	1.	6.	0.	1.	0.	2.	0.	1.	0.	9.	1.	10.	0.	0.	0.	0.	10.
12A	1.	4.	0.	3.	8.	0.	0.	1.	0.	1.	0.	0.	8.	0.	8.	2.	0.	1.	0.	7.
F4D	3.	7.	1.	5.	12.	0.	1.	0.	0.	2.	0.	1.	15.	1.	15.	0.	3.	9.	0.	11.
55D	3.	5.	6.	5.	12.	4.	2.	0.	1.	0.	0.	0.	18.	1.	18.	2.	1.	7.	0.	12.
NND	74.	117.	92.	91.	232.	55.	13.	15.	19.	14.	22.	4.	331.	42.	314.	27.	34.	126.	18.	230.
VNC	11.																			
DEV	0.	76.																		
IMP	0.	0.	76.																	
PBU	0.	0.	0.	61.																
PRM	0.	0.	0.	0.	20.															
DF	0.	0.	0.	0.	0.	73.														
ETA	0.	0.	0.	0.	0.	0.	62.													
TPS	0.	0.	0.	0.	0.	0.	25.													
PS1	0.	0.	0.	0.	0.	0.	0.	20.												
PS2	0.	0.	0.	0.	0.	0.	0.	0.	16.											
CTR	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	87.							
ABS	1.	38.	0.	36.	0.	3.	56.	18.	12.	1.	6.	33.	204.							
DE	0.	0.	0.	1.	0.	0.	0.	4.	0.	0.	0.	1.	0.	6.						
A	4.	7.	1.	3.	0.	6.	2.	9.	0.	0.	1.	4.	0.	0.	29.					
PAR	0.	1.	0.	2.	0.	0.	0.	6.	0.	0.	0.	8.	0.	0.						
PR5	10.	34.	75.	19.	26.	11.	15.	25.	13.	19.	9.	41.	0.	0.	0.	17.	0.	0.	297.	
11D	1.	14.	0.	7.	0.	1.	5.	15.	3.	1.	1.	14.	48.	3.	8.	5.	2.			
12D	0.	1.	0.	0.	0.	0.	0.	0.	1.	0.	1.	2.	4.	0.	0.	1.	0.			
13D	0.	4.	0.	3.	0.	1.	9.	3.	0.	0.	0.	5.	21.	0.	0.	1.	3.			
21D	0.	1.	0.	0.	0.	0.	0.	0.	2.	0.	0.	2.	11.	0.	2.	1.	2.			
22D	0.	1.	0.	3.	0.	1.	8.	1.	0.	0.	1.	1.	13.	0.	1.	1.	1.			
23D	0.	0.	0.	1.	0.	0.	8.	0.	0.	0.	0.	1.	8.	0.	1.	0.	1.			
12A	0.	0.	0.	3.	0.	0.	2.	0.	0.	0.	0.	3.	6.	0.	1.	1.	0.			
F4D	0.	0.	0.	2.	0.	0.	2.	3.	1.	0.	0.	3.	4.	0.	2.	0.	10.			
55D	0.	3.	3.	5.	0.	1.	2.	0.	0.	1.	4.	9.	0.	2.	0.	8.				
NND	10.	44.	71.	34.	26.	13.	35.	36.	18.	19.	12.	52.	80.	3.	13.	6.	272.			

Tableau 2

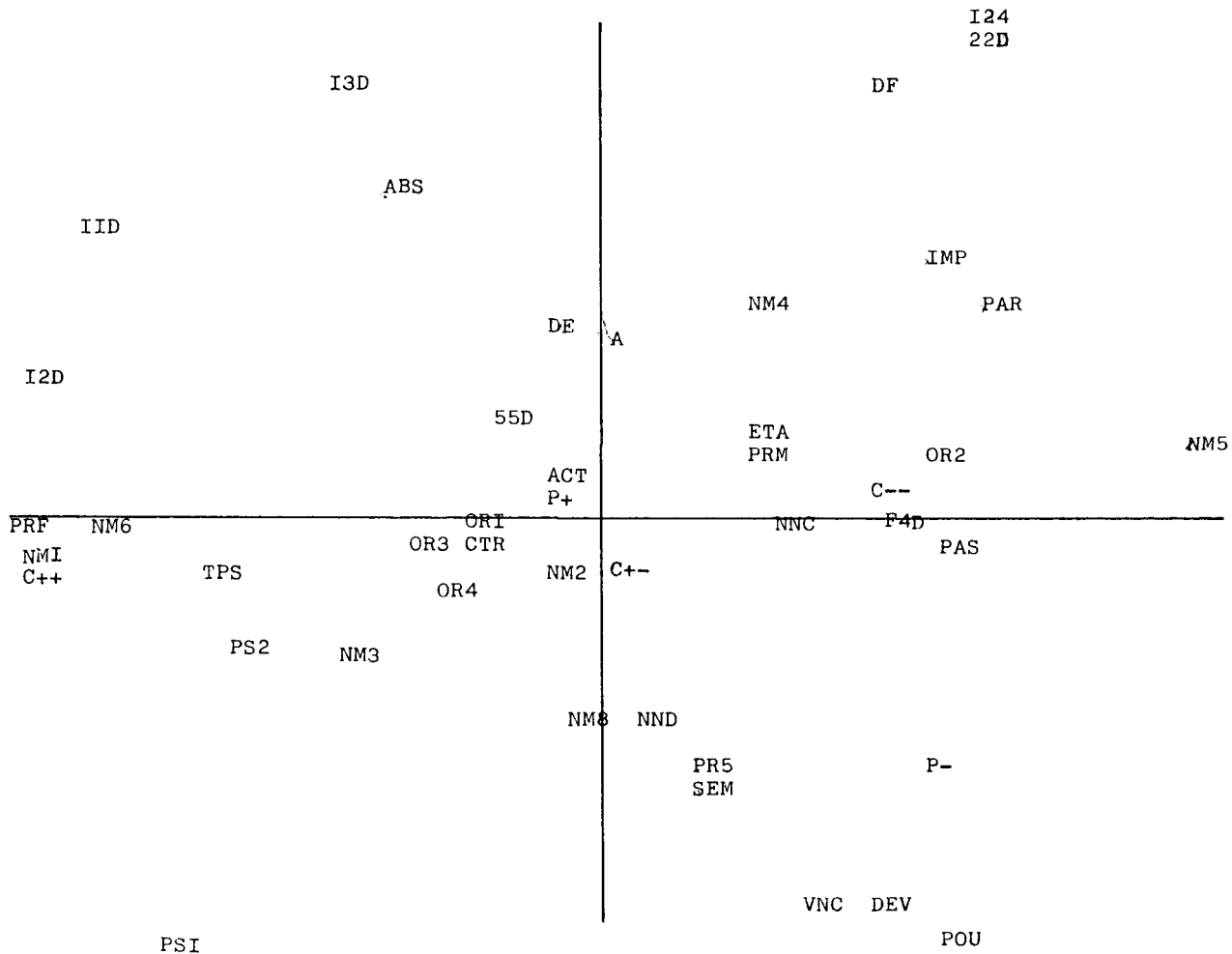


Tableau 3

		VNC		A		
	NM2		55D			
			FAD		I24	
		PRM OR4				
					, IID	
	DEV	NMI	IMP PRF			PAS PAR
			NM5 C--P+	NNC	CTR	SEM ETA
	NM8	AET PR5		C++		
	22D	OR3 NND				
		NM4 ABS OR2				DE
			I3D	PS2		
	DF			21D	ORI	
				TPS		I2D
				P-		
23D		PSI			NM3	
					NM6	

Les résultats de l'analyse factorielle établie à partir de ces données sont indiqués sur les tableaux 2 et 3 qui représentent les projections sur les quatre axes principaux. Le tableau 2 représente la projection sur l'axe 1 (horizontal; valeur propre: 6,23 %) et l'axe 2 (vertical; valeur propre: 5,34 %) et le tableau 3 représente la projection sur l'axe 3 (horizontal; valeur propre: 4,82 %) et l'axe 4 (vertical; valeur propre: 4,08 %).

Interprétation

L'examen des chiffres du tableau 1 et des oppositions entre les modalités les mieux représentées sur les différents axes permet d'observer les faits suivants :

Relations N1/V1

Il y a une corrélation entre la classe des verbes performatifs (PRF) et la classe NM1 (*la CFDT, le congrès*) et aussi, dans une moindre mesure, la classe NM3 (*les travailleurs, etc.*). Ces deux classes constituent les deux *énonciateurs* posés comme tels dans le discours. Ni la classe NM2 (les «composantes» de la classe NM1) ni la classe NM6 (les adversaires) ne possèdent ce statut. C'est la confédération qui parle, et à travers elle l'ensemble de ceux qu'elle entend représenter. En revanche la classe NM2 est significativement corrélée au verbe *devoir* (et à la classe DEV): le congrès définit les obligations des individus et des organisations membres de la confédération.

La classe NM4 (*action (s), lutte (s), combat*) est corrélée avec la classe DEV mais aussi avec la classe PRM: les *luttés* (cette forme apparaît souvent en position de sujet) sont le seul «moteur» qui permette le changement. Les termes abstraits au projet et à la transformation sociale (classe NM5) sont corrélés avec les verbes implicatifs (classe IMP). Enfin, les classes NM1 et NM6 sont l'une et l'autre corrélées avec les classes TPS, PS1 et CTR: les classes de verbes actifs et les classes de verbes marquant la volition (auxquels on peut ajouter les verbes de la classe TPS lorsqu'ils sont marqués (+ *contrôlable*)) sont donc significativement cooccurrents avec les deux séries nominales qui indiquent les forces qui, selon la Résolution, s'opposent le plus. On peut en conclure que la fonction d'agent est assumée de manière privilégiée par la CFDT d'une part et ses adversaires (*patronat, gouvernement, capitalisme*) d'autre part — et non par les classes NM2, NM3 ou NM4.

Relations N1/contrôle

L'étude des corrélations entre les diverses classes de noms N1 et la propriété « contrôle » vient largement corroborer les remarques précédentes.

En effet, la propriété C++ va (significativement) de pair avec les classes NM1 et NM6 et elles seules. En revanche, la classe NM3 est corrélée avec la propriété C+-, et les classes NM2, NM4 et NM5 avec la propriété C--.

Relations N1/actif-passif

Toutes les classes de N1 sont corrélées avec ACT. Seule la classe NM5 apparaît relativement souvent en cooccurrence avec la propriété « SE-MOYEN ».

Relations N1/positif-négatif

Si toutes les classes de N1 sont corrélées avec la forme positive, il est significatif que la forme négative soit plus nettement exclue encore pour les classes NM1 et NM4 qui ne peuvent pratiquement jamais introduire une phrase négative. La CFDT comme les luttes, en tant que « moteurs » du changement social, n'induisent aucun processus de forme négative.

Relations V1/structure modale

Les classes PRF et ETA sont corrélées avec la structure 11 D : les phrases descriptives ou analytiques ont largement recours aux définis spécifiques.

Ces phrases s'opposent aux définitives dont le verbe (de la classe DF) définit une structure énonciative qui est corrélée avec des déterminants définis non spécifiques ou génériques. Elles s'opposent également aux phrases dont le verbe appartenant aux classes DEV, IMP ou PRM constitue l'un des champs permettant les interprétations non spécifiques ou génériques des indéfinis, ou la présence d'un déterminant générique. Le cas de la classe CTR est particulier puisque cette classe va de pair avec, à la fois, des définis spécifiques et des indéfinis non spécifiques ou génériques. Cela s'explique probablement par le fait que cette classe est la seule qui ne corresponde pas de manière univoque à l'un des types énonciatifs exprimés par les modalités de la colonne 6 (dans tous les autres cas, des analyses similaires donnent des corrélations très fortes entre ces modalités de la colonne 6 et la classe

du verbe, ce qui n'est pas étonnant puisque c'est essentiellement en référence au verbe V1 qu'on a défini les types énonciatifs).

Relations ordre/V1

La seconde partie (OR2) (traitant de la stratégie) s'oppose significativement aux trois autres par des corrélations avec les classes POU, IMP et DF qu'on ne trouve pas dans les autres cas. Les classes PRF et ETA caractérisent la première partie (OR1) (qui traite de l'analyse); les classes TPS et PS1 (volition) caractérisent la troisième partie (OR3) (qui traite de l'action) cependant que cette même classe PS1 et la classe DEV caractérisent la quatrième (OR4) (qui traite de l'organisation, et impose donc de manière significative des obligations aux instances et aux membres de l'organisation).

Relations ordre/structure modale

Ce qui a été dit sur les corrélations entre classes de verbe V1 et structures modales permet de prédire correctement les relations entre ce dernier critère et l'ordre. OR2 est corrélé avec 21 D, 22 D et F4 D, 124, 55 D; OR1 et OR3 sont corrélés avec 11 D (et, dans le cas de OR3, 13 D); OR4 est corrélé avec 130, F4 D et 55 D. Le fait que les résultats corroborent les prédictions est un indice de la validité de la méthode.

Relations ordre/contrôle

La première partie exprime une analyse; la seconde des principes stratégiques; la troisième et la quatrième sont tournées vers l'action.

Il est donc cohérent que cette fois-ci les parties s'opposent deux à deux: OR1 et OR2 sont corrélés avec C+- et C--; et OR3 et OR4 sont corrélés avec C++.

Relations ordre/N1

Il est logique que la classe NM6 (les adversaires) soit corrélée avec OR1 (l'analyse), que la classe NM5 (le changement social) soit corrélée avec OR2 (la stratégie), que NM1, NM3 et NM4 soient corrélées avec OR3 et enfin que NM1 et NM2 le soient avec OR4.

— Les tableaux 4 et 5 permettent de représenter ces relations et ces oppositions qui sont également mises en lumière par l'analyse des tableaux 2 et 3 — analyse dans laquelle on

Tableau 4

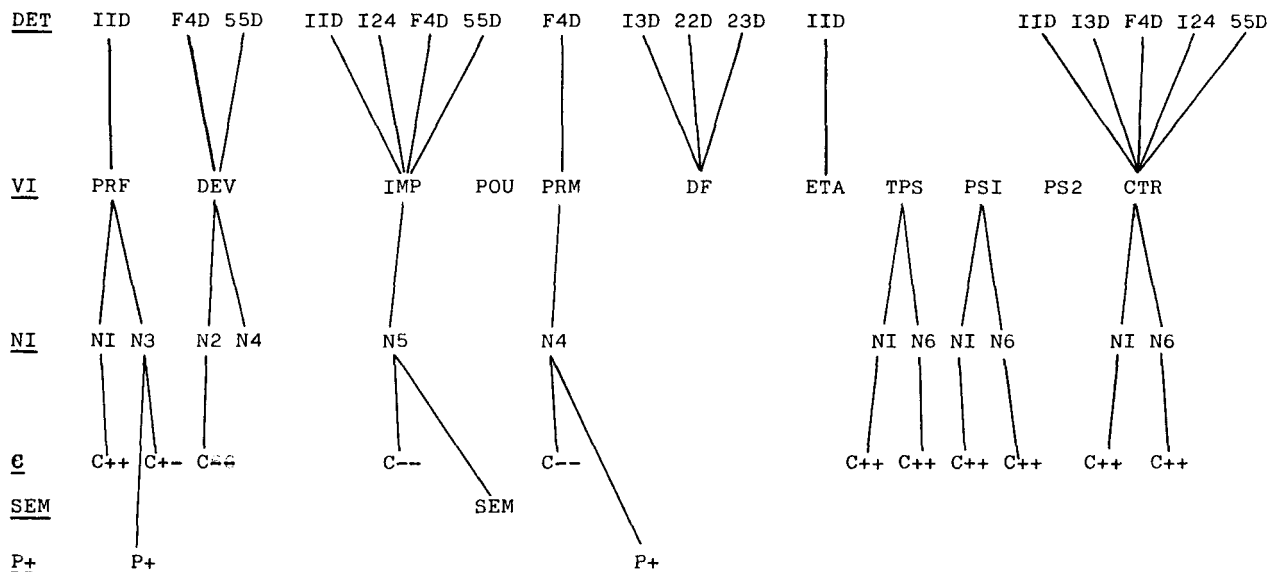
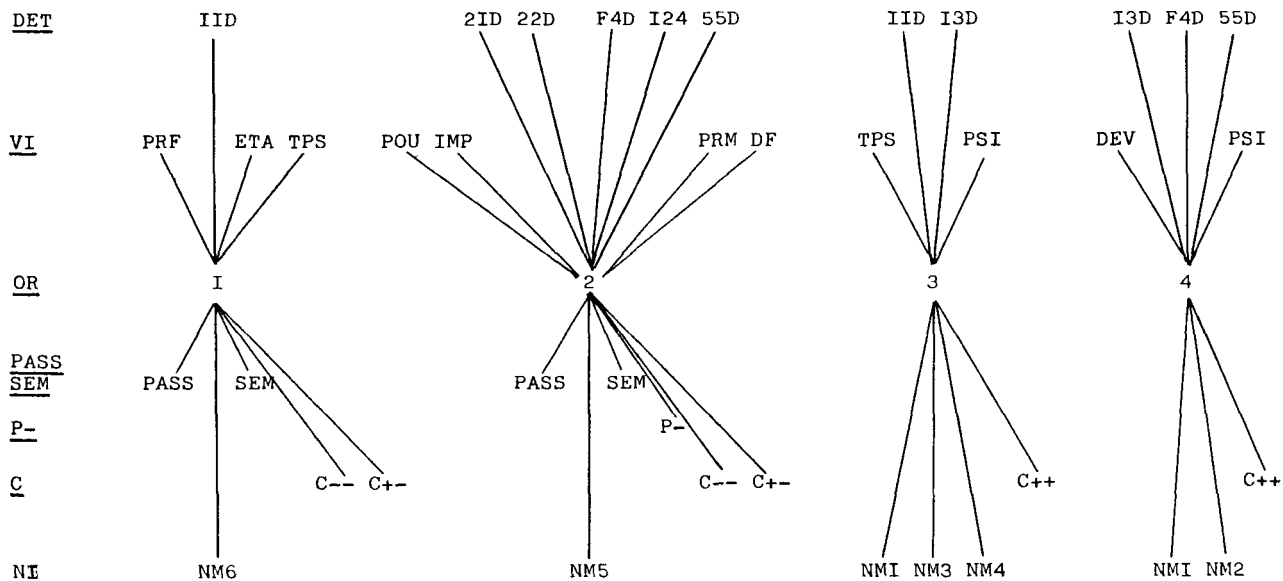


Tableau 5



ne considère pour chaque axe que les modalités dont la contribution à la définition de l'axe est suffisamment importante¹⁵.

— L'axe 1 oppose clairement OR2 d'une part (et par conséquent C--, PASS, F4 D, 124, 22 D, DEV, POU) et, d'autre part, OR1, OR3 et OR4 (et par conséquent C++, 11 D, TPS, PRF, PS1, PS2, NM1, NM6).

— L'axe 2 oppose les verbes modaux (DEV, POU) et la série des déterminants qui en sont antinomiques (dans le corpus), c'est-à-dire essentiellement les définis.

— L'axe 3 oppose les verbes DEV et DF d'une part et d'autre part, les verbes ETA et OR1 (modalités corrélées).

— L'axe 4 oppose OR4 à OR1. Cette opposition est particulièrement intéressante. En effet, la partie OR2 n'étant que peu représentée sur cet axe, on observe que OR1 (et OR3) s'opposent à OR4 de la même manière que, sur l'axe 1, OR1, OR3 et OR4 s'opposaient à OR2. Cela s'explique sans doute partiellement par la présence à la fin de OR4 d'une conclusion qui reprend la thématique de OR2. On retrouve ainsi du côté de OR4: PRM, IMP, DEV; PRF, SSD, F4 D, 124, et du côté de OR1, 12 D; 21 D, C+-, etc. Les modalités qui vont de pair avec OR2 sur l'axe 1 et OR4 sur l'axe 4 définissent un type d'énonciation « législatif-possibiliste » qui s'oppose au type analytique de OR1 et au type « législatif-volitionnel » qui est commun à OR3 et OR4 (sauf sur l'axe 4). Le double statut de OR4 montre que ces oppositions typologiques sont essentiellement relatives, et n'existent pas indépendamment des oppositions définies à l'intérieur d'un corpus ou entre des corpus.

Relations V1/V2

L'étude des corrélations entre les classes de V1 et de V2 a été menée selon la même méthode. Les classes retenues n'ont pas été les mêmes pour V1 et V2, puisque la distribution des verbes apparaissant dans ces deux positions présente de notables différences (le tableau 6 représente une projection des verbes les plus fréquents — indiqués selon la liste reproduite

15. Le facteur « préposition » ne jouant pas de rôle décisif dans le réseau des relations entre les précédents facteurs n'a pas été inclus dans le commentaire. (Il a néanmoins été mentionné dans les tableaux puisqu'il a effectivement été intégré dans les calculs). Ce facteur joue un rôle lorsqu'on le croise avec d'autres configurations de facteurs (incluant N2, par exemple) et lorsqu'on l'insère dans l'ensemble des marques qui définissent les cadres syntaxiques.

Tableau 7

PRM	V72	IMP ETA	
DFI		PSI PS2 CT2	
PRF	V82 V52	TP2 DEV	CTR IP2 772
POU		ET2	PO2 PSY
PR2		DF2	
TPI			

en annexe — en fonction du nombre de leurs occurrences en position de V1 (modalité V1) ou de V2 (modalité INF).

On a donc les classes suivantes (le nombre d'occurrences dans les chaînes majeures est indiqué entre parenthèses):

V1	V2
PRF: cl. 01-03 (6)	PR2: cl. 01-03 (13)
DEV: cl. 10 (73)	IP2: cl. 11-14 (6)
IMP: cl. 13-15 (7)	PO2: cl. 20-21 (24)
POU: cl. 20 (24)	DF2: cl. 30-32 (11)
PRM: cl. 21 (19)	ET2: cl. 40-45 (16)
DF1: cl. 30-34 (9)	V52: cl. 50-52 (12)
ETA: cl. 44-45 (7)	TP2: cl. 56 (6)
TP1: cl. 55-56 (6)	V72: cl. 72-75 (16)
PS1: cl. 60 (19)	772: cl. 76-77 (5)
PS2: cl. 61-63 (7)	V82: cl. 80-81 (13)
CTR: cl. 80-88 (4)	CT2: cl. 87-88 (59)

On constate qu'il y a une corrélation entre:

— la position V1 et les classes DEV, POU, PRM, PS1, DF;

— la position V2 et les classes V72, 772, V82, CT2;

c'est-à-dire l'ensemble des verbes situés dans le tableau de l'annexe dans les classes 70-89.

— Les tableaux 7 et 8 représentent les projections des diverses modalités: le tableau 7, les projections sur l'axe 1 (horizontal; valeur propre: 29,21%) et l'axe 2 (vertical; valeur propre: 26,01%); le tableau 8 les projections sur l'axe 3 (horizontal; valeur propre: 9,68%) et l'axe 4 (vertical; valeur propre: 8,73%).

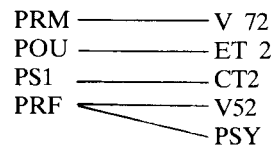
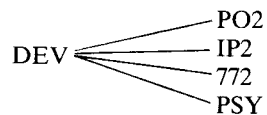
— Sur l'axe 1 se rapprochent notamment DEV d'une part et IP2, PO2, 772, PSY d'autre part. Sur le même axe s'opposent d'un côté TP1 et DF1 et de l'autre PSY, 772, IP2, PO2.

— Sur l'axe 2 se rapprochent PRM et V72 ainsi que PS1 et CT2, et aussi TP1 et PR2. Sur cet axe PRM, IMP et ETA s'opposent à DF2 et ET2 — et TP1 s'oppose à V72.

— Sur l'axe 3 PS2 et V82 se rapprochent ainsi que PRF d'une part et V52 et PSY d'autre part. Sur cet axe PS2 s'oppose à PSY et PRF à IP2 et à V82.

— Sur l'axe 4 PS2 et CTR s'opposent à TP2.

En conclusion, on observe principalement les relations suivantes :



Deux regroupements de verbes V1 s'opposent :

— les verbes TP1/PS1, PRF, PRM qui introduisent les verbes marqués (+ contrôlables) des classes 70-89 ;

— les verbes POU qui introduisent la classe ETA et qui, avec d'autres verbes CDF, IMP, etc.), apparaissent principalement dans les énoncés d'analyse.

La classe DEV relève à la fois de ces deux regroupements selon les contextes. Ainsi, il est significatif que les verbes exprimant un processus (classes 71-78) soient corrélés avec PRM s'il s'agit d'un processus positif et avec DEV s'il s'agit d'un processus négatif. On a « ... *permet de construire...* » et « ... *doit mettre fin...* ».

* * *

Ces analyses n'ont porté que sur ce que j'ai appelé la « grammaire des fréquences ». L'étude de la « grammaire des séquences » permettrait d'observer comment les diverses propriétés se répartissent par rapport à la structure linéaire du texte, et de relever notamment les discordances entre les types énonciatifs des phrases et les quatre parties de la résolution qui semblent pourtant chacune correspondre à un type énonciatif privilégié. Cette étude permettrait également de montrer à quel point les textes de résolution s'opposent à d'autres types de discours (par exemple les récits) pour ce qui est de la répartition des déterminants, et donc des structures modales, et aussi des anaphores.

Ainsi, on ne passe pas de l'indéfini aux définis pour la première occurrence d'un nom et celles qui suivent selon la logique du « fléchage », telle qu'elle a été, notamment, décrite par Fuchs et Pécheux (1970). Mais chacun de ces deux types de déterminant semble plutôt lié à certains types de SN et à certaines places syntaxiques (la position N1 étant corrélée avec les définis, et les indéfinis étant plus fréquents en position N2 ou à l'intérieur des compléments de but, etc.). Notons que le codage effectué doit permettre de décrire ces faits sur la base de corrélations entre les propriétés des cartes successives considérées deux à deux (comparaison entre les modalités de la carte n et de la carte $n + 1$, puis de la carte $n + 1$ et de la carte $n + 2$, etc.).

D'autre part, l'application de la même méthode à d'autres corpus permettra d'effectuer les analyses contrastives qui seront nécessaires à l'élaboration d'une typologie des discours. Cette typologie pourra opposer des ensembles de modalités corrélées (lexicales, syntaxiques, énonciatives), chaque ensemble définissant un type énonciatif et/ou un locuteur. Il semble en tout cas que l'étude statistique de ces trois sortes de régularités qui constituent le discours, et donc la mise en évidence des connexions entre les unes et les autres sur une base probabiliste, constitue une étape, indispensable, vers la constitution d'une grammaire de discours.

mars 1982

BIBLIOGRAPHIE

- BELLERT I., 1970, «On a condition of the coherence of texts», *Semiotica*, 2 (4), p. 335-363.
- BENVENISTE E., 1965, *Problèmes de linguistique générale*, Paris, Gallimard.
- CHOMSKY N., 1972, *Studies on semantics in generative grammar*, La Haye, Mouton ; Paris, Le Seuil, 1974.
- CHOMSKY N., 1980, *Essais sur la forme et le sens*, Paris, Le Seuil.
- DUBOIS J., 1967, *Grammaire structurale du français: le verbe*, Paris, Larousse.
- FISCHER S.D. et MARSHALL B.A., 1969, *The examination and abandonment of the theory of begin of D.M. Perlmutter*, Cambridge, MIT Press, inédit.
- FUCHS C. et PECHEUX M., 1970, «Lexis et méta-lexis. Applications au problème des déterminants», *Considérations théoriques à propos du traitement formel du langage*, *Documents de linguistique quantitative*, 7, Paris, Dunod.
- GEFFROY A., LAFON P. et TOURNIER M., 1973, «L'indexation minimale», *Actes du Colloque sur l'analyse des corpus linguistiques*, Université de Strasbourg, repris dans TOURNIER (1975), vol. 3, p. 120-157.
- GIRARD L., SERMAN W., CADET E., GOSSEZ R., 1976, *La Chambre des députés en 1837-1839*, Paris, Publications de la Sorbonne.
- GROSS M., 1975, *Méthodes en syntaxe*, Paris, Hermann.
- GROSS M., 1976, Présentation de BOONS, GUILLET et LECLERE, *La Structure des phrases simples en français*, Genève, Droz.
- GUESPIN L., 1975, *L'analyse du discours: problèmes et perspectives*, Paris, Eds de la Nouvelle Critique.
- GUILHAUMOU J., 1977, «Sur les discours politiques contemporains: orientations actuelles en analyse de discours», photocopié à l'ENS de Saint-Cloud.
- JACKENDOFF R.S., 1972, *Semantic interpretation in generative grammar*, Cambridge, MIT Press.
- JAKOBSON R., 1963, *Essais de linguistique générale*, Paris, Minuit.
- KATZ J.J. et FODOR J.A., 1963, «The structure of a semantic theory», *Language*, 39, p. 170-210.
- KATZ J.J. et POSTAL P.M., 1964, *An integrated theory of linguistic description*, Cambridge, MIT Press.
- LAFON P. et TOURNIER M., 1978, «Une nouvelle approche lexicométrique des cooccurrences dans un texte», communication au colloque de l'ALLC, Birmingham, 6 avril 1978, repris dans *Travaux de lexicométrie et de lexicologie politique*, 3, p. 135-147.
- LEBART, MORINEAU et TABARD, 1977, *Techniques de la statistique descriptive*, Paris, Dunod.
- MAINGUENEAU D., 1976, *Initiation aux méthodes de l'analyse de discours*, Paris, Hachette.

- MALDIDIER D., 1971 a, «Le discours politique de la guerre d'Algérie: approche synchronique et diachronique», *Langages*, 23, p. 57-86.
- MALDIDIER D., 1971 b, «Lecture des discours de de Gaulle par six quotidiens parisiens: 13 mai 1958», *Langue française*, 9, p. 34-46.
- MARCELLESI J.B., 1971, *Le congrès de Tours. Etudes sociolinguistiques*, Paris, Le Pavillon, R. Martin.
- MARCELLESI J.B., 1976, *L'individuation linguistique des groupes politiques; communistes et socialistes de 1920 à 1925; analyse de discours à entrée lexicale*, Thèse de doctorat d'Etat, Université de Paris X-Nanterre.
- MARCUS S., 1978, *La sémantique formelle du folklore*, Paris, Klincksieck.
- MULLER C., 1968, *Initiation à la statistique linguistique*, Paris, Larousse.
- MULLER C., 1977, *Principes et méthodes de statistique lexicale*, Paris, Hachette.
- MULLER C., 1979, *Langue française et linguistique quantitative*, Genève, Slatktine.
- PÊCHEUX M., 1969, *L'Analyse automatique de discours*, Paris, Dunod.
- PÊCHEUX M. et FUCHS C., 1975, «Mises au point et perspectives à propos de l'analyse automatique de discours», *Langages*, 37, p. 7-80.
- PROST A., 1974, *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*, Paris, PUF.
- ROBIN R., 1973, *Histoire et linguistique*, Paris, A. Colin.
- RUWET N., 1967, *Introduction à la grammaire générative*, Paris, Plon.
- RUWET N., 1972, *Théorie syntaxique et syntaxe du français*, Paris, Le Seuil.
- SALKOFF M., 1973, *Une grammaire en chaînes du français; analyse distributionnelle*, Paris, Dunod.
- SIMONIN-GRUMBACH J., 1975, «Pour une typologie des discours», dans: *Langue, discours, société (pour Emile Benveniste)* Paris, Le Seuil, p. 85-121.
- SUEUR J.-P., 1977, «A propos des restrictions de sélection: les infinitifs *devoir* et *pouvoir*», *Linguisticae investigationes*, 1 (2), p. 375-410.
- SUEUR J.-P., 1978, «Sur la syntaxe du récit oral», inédit.
- SUEUR J.-P., 1980, «Syntaxe, lexique et énonciation dans quelques résolutions syndicales», communication au colloque Lexicologie politique du français moderne, Saint-Cloud, 15-20 septembre 1980.
- TOURNIER M., 1975, *Un vocabulaire ouvrier en 1848, essai de lexicométrie*, Thèse de doctorat d'Etat, ENS de Saint-Cloud.
- VAN DIJK T.A., 1972, *Some aspects of text grammars*, La Haye, Mouton.

ANNEXE 1

Liste des noms codés

(Le premier nombre indique le numéro de code, le second la fréquence)

1. travailleurs 124	30. type 24	59. syndicats 14	88. besoins 10
2. action 112	31. pratique 23	60. tiers-monde 14	89. force 10
3. CFDT 101	32. droit 22	61. adhérents 13	90. cadre 10
4. travail 74	33. emploi 21	62. exploitation 13	91. conscience 10
5. luttes 63	34. formation 21	63. maintien 13	92. élaboration 10
6. développement 61	35. cause 20	64. perspectives 13	93. expression 10
7. lutte 57	36. entreprise 20	65. temps 13	94. gouvernement 10
8. société 56	37. problème 20	66. base 12	95. intérêts 10
9. classe 54	38. production 20	67. charge 12	96. militants 10
10. conditions 47	39. état 20	68. confédération 12	97. participation 10
11. forces 46	40. prise 19	69. fait 12	98. partis 10
12. organisation 42	41. rôle 19	70. femmes 12	99. positions 10
13. organisations 38	42. formes 18	71. immigrés 12	100. renforcement 10
14. pays 38	43. syndicat 18	72. responsabilité 12	101. sections 10
15. transformation 38	44. droits 17	73. salaire 12	102. suppression 10
16. rapports 35	45. union 17	74. salariés 12	103. activité 9
17. revendications 35	46. peuples 16	75. secteurs 12	104. amélioration 9
18. moyens 34	47. rapport 16	76. unions 12	105. but 9
19. objectifs 34	48. contrôle 15	77. capacité 11	106. cas 9
20. crise 33	49. domination 15	78. construction 11	107. catégories 9
21. socialisme 32	50. information 15	79. groupes 11	108. chômage 9
22. congrès 31	51. libertés 15	80. indépendance 11	109. compte 9
23. masse 30	52. actions 14	81. part 11	110. contestation 9
24. situation 30	53. division 14	82. période 11	112. durée 9
25. structures 30	54. entreprise 14	83. processus 11	113. fois 9
26. vie 30	55. Europe 14	84. réduction 11	114. institutions 9
27. capitalisme 26	56. fonction 14	85. répression 11	115. investissements 9
28. stratégie 25	57. mouvement 14	86. syndicalisme 11	116. monde 9
29. système 25	58. nécessité 14	87. unité 11	117. place 9

- | | | | |
|------------------------|---------------------|-------------------------|---------------------|
| 118. projet 9 | 149. contexte 7 | 181. objectif 6 | 214. hommes 5 |
| 119. réalisation 9 | 150. création 7 | 182. plan 6 | 215. individus 5 |
| 120. salaires 9 | 151. débat 7 | 183. planification 6 | 216. insertion 5 |
| 121. sécurité 9 | 152. démocratie 7 | 184. possibilité 6 | 217. interdiction 5 |
| 122. services 9 | 153. élément 7 | 185. présence 6 | 218. justice 5 |
| 123. achat 8 | 154. évolution 7 | 187. progression 6 | 219. libération 5 |
| 124. analyse 8 | 155. fédérations 7 | 188. propriété 6 | 220. limites 5 |
| 125. cadres 8 | 156. France 7 | 189. reconnaissance 6 | 221. loi 5 |
| 126. décision 8 | 157. issue 7 | 190. secteur 6 | 222. maîtrise 5 |
| 127. défense 8 | 158. jeunes 7 | 191. socialisation 6 | 223. manière 5 |
| 128. effort 8 | 159. majorité 7 | 192. structure 6 | 224. minimum 5 |
| 129. égalité 8 | 160. matières 7 | 193. syndicalisation 6 | 225. modèle 5 |
| 130. forme 8 | 161. mode 7 | 194. tâches 6 | 226. modification 5 |
| 131. heures 8 | 162. nombre 7 | 195. accroissement 5 | 227. niveau 5 |
| 132. liaison 8 | 163. orientations 7 | 196. affiliation 5 | 228. passage 5 |
| 133. mobilisation 8 | 164. qualité 7 | 197. âge 5 | 229. perspective 5 |
| 134. moyen 8 | 165. revenus 7 | 198. alternative 5 | 230. peuple 5 |
| 135. ordre 8 | 166. santé 7 | 199. années 5 | 231. public 5 |
| 136. orientation 8 | 167. terrains 7 | 200. appropriation 5 | 232. réflexion 6 |
| 137. patronat 8 | 168. CMT 6 | 201. aspiration 5 | 233. refus 5 |
| 138. phase 8 | 169. changement 6 | 202. capital 5 | 234. région 6 |
| 139. priorité 8 | 170. combat 6 | 203. champ 5 | 235. retraités 5 |
| 140. rassemblement 8 | 171. condition 6 | 204. changements 5 | 236. sens 5 |
| 141. responsabilités 8 | 172. consommation 6 | 205. conflits 5 | 237. service 5 |
| 142. restructuration 8 | 173. coopération 6 | 206. croissance 5 | 238. sociétés 5 |
| 143. satisfaction 8 | 174. coordination 6 | 207. décentralisation 5 | 239. soutien 5 |
| 144. solidarité 8 | 175. emplois 6 | 208. définition 5 | 240. statut 5 |
| 145. transition 8 | 176. expérience 6 | 209. embauche 5 | 241. voie |
| 146. armée 7 | 177. extension 6 | 210. équipements 5 | 242. politique ? |
| 147. caractère 7 | 178. instauration 6 | 211. exercice 5 | 243. politiques ? |
| 148. confrontation 7 | 179. jour 6 | 212. gauche 5 | 244. pouvoir ? |
| | 180. niveaux 6 | 213. gestion 5 | |

ANNEXE 2

*Classes de noms**Classe 1 :*

CFDT
congrès
confédération (?)

Classe 2 :

adhérents
militants
sections
syndicats
unions
fédération
organisation (?)
organisations (?)

Classe 3 :

travailleurs
salariés
classe ouvrière
peuples
femmes
immigrés
jeunes
masse

Classe 4 :

action(s)
lutte(s)
combat

Classe 5 :

alternative
projet
autogestion
socialisme
transformation
construction
changement
socialisation
décentralisation
appropriation
perspective
libération

Classe 6 :

patronat
gouvernement
impérialisme
capitalisme
forces dominantes

ANNEXE 3

Liste des verbes codés

(Le premier nombre indique le numéro de code, le second la fréquence des verbes, qui sont tous lemmatisés.)

301. être 100	328. rechercher 12	355. déboucher 7	382. révéler 6
302. devoir 74	329. adapter 11	356. définir 7	383. vivre 6
303. permettre 43	330. concerner 11	357. devenir 7	384. attacher 6
304. mettre en œuvre (?)	331. constituer 11	358. élaborer 7	385. autogérer 5
305. développer 31	332. accroître 11	359. priver 7	386. contester 5
306. pouvoir 31	333. montrer 11	360. renforcer 7	387. continuer 5
307. réaliser 31	334. œuvrer (?)	361. utiliser 7	388. dépasser 5
308. faire 29	335. réunir 10	362. traduire 7	389. dire 5
309. avoir (?)	336. être possible (?)	363. combattre 6	390. élire 5
310. donner 24	337. falloir 9	364. contenir 6	391. établir 5
311. assurer 23	338. favoriser 9	365. décider 6	392. éviter 5
312. prendre en charge (?)	339. il y a (?)	366. changer 6	393. exister 5
313. être nécessaire (?)	340. mettre (?)	367. entendre 6	394. garantir 5
314. imposer 18	341. mettre fin (?)	368. entraîner 6	395. instaurer 5
315. agir 16	342. prendre (?)	369. exiger 6	396. intégrer 5
316. nécessiter 16	343. prendre conscience (?)	370. exploiter 6	397. jouer 5
317. poursuivre 16	344. faire face (?)	371. fonder 6	398. marquer 5
318. être indispensable (?)	345. acquérir 8	372. inscrire 6	399. opposer 5
319. savoir 15	346. industrialiser 8	373. lier 6	400. parvenir 5
320. remettre en cause (?)	347. mener 8	374. lutter 6	401. préciser 5
321. contribuer 14	348. organiser 8	375. mandater 6	402. reconnaître 5
322. déterminer 14	349. voir 8	376. participer 6	403. répondre 5
323. impliquer 14	350. approprier 7	377. passer 6	404. reposer 5
324. dominer 13	351. conditionner 7	378. produire 6	405. tendre 5
325. maintenir 13	352. conduire 7	379. progresser 6	406. travailler 5
326. améliorer 12	353. confédérer 7	380. réduire 6	
327. exercer 12	354. construire 7	381. refuser 6	

ANNEXE 4

Classes de verbes

0. PERFORMATIFS

01. *Positifs* (SN animé complément facultatif ou impossible)
accepter, confirmer, constater, considérer, conclure, communiquer, (re) définir, demander, décider, déclarer, décréter, dire, dicter, estimer, enregistrer, expliquer, exiger, exprimer, formuler, inscrire, insister, montrer, manifester, opter, préciser, proposer, se prononcer, reconnaître, révéler, rappeler, souligner, tirer les enseignements.
02. *Positifs* (SN animé complément obligatoire)
appuyer, appeler, encourager, inviter, mandater, soutenir.
03. *Négatifs*
contester, condamner, dénoncer, interdire, récuser, refuser, remettre en cause, supprimer.

1. LÉGISLATIFS

10. *Opérateurs de phrase*
devoir, falloir, ne savoir (conditionnel), ne pouvoir que, convenir, y avoir lieu.
11. *Prédicats à deux (trois) places*
exiger, imposer, contraindre.
12. *Implicatifs* (conséquence-cause ; condition suffisante)
découler, reposer, résulter.
13. *Implicatifs* (conséquence-cause ; condition nécessaire)
impliquer, nécessiter, dépendre, exiger, supposer, refléter, passer par, appeler.
14. *Implicatifs* (cause-conséquence ; condition suffisante)
déterminer, engendrer, conduire, finir par, aboutir, avoir des effets, avoir pour conséquence, engendrer, fonder, baser, déboucher.
15. *Implicatifs* (cause-conséquence ; condition nécessaire)
conditionner, être nécessaire, être indispensable.

2. POSSIBILITÉ

20. *Opérateurs de phrase*
pouvoir, être possible, être susceptible, avoir la possibilité.

21. *Prédicats à deux (trois) places*

permettre, rendre possible, offrir des possibilités.

3. DÉFINITOIRES

30. *Définitions exclusives*

être, constituer.

31. *Définitions non exclusives classificatoires*

être, représenter.

32. avoir, comporter, comprendre, contenir.

33. c'est.

34. il y a.

4. ANALYTIQUES

41. *Interprétation + contrôlable possible*

manifester, traduire, montrer, révéler, représenter, caractériser, marquer, mettre en lumière, toucher, atteindre.

42. *Interprétation + contrôlable impossible*

s'avérer, apparaître, se manifester, se faire jour.

43. *Coordination*

conjuguer, lier, concilier, coordonner, rejoindre, accompagner, coexister.

44. *Analytiques divers*

correspondre, concerner, opposer, voir, etc.

45. *Verbes d'état*

5. ANALYTIQUES (ESPACE/TEMPS)

50. *Extension non linéaire positive*

développer, étaler, accroître, élargir, multiplier, approfondir, étendre, intensifier.

51. *Extension non linéaire négative*

réduire, concentrer, endiguer.

52. *Extension linéaire positive*

avancer, progresser, dépasser, parvenir, ouvrir la voie.

53. *Extension linéaire négative*

reculer, régresser, baisser.

54. *Temps*

précéder.

55. entreprendre, commencer, aller.

56. poursuivre, continuer, perpétuer.

6. PSYCHOLOGIQUES

60. *Tension positive (+ V2)*

essayer, entendre, s'efforcer, (re) chercher, s'attacher, souhaiter, vouloir, s'employer, tendre, viser, envisager, avoir pour objectif (but), aspirer.

61. *Autres*

accepter, admettre, se contenter, se limiter.

62. éviter, hésiter, menacer, tenter.

63. divers.

7. PROCESSUS + CONTROLABLE

71. préparer.

72. créer, élaborer, mettre en place, mettre en œuvre, susciter, instaurer, établir.

73. bâtir, construire, réaliser, effectuer, assurer, pratiquer, assumer, œuvrer.

74. améliorer, modifier, adapter.

75. maintenir, conserver, consolider, affermir, renforcer.

76. mettre un terme, mettre fin, freiner, détruire, éliminer, briser, supprimer, abroger, renverser, confisquer, faire disparaître.

77. rétablir, rénover, restaurer.

78. transformer.

8. AUTRES VERBES + CONTROLABLE

80. lutter, agir, combattre, œuvrer, résister, être dans l'action.

81. conquérir, vaincre, obtenir, libérer.

82. rassembler, regrouper, unifier, unir, réunir.

83. mener, conduire, mobiliser.

84. divers (lutte).

85. exploiter, dominer, aliéner, extorquer.

86. divers (contre-lutte).

87. faire.

88. divers.

Résumé de l'article /Abstract

POUR UNE GRAMMAIRE DU DISCOURS. ÉLABORATION D'UNE MÉTHODE ; EXEMPLES D'APPLI-CATIONS

Ce texte se propose de constituer une grammaire du discours à base statistique intégrant des faits lexicaux syntaxiques et énonciatifs. Ont été étudié systématiquement, à partir de la Résolution du congrès de la CFDT de 1976, et à l'aide de l'analyse factorielle des correspondances (chaque phrase étant codée selon divers critères retenus) : l'ordre, la structure syntaxique, la classe du nom sujet superficiel, le rapport entre le nom et le verbe principal, la classe du verbe principal, la structure énonciative, la forme active ou passive, la forme affirmative ou négative, la structure modale. Une telle grammaire a deux dimensions. Elle est à la fois « grammaire de fréquences » et « grammaire de séquences ». La « grammaire de fréquences » considère le texte comme un ensemble homogène et étudie les fréquences syntaxiques, lexicales et énonciatives et les corrélations entre les unes et les autres d'une manière globale, sur l'ensemble du corpus. Le corpus ainsi homogénéisé est une abstraction par rapport au corpus réel, mais l'idée même d'une grammaire probabiliste conduit nécessairement à ce type d'abstraction. D'où la nécessité d'une « grammaire de séquences » qui étudie la distribution des règles précédentes à l'intérieur de la séquence des phrases, et donc dans le déroulement linéaire du texte.

CONSIDERING A GRAMMAR OF DISCOURSE. THE DEVELOPMENT OF A METHOD AND SOME EXAMPLES OF APPLICATION

This text sets out to form a statistically based grammar of discourse integrating lexical, syntactic and enunciative facts. With reference to the General Resolution of the 1976 CFDT congress and with the help of the factorial analysis of correspondances (where each sentence is coded according to different selected criteria), the following features have been studied: the order, the syntactic structure, the grammatical category of the noun as superficial subject, the relationship between the noun and the category of the main verb, the enunciative structure, the active or passive form, the affirmative or negative form and the modal structure. Such a grammar has two dimensions. It is both a "grammar of frequencies" and a "grammar of sequences". The "grammar of frequencies" considers the text as a homogeneous whole and makes a global study of the syntactic, lexical and enunciative frequencies and the correlations between them throughout the whole corpus. The corpus, homogeneized in such a way, is an abstract projection of the real corpus but the very concept of a probabilistic grammar leads necessarily to this kind of abstraction. Hence the need for a "grammar of sequences", which studies the distribution of preceding rules within the sequence of sentences and therefore in the linear unfolding of the text.