

SPÉCIAL

REPÈRES

POUR LA RÉNOVATION
DE L'ENSEIGNEMENT DU FRANÇAIS
À L'ÉCOLE ÉLÉMENTAIRE

BULLETIN DE LIAISON ET D'ÉCHANGE DES ÉQUIPES EXPÉRIMENTALES DE FRANÇAIS

**analyse
de textes d'enfants**

**Séminaire
des 24 - 25 - 26 avril 1974**

1976

INSTITUT NATIONAL DE RECHERCHE ET DE DOCUMENTATION PÉDAGOGIQUES

Service des Études et Recherches Pédagogiques

Français 1^{er} degré



ANALYSE DE TEXTES D'ENFANTS

Séminaire des 24-25-26 avril 1974

"Ces documents de travail sont la propriété de l'Institut National de Recherche et de Documentation Pédagogiques. Les droits de reproduction et de traduction sont réservés pour tous les pays. Toute reproduction est interdite. Une copie ou reproduction, par quelque procédé que ce soit, photographie, microfilm, bande magnétique, disque ou autre, constitue une contrefaçon passible des peines prévues par la loi du 11 Mars 1957 sur la protection des droits d'auteurs".

§

"Ces documents de travail à l'état "BRUT" sont en principe destinés au seul usage des équipes de recherche. Ils représentent au moment de leur rédaction une étape dans une recherche donnée. Ils ne visent pas à donner une idée d'ensemble des questions abordées mais à dégager une problématique, à susciter la discussion, l'échange dans les équipes mêmes. On voudra donc bien les lire, non pas en tant que documents achevés, fixant une quelconque "doctrine", mais comme base de discussion possible dans le cadre de recherches en cours".

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

THE UNIVERSITY OF CHICAGO
LIBRARY

SOMMAIRE

- Présentation du Séminaire
H. ROMIAN - I.N.R.D.P. p. 5
- Analyse automatique de rédactions "Libres"
F. CHARPIN - Université de Limoges p. 11
- Vers une description linguistique de production écrites
d'élèves de l'école élémentaire
R. TOMASSONE - Université de Nancy II p. 21
- Remarques sur la spécificité du langage parlé
D. WEILL - Université de Strasbourg p. 51
- Norme et surnorme
F. FRANCOIS - Université de Paris V p. 69
- Une pré-enquête syntaxique sur ordinateur
R. JOLIVET - Université de Lausanne p. 91

Bulletin réalisé par :

Hélène ROMIAN
Responsable de l'Unité de Recherche
Français 1er Degré

Secrétariat : Françoise TOSSARD-DELAIR
Dactylographie : Suzanne OMAR

The first part of the report
 is devoted to a description of
 the general character of the
 country, and to a statement
 of the principal features of
 the topography, geology, and
 natural resources.

The second part of the report
 is devoted to a description of
 the principal cities and towns
 of the State, and to a statement
 of the principal features of
 their history, commerce, and
 population.

Deux ans après.....

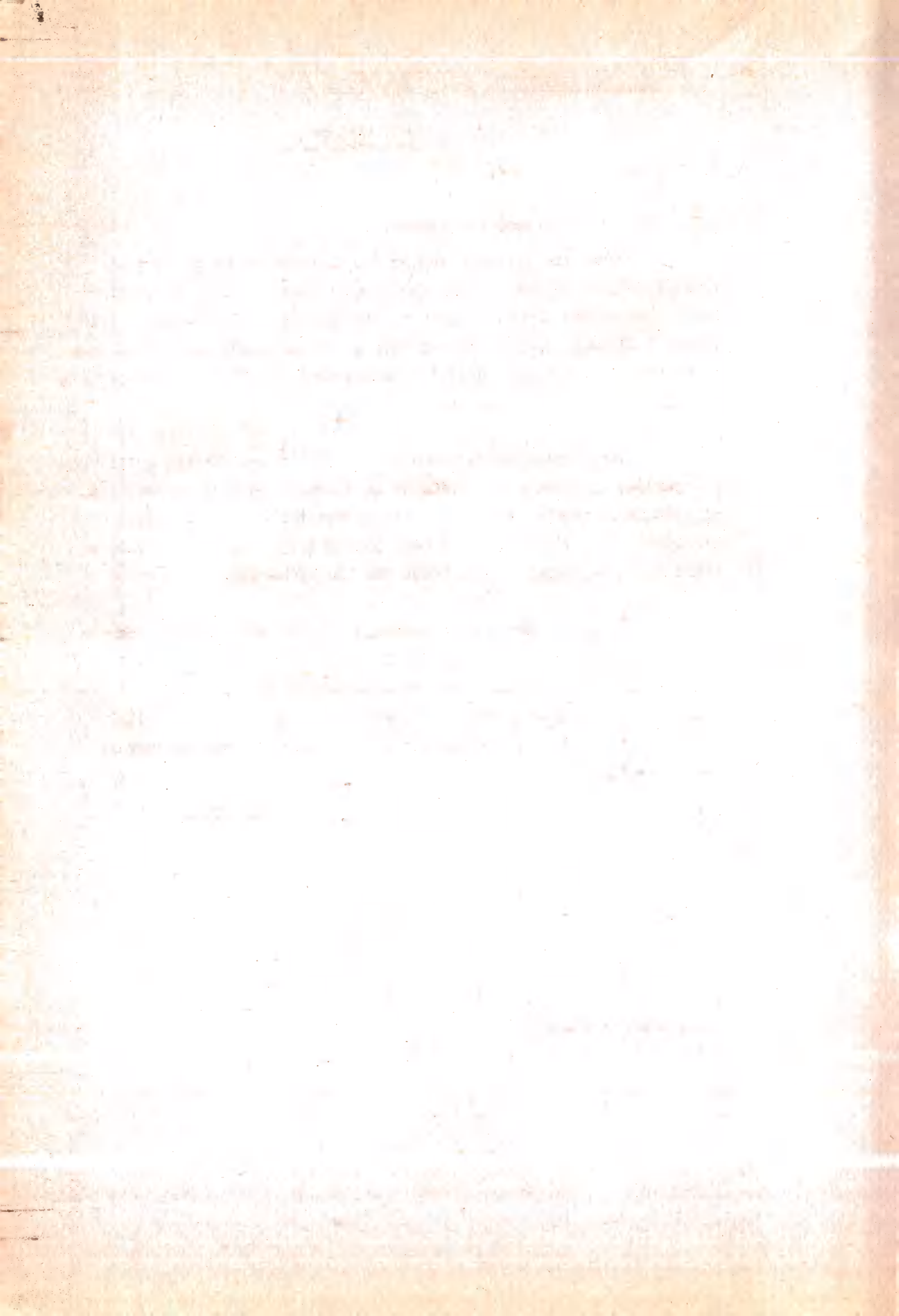
Nous nous excusons auprès des universitaires qui ont bien voulu informer nos équipes sur leurs recherches en cours de publier leurs interventions plus de deux ans après. Nous les remercions très vivement d'avoir accepté d'enfreindre la règle tacite qui filtre soigneusement l'information jusqu'au moment de la publication, règle très préjudiciable à la recherche.

Nous remercions également le CRAL (1) pour l'aide qu'il nous a apportée. Mais les moyens réduits dont nous disposons ne nous ont pas permis de publier rapidement des informations dont la richesse et l'intérêt sont évidents. On ne dira jamais trop l'importance vitale d'une rapide circulation des idées pour la recherche.

Nous demandons à nos lecteurs de bien vouloir tenir compte du fait que, d'une part, il s'agit d'interventions orales en situation de séminaire et que, d'autre part, depuis 2 ans, les uns et les autres ont progressé dans leurs travaux. Tel est cependant l'état de la question que - malheureusement - ces informations conservent leur intérêt...

H. ROMIAN

(1) Centre de Recherches Appliquées à la Linguistique - Nancy II



L'analyse de textes d'enfants. Pourquoi ? Pour quoi faire ?

H. ROMIAN

En avril 1974, l'Unité de Recherche Français 1er Degré organise un séminaire centré sur l'analyse de textes d'enfants. Y participent, d'une part des Universitaires engagés dans des recherches sur ce thème, et d'autre part des animateurs d'équipe I.N.R.D.P. dont certains ont engagé des études empiriques sur l'analyse de textes d'enfants, d'autres travaillent sur une épreuve de syntaxe CM1, d'autres enfin participent à un groupe de réflexion⁽¹⁾ sur l'analyse automatique de textes d'enfants menée par le CRAL dans le cadre d'un essai de validation expérimentale des hypothèses pédagogiques du Plan de Renovation. L'objectif majeur du séminaire est de "recentrer" les travaux I.N.R.D.P. à partir des informations reçues et - si possible - de déboucher sur l'organisation d'un groupe national de recherche I.N.R.D.P. sur la description de la langue écrite des enfants d'école élémentaire, du CP au CM2. Effectivement, le séminaire d'avril 1974, marque l'acte de naissance d'un "Groupe Analyse de textes" qui deviendra "Groupe Langue Ecrite" (GLE)⁽²⁾

Les questions qui se posent à l'Unité de Recherche sont nombreuses et complexes. Nous avons estimé que la rénovation de l'enseignement du français doit viser la maîtrise de la langue orale et écrite pour tous les enfants. Mais cette notion de maîtrise devrait être cernée, analysée : quels objectifs pédagogiques explicites et implicites recouvre-t-elles ? quelles capacités veut-on développer et/ou créer par l'apprentissage de l'écriture (au sens large du terme) ? et pourquoi celles-là ? ces capacités sont-elles essentiellement d'ordre linguistique et/ou d'ordre psychologique, et/ou sociologique ? Selon quels indices objectifs ces capacités se manifestent-elles dans les textes d'enfants ? Si l'on considère par exemple la capacité à produire un discours cohérent d'un point de vue logique, et/ou rhétorique, et/ou énonciatif etc... quels aspects formels peut-on considérer comme des indices de cohérence, au niveau du texte ? de ce point

----- ./.
(1) Les réunions de ce groupe sont trop espacées pour être réellement fructueuses. **Plais d'argent est mortelle...**

(2) Depuis janvier 1974, existe un Groupe Langue Orale, centré sur l'élaboration d'un sondage phonologique C.P

de vue, y-a-t-il des productions écrites enfantines que l'on puisse considérer comme des "non-textes" ? sur quel(s) modèle(s) théorique(s) peut-on s'appuyer pour analyser du discours écrit, compte-tenu du fait que la linguistique fonctionne au niveau de la phrase et non du texte ?

Si l'on raisonne sur tout autre aspect, par exemple, la capacité à utiliser la langue de manière fonctionnelle (selon les diverses fonctions du langage), se posent des questions de même ordre.

Si l'on part, inversement d'indices de maîtrise déterminés empiriquement comme les types de phrases, les coordonnants, les subordinants, l'expression de la causalité, de la simultanéité, le nombre de mots différents par rapport au nombre total de mots du texte etc..., comment interpréter les décomptes effectués ? que prouve par exemple la non-utilisation de telle ou telle structure, la longueur moyenne des phrases ou la longueur d'un texte ? Apparaît ici l'importance capitale des situations de production des textes en tant qu'elles induisent probablement un certain nombre de structures, et qu'elles en excluent probablement d'autres ; mais selon quelles données définir ces situations, et la consigne scolaire de rédaction y suffit-elle ? à quels indices objectifs dans le texte peut-on voir que l'enfant donne la situation de communication écrite qui a induit le texte ? ou qu'il ne la donne pas, et quels apprentissages en découleraient ? Existe-t-il des stades d'acquisition de la langue écrite ? entre les textes de CM1 et ceux de CM2, peut-on saisir des différences significatives ? comment procéder pour observer les événements d'ordre générique, compte tenu de la multiplicité des facteurs qui peuvent influencer les productions écrites scolaires ? Quelles relations existe-t-il entre le développement (ou l'acquisition) des capacités requises par l'acte de production de textes écrits et la formation psychologique, sociale de l'enfant ? Quelle aide peut apporter l'orthographe aux recherches nécessaires pour répondre à ces questions complexes ? peut-il, dans l'état actuel des connaissances, effectuer autre chose que des comptages à partir desquels on peut espérer affiner les hypothèses de recherche et, peut-être, en apercevoir d'autres ? Il est certain que de longs et difficiles travaux d'analyse "manuelle" sont nécessaires pour y voir plus clair, dont l'issue et le déroulement sont imprévisibles actuellement. En tous cas, pour le moment

l'analyse automatique est impuissante à distinguer des énoncés du type "la table est sous le parapluie" des énoncés sémantiquement corrects... etc... De toute façon se pose le problème de la finalité des analyses de textes faites ou à faire : les équipes IIRDP ne font pas de la recherche en linguistique, et pas même de linguistique appliquée ; leurs finalités sont, de toute évidence, pédagogiques. De ce point de vue, nous savons ce dont nous ne voulons plus : l'analyse - "correction" normative - et même surnormative, selon l'expression de P. François - le relevé de fautes tel que nous l'avons tous pratiqué, selon les fatidiques "impr", "m.d.", "inc". ; autant de "renforcements négatifs" plus propres à décourager les enfants qu'à leur donner la moindre envie de se "corriger". Restent à construire - et c'est là une affaire de longue haleine - d'autres "grilles" d'analyse des textes d'enfants, propres à évaluer ceux-ci positivement, en termes de réussites et non d'échecs. Ces grilles, conçues selon des critères pédagogiques, devraient pouvoir servir à la fois d'outils de recherche descriptive et d'outils de formation des maîtres. La formation des maîtres n'est-elle pas en effet le meilleur canal de diffusion possible des "produits" de la recherche pédagogique ?

Cette présentation, rédigée en mai 1976 d'après les interventions d'ouverture et de conclusion du séminaire "Analyse de textes d'enfants", a tenté de ne pas déborder de la problématique qui était celle des équipes en 1974. C'est dire que bien du chemin restait à faire pour parvenir à élaborer un cadre théorique de référence, des hypothèses et un plan de recherche opérationnels. C'est dire aussi l'extraordinaire difficulté de l'entreprise.

Mais c'est dire enfin son utilité : nous vivons, en matière d'"évaluation" des productions écrites des enfants, et en matière de "mise au point", de travail sur ces textes, en plein empirisme. Sans parler des apprentissages nécessaires tout au long de la scolarité, qui ne se font que de manière diffuse et occasionnelle. Les pédagogues ne peuvent pas attendre que linguistes, psycho et sociolinguistes aient mis au point un (des) modèle(s) théorique(s) au(x) quel(s) ils puissent se référer. La recherche en pédagogie peut tenter de répondre, ou du moins d'élaborer des outils de travail qui permettent - pas plus, mais pas moins - de rationaliser l'apprentissage

de la langue écrite tout au long de la scolarité, selon des objectifs et des démarches plus explicites, plus cohérents, et de la dynamiser en donnant aux maîtres la possibilité de passer d'évaluations normatives à des évaluations objectives faites, selon des critères positifs.

ANALYSE AUTOMATIQUE DE REDACTIONS "LIBRES"

F. CHARPIN

Centre de Linguistique Appliquée de l'Université de Limoges

Le Centre de Linguistique Appliquée réunit une équipe inter-disciplinaire qui travaille dans le cadre de l'Université de Limoges. Malgré l'absence de chercheurs à temps plein, malgré l'absence de crédits spécifiques, plusieurs enseignants se sont groupés pour effectuer une enquête sur le traitement automatique de la "performance" dans le domaine de la langue française. Ont travaillé en collaboration étroite tous les professeurs du Centre de Calcul de l'Université, deux enseignants de la Faculté des lettres, des enseignants de l'Ecole Normale de Jeunes Filles (notamment le professeur de psychologie) ; l'inspection Académique, le Centre Régional de Documentation Pédagogique nous ont apporté leur concours. Plusieurs centaines d'instituteurs de la Haute-Vienne se sont joints à l'équipe de fondation et nous ont transmis 12 000 rédactions "libres" qui ont été rédigées par les élèves de leurs classes en Mars et en Avril 1972. Les textes ont été réclamés par Monsieur l'Inspecteur d'Académie qui avait adressé une circulaire à toutes les écoles. L'ensemble ainsi constitué manque sans doute de cohérence. Plusieurs textes sont réellement des rédactions "libres" c'est-à-dire composées par les enfants en dehors de toute contrainte, selon leur inspiration, chez eux ou à l'école... D'autres textes ont été revus, voire même suggérés par l'instituteur et ne méritent absolument pas le nom de rédactions "libres". Parallèlement à l'envoi des textes les instituteurs nous ont adressé une documentation sur chaque enfant : ils nous ont indiqué les noms, prénoms, âge, rang dans la fratrie, profession des parents, résultats scolaires, zone d'habitation, niveau de ressources des parents, santé, troubles du langage, comportement dans le groupe, origine de la famille (Limousin, autres régions, autres pays)... L'ensemble de ces données a été codé et se trouve mis en rapport avec les textes que nous possédons. Pour les textes proprement dits, nous les avons traités d'après les principes qui ont été établis par le Laboratoire d'Analyse Statistique des langues

Anciennes de l'Université de Liège : nous avons décidé d'attribuer une carte perforée pour chaque mot : la carte comporte quatre zones ; dans la première nous avons laissé le mot tel qu'il a été écrit par l'enfant ; dans la deuxième nous avons écrit le mot en le corrigeant si besoin était ; dans la troisième nous avons placé le lemme c'est-à-dire le mot tel qu'il est écrit dans le dictionnaire ; dans la quatrième sont marqués les différents codages morphologiques et syntaxiques, les références qui permettent d'identifier le mot dans la rédaction, la rédaction dans l'ensemble des rédactions de l'enfant et l'enfant dans son groupe.

Nous disposons d'une documentation tout à fait originale parce qu'elle nous autorise l'étude du Français tel qu'il est écrit par l'ensemble des enfants d'un Département. Il suffit en effet de comparer les zones 1 et 2 de chaque fiche pour obtenir un dictionnaire des fautes, de comparer les zones 2 et 3 pour obtenir un dictionnaire du vocabulaire des enfants dans l'absolu et en fonction du contexte. En manipulant les données syntaxiques transcrites dans le codage nous pouvons dresser l'inventaire des modèles de phrases employées. Nous pouvons enfin mettre en rapport le vocabulaire et les structures syntaxiques. Dès maintenant nous avons obtenu quelques résultats partiels. C'est ainsi que nous avons constaté que tous les élèves d'une école située en plein centre de la ville de Limoges sont susceptibles d'employer 1800 mots. Sur ce total, les enfants du Cours Préparatoire n'utilisent que 340 mots, ceux du Cours élémentaire 1ère année 980, ceux du Cours élémentaire 2ème année 945, ceux du Cours moyen 2ème année 900. Dans chaque classe les enfants n'utilisent que la moitié du vocabulaire de l'école ; dès le Cours élémentaire 1ère année ils possèdent la connaissance du Français-standard et ne progressent pratiquement plus dans les trois classes suivantes. Du point de vue de la syntaxe, tous les enfants de la même école bâtissent la moitié de leurs énoncés sur 4 modèles de phrases :

Substantif sujet Verbe Substantif objet

Circonstant Substantif sujet Verbe Substantif objet

Substantif sujet Verbe Circonstant

Circonstant Substantif sujet Verbe Circonstant

Le tableau suivant donne la répartition selon les classes.

Classe	nombre de types de c.p.r. attestés	142	5142	145	5146	Total
C.P.	46	15,81 %	4,42 %	8,84 %	15,04 %	45,11 %
C.E. 1	55	8,57 %	8,57 %	5 %	7,14 %	29,28 %
C.E. 2	143	8,42 %	5 %	7,36 %	8,15 %	28,93 %
C.I. 1	112	11,62 %	5,81 %	8,52 %	4,65 %	30,60 %
C.I. 2	197	7,62 %	5,78 %	8,67 %	5,61 %	27,68 %

Il est surprenant que les enfants préfèrent des phrases telles que : Avec mes amis je suis allé à la campagne et n'emploient pas de phrases telle que : je suis allé à la campagne avec mes amis. L'explication se trouve peut-être dans les modèles de phrases qui sont appris par l'intermédiaire des exercices scolaires.

Nous avons la possibilité d'étudier les différents champs sémantiques qui structurent le vocabulaire des enfants. Nous avons ainsi constaté que dans une école de garçons, les mots qui désignent un personnage féminin (Maîtresse, dame, soeur, fille...) sont deux fois plus nombreux que les mots qui désignent un personnage masculin (Maître, Monsieur, frère, fils.). Dans une école de filles, la proportion est inverse. D'après les devoirs nous obtenons un tableau très fidèle des principaux centres d'intérêt d'un écolier : le tiers du vocabulaire est consacré à la nourriture, un autre tiers aux choses de la nature (animaux, plantes, paysages..) le reste concerne divers éléments de civilisation. La comparaison du vocabulaire de tous les enfants d'une même classe fait ressortir des éléments qui sont parfois très significatifs. Par exemple dans certaines copies le JE est cité plus de douze fois ; dans d'autres il n'apparaît jamais. Fait plus troublant nous avons trouvé deux narrations où les enfants ne se réfèrent à aucun pronom personnel, mais utilisent exclusivement des substantifs. Le psychanalyste que nous avons consulté considère les deux copies comme l'illustration de troubles graves dans la représentation du Moi. D'un point de vue ethnologique, les devoirs renferment un nombre considérable de notations sur la vie dans le Limousin. Nous savons ainsi que,

pour les enfants, une maison ne comporte que deux pièces principales : la cuisine et la chambre ; que le mobilier se compose essentiellement de chaises et de tables ; nous savons le menu exact d'un repas-type ; deux boissons seulement sont citées : l'eau pour les repas ordinaires et le champagne pour les repas fêtes... Les études que nous avons commencées, nous permettront de publier des monographies sur de très nombreux sujets par exemple sur l'autorité des parents, le mobilier, la famille, la nature dans les devoirs des écoliers du Limousin.

Nous aurons la possibilité de mettre toutes ces enquêtes en rapport avec les renseignements que nous possédons sur chaque enfant. Nous obtiendrons alors une description précise de la langue des élèves que les instituteurs classent comme "bons élèves" ou comme "élèves médiocres". Nous pensons isoler les traits spécifiques des micro-langages ainsi définis. Est-ce qu'un bon élève est celui qui emploie des mots originaux ou des mots banals, qui utilise un très petit nombre ou au contraire un grand nombre de modèles de phrases ? Est-ce qu'un mauvais élève est celui qui emploie des formes patoisantes ? Est-ce qu'un fils de notaire écrit comme un fils d'ouvrier ; Est-ce qu'un fils d'ouvrier écrit comme un fils d'employé de commerce ? Dans ce domaine nous avons pu constater que la profession des parents est sans doute moins importante que leur lieu d'habitation. Il est somme toute assez normal que, dans une époque où la séparation entre les quartiers riches et les quartiers pauvres est fortement accentuée par l'urbanisme, la différence entre la langue des enfants qui habitent une ZUP et la langue des enfants qui habitent les quartiers résidentiels, soit elle-même fortement marquée mais que toutes les différences s'estompent à l'intérieur de chacune de ces deux zones d'habitation. Si on habite un quartier résidentiel on parle la langue-standard du quartier résidentiel que l'on soit fils de notaire, d'ingénieur ou de gros commerçant ; si on habite une ZUP on parle la langue-standard de la ZUP, que l'on soit fils d'ouvrier, d'employé de commerce ou de petit fonctionnaire. Les différences linguistiques et les différences sociales nous ont semblé inscrites dans la géographie des villes.

DISCUSSION

- Existe-t-il une procédure pour vérifier le rapport des mots étudiés avec le contexte ?

- L'avantage fondamental de l'ordinateur est que l'on peut chaque fois et pour toute étude vérifier le contexte à la demande, qu'il s'agisse du contexte d'un mot, du contexte d'une phrase ou même du contexte d'une rédaction parmi les autres rédactions de l'enfant, ou de tous les enfants de la classe, ou encore de l'ensemble des rédactions ayant un thème donné ... Nous avons en effet choisi le système d'entrée des données le plus astreignant puisque nous utilisons la carte perforée et que nous consignons un mot par carte ou un signe de ponctuation par carte. Le travail est considérable. Nous disposons déjà de 180 000 cartes. Quand toute la documentation sera traitée, nous aurons plus de 2 000 000 de cartes. Elles sont toutes repérées et classées ; pour faire sortir le contexte, il suffit d'une manipulation très simple de l'ordinateur. Il va de soi que pour obtenir ce résultat, nous avons utilisé des systèmes de codage extrêmement simplifiés. En particulier nous repérons une phrase parce qu'elle est séparée par deux cartes contenant un point dans le fichier-machine. Nous distinguons les fonctions syntaxiques parce que nous utilisons une terminologie actantielle empruntée à Tesnière qui nous donne une première approximation de la structure de l'énoncé. Cette terminologie a le mérite d'être très simple, donc facilement compréhensible pour un codificateur non-informé ; elle possède également le mérite de pouvoir être ré-interprétée dans le cadre d'une grammaire générative transformationnelle. Grâce à ces commodités que nous nous sommes données, le texte que nous avons pris la précaution de conserver dans sa forme originale demeure constamment accessible. Nous avons évidemment renoncé, dans un premier temps, à procéder à une analyse syntaxique automatique. Nous avons en effet considéré que nous perdriions un temps considérable si nous programmions un système logique a priori qui, dans l'état actuel de la science grammaticale, ne serait pas fiable et comporterait dans la meilleure hypothèse 45 % d'erreurs. Je reconnais personnellement que notre système est un bricolage artisanal mais l'essentiel est qu'il n'interdise pas les enquêtes ultérieures.

- Par qui la fiche signalétique a-t-elle été rédigée, par l'enfant ou par l'instituteur ?

- Elle a été rédigée par l'instituteur. Il était pour nous essentiel que le responsable de l'enseignement dans une classe nous indique lui-même ce qu'il sait et ce qu'il pense d'un élève. Notre but n'est pas de décrire le comportement intellectuel de l'enfant mais de décrire les structures linguistiques et à partir d'elles les structures mentales d'un élève qui suit le cursus régulier d'une école. Tout enseignant sait très bien qu'un élève du cours préparatoire possède une imagination et une créativité qui se manifestent dans ses dessins, dans les thèmes qui sont traités dans ses rédactions libres... Statistiquement nous avons en gros et avec toutes les réserves qu'il faut apporter à ce genre d'expertise, relevé une cinquantaine de sujets que développent les jeunes enfants. Au cours moyen 2ème année toute cette richesse a disparu. Les devoirs, qui étaient très courts en début de scolarité, deviennent beaucoup plus longs. Certains font une page entière. Ils tournent tous autour de deux ou trois lieux communs tels que la promenade en famille, le bon dimanche, un repas de fête... Les thèmes sont tellement figés que nous sommes en mesure de programmer la rédaction-type de chacun de ses sujets : Nous savons par exemple que la phrase d'introduction pour une sortie en famille sera toujours du type Papa, Maman, mon frère et moi ... Et que la phrase de conclusion sera Ah quelle bonne journée ! Le procédé est tellement automatique qu'un enfant qui venait de raconter un accident où il y avait eu deux morts n'a pas manqué de terminer son devoir en s'extasiant sur le plaisir que lui avait donné sa promenade. Il est bien évident que ce ne sont pas les enfants qui sont responsables de cet état de choses mais bien plutôt l'institution scolaire. C'est pourquoi nous souhaitons que ce soit l'instituteur et non pas l'enfant qui remplisse la fiche signalétique. Qu'on nous comprenne bien : Nous ne pensons pas à l'instituteur en tant qu'individu mais en tant que participant dans le cadre du Ministère de l'Éducation à toute une expérience de formation où se trouvent également engagés l'enseignement secondaire et l'enseignement supérieur. Les enseignants du Lincousin ne s'y sont pas trompés : Il n'est pas indifférent de vous signaler que mon bureau a été plusieurs fois forcé par des collègues de la Faculté soucieux de récupérer les travaux de leur progéniture !

- Comment les textes ont-ils été produits ?

- Tous les textes nous ont été donnés comme "textes libres". Leur récolte a été opérée grâce à la collaboration de Monsieur l'Inspecteur d'Académie qui les a réclamés par une circulaire et qui nous les a fait parvenir par la voie hiérarchique. Dans certaines écoles, l'instituteur disposait de quelques rédactions libres c'est-à-dire écrites par l'enfant, chez lui ou à l'école, sur un thème qui lui plaisait, à un moment où il avait envie d'écrire. Dans d'autres écoles l'instituteur s'est contenté de demander à ses élèves de rédiger en deux heures une rédaction libre ; il s'agit alors d'un sujet libre. Certains instituteurs ont revu et corrigé les devoirs, ont obligé les enfants à changer une expression qui leur déplaisait. Dans d'autres cas nous avons le devoir dans sa forme originale. La collection de textes ainsi obtenue est extrêmement disparate. Elle contient des travaux effectués dans une classe Freinet, des travaux inspirés par le maître, des travaux inspirés par le groupe, des travaux inspirés par les lectures, des travaux inspirés par le milieu familial ... Nous avons tout aussi bien une "chasse au requin" à Saint-Léonard-de-Moblat que trente devoirs d'une même classe qui décrivent le hamster de l'école... Au risque de scandaliser les pédagogues nous pensons que cela n'a aucune importance pour notre étude. Celle-ci n'est possible que dans la mesure où nous pensons justement que tous les textes sont déterminés par le milieu familial, l'école, les lectures... Nous ne cherchons pas ce qu'ils ont d'original ; nous essayons de définir au contraire ce qu'ils ont de commun et de voir si dans une même classe, à un même âge, dans une même école ou dans toutes les écoles du Limousin, ils obéissent à des modèles facilement identifiables que nous pourrions présenter comme les modèles qu'impose la Société dont l'école ne serait qu'un des éléments mais non le moindre puisqu'elle détermine la pratique linguistique de l'enfant dans l'écrit.

- Il n'est pas facile de savoir si les textes appelés textes libres par le maître correspondent réellement à la définition que l'on propose habituellement : est-ce que l'enfant écrit quand il veut, comme il veut, écrit pour lui ou pour le maître ? Souvent un texte que l'on prétend libre n'est qu'un sujet libre. Est-ce que vous ne croyez pas nécessaire de contrôler de façon précise les conditions de production des textes ?

Dans l'immédiat ce problème ne nous préoccupe pas. Nous souhaitons d'abord, en utilisant un corpus hétérogène, vérifier s'il présente des régularités dans le domaine de la syntaxe et du vocabulaire. Quand l'enfant écrit il emploie des phrases, il emploie des mots. Il nous paraît plus important de retrouver des structures identiques à travers des textes dont certains sont des textes libres au sens strict du terme plutôt que d'établir des distinctions qui seront forcément arbitraires à l'intérieur de notre fichier. Bref, nous nous intéressons à la pratique de l'enfant qui écrit avant de nous intéresser à la pratique de l'enfant qui écrit dans des conditions de contrainte plus ou moins grandes. Ultérieurement, et en fonction des résultats de cette enquête, nous aurons les moyens, grâce à la collaboration des enseignants mais aussi en nous référant aux études déjà entreprises par Monsieur KILMER sur les instituteurs limousins, de mettre en rapport les méthodes pédagogiques des enseignants et les résultats de notre enquête linguistique. Pour l'instant nous croyons suffisant d'analyser dans son ensemble les écrits des écoliers à travers des textes qu'on nous a présentés comme "libres" et que nous avons interprétés comme n'appartenant pas à des exercices tels que les exercices grammaticaux ou les exercices techniques dans lesquels la structure de la phrase et le vocabulaire sont strictement imposés par la nature-même de la discipline. Le degré de spontanéité des textes ne nous paraît pas encore un critère déterminant pour que l'on étudie l'originalité du vocabulaire et de la syntaxe.

- La tâche qui consiste à écrire se présente de façon très différente pour les enfants selon le milieu familial auquel ils appartiennent. Il y a une différence entre l'usage de la langue écrite à l'école et l'usage de la langue écrite à la maison.

Pouvez-vous mettre en évidence ces différences ? Pouvez-vous montrer l'interférence des récitations, de ce qui a été appris et vu, par exemple les émissions télévisées ...?

- Nous nous intéressons principalement à la langue écrite par les enfants dans le cadre scolaire. Nous envisageons certainement d'étudier ultérieurement les modifications linguistiques apportées par une émission télévisée... Dans l'immédiat nous avons une enquête suffisamment vaste. Nous avons décidé de laisser de côté toute une série d'analyses sur les rapports entre la langue écrite et la langue parlée, entre la langue écrite à l'école et la langue écrite à la maison. Notre ambition, encore une fois, est de montrer comment d'année en année l'école (ou la société, ou les événements...) et tout ce qui se manifeste à travers l'école, impose aux enfants des modèles syntaxique et sémantique de plus en plus contraignants. Dans la limite des renseignements qui nous sont fournis par les fiches signalétiques, nous mettrons en relation la linguistique et les catégories sociales... Nous ne pouvons faire plus dans l'immédiat. Il va de soi que nous avons l'intention d'interpréter, si possible, les modèles ainsi obtenus.

- Ecrire c'est se trouver devant une page blanche, organiser un cadre typographique, présenter des paragraphes etc... Certains textes comportent en effet des dessins, sont parfois même entre-coupés de dessins, sont écrits en vers... Nous avons dû laisser de côté cet aspect de leur organisation tout du moins dans l'étude que nous entreprenons actuellement.

Est-ce que vous vous intéressez à la rhétorique des textes, et plus spécialement à l'expression des relations spatio-temporelles ou à l'expression des relations personnelles ?

Dans la mesure où l'ordinateur donne le contexte de tout mot ou de toute phrase, nous pouvons étudier très précisément comment les phrases sont enchaînées les unes aux autres et comment les référents sont organisés dans un même texte. Toutefois dans l'immédiat, nous ne croyons pas qu'il soit nécessaire d'orienter l'analyse sur des notions trop sémantiques. On ne gagne rien à admettre des principes logiques a priori qui permettraient de juger de la valeur d'un énoncé. Bref nous entendons rester au niveau d'une description linguistique et nous refusons de nous laisser entraîner dans des distinctions philosophiques.

Il est inutile de savoir si la conjonction OU a valeur exclusive pour juger de la grammaticalité d'un texte. Nous abandonnons à la stylistique le soin de juger de ces problèmes. Quel est votre projet au-delà de cette description des modèles sémantique et syntasique ?

Nous souhaitons essentiellement décrire la pratique linguistique des enfants qui écrivent et qui apprennent à écrire à l'école, en utilisant l'inventaire le plus vaste possible, en conservant le plus grand nombre de critères possibles tout au long de l'analyse, en procédant le plus longtemps possible à une investigation exhaustive de notre corpus. De la documentation très abondante que nous employons, nous pensons dégager les lignes de force qui, d'année scolaire en année scolaire, caractérisent spécifiquement l'action de l'Ecole en tant qu'institution insérée dans une société donnée. C'est pourquoi nous accorderons la plus grande importance à toutes les sciences qui permettent l'interprétation des modèles que nous découvrirons (notamment la sociologie et la psychanalyse). En gros, et avec toutes les réserves qu'il convient d'apporter à ces deux termes, notre projet n'est pas "pédagogique" mais en quelque sorte "ethnologique".

VERS UNE DESCRIPTION LINGUISTIQUE DE PRODUCTIONS ÉCRITES
D'ÉLÈVES DE L'ÉCOLE ÉLÉMENTAIRE

R. TOMASSONE - Université de Nancy II

La recherche que je présente ici aujourd'hui a été entreprise il y a deux ans, dans le cadre du C.R.A.L. (Centre de Recherches et d'Applications Linguistiques), et en relation étroite avec l'I.N.R.D.P. C'est ce qui explique son double aspect de recherche "soumise", en quelque sorte, et de recherche autonome.

Recherche soumise tout d'abord : le premier objectif était défini dans le cadre du plan de validation du projet de rénovation de l'enseignement du français à l'école élémentaire. Il s'agissait avant tout de mettre au point une méthode d'analyse de textes par ordinateur qui permette de fournir rapidement des documents et des résultats aux équipes de l'I.N.R.D.P. Ceci explique la nature et les dimensions de notre corpus. Ce corpus est constitué de 2 200 copies environ, qui ont été recueillies en mai 1972 dans diverses écoles réparties dans toute la France. Ces copies étaient rédigées à partir de trois sujets différents : un sujet de rédaction, une histoire à compléter, et un sujet libre. Sur ces 2 200 copies, ou du moins sur un certain nombre d'entre elles, nous nous engageons à fournir au niveau du vocabulaire, au niveau de la morphologie, au niveau de la syntaxe, tous les renseignements utiles. Nous devons fonctionner essentiellement comme un libre-service de données linguistiques : et nous sommes à même de le faire à cette heure. Je préciserai simplement la façon dont nous avons choisi de préparer ce travail. Nous avons écarté de propos délibéré, dès le début, l'idée de tenter une analyse automatique. Etant donné les moyens matériels et le personnel dont nous disposions, il eût été utopique de l'entreprendre. Comme il fallait envisager de fournir le plus vite possible des index, des résultats, des dénombrements nous avons opté pour une codification. Nous avons travaillé sur des cartes-textes, c'est-à-dire que nous avons enregistré le texte en séquence, le rendant simplement lisible,

c'est-à-dire introduisant des corrections au niveau de la morphologie seulement. Toutefois, le désir d'enregistrer scrupuleusement le texte tel qu'il avait été écrit par l'élève nous a conduits à assortir le texte corrigé d'une "zone de fautes", à garder la trace de toutes les modifications apportées. Par ailleurs, nous avons codé sur deux zones de commentaires, un certain nombre de renseignements concernant la morphologie et la syntaxe. Concernant la morphologie, il s'agit bien sûr, des catégories grammaticales, avec quelques précisions pour chacune d'elles, pour les substantifs et adjectifs le nombre, pour le verbe : temps, mode, personne ; etc...

Au niveau de la syntaxe, le code indique au début de chaque phrase et de chaque proposition le type de phrase et les transformations appliquées. Sont indiquées en outre, les bornes des propositions subordonnées et les limites des groupes, ainsi que leur nature et leur fonction. Une dernière zone, de lemmatisation, permet de regrouper toutes les formes d'un même mot. (1)

A l'heure actuelle, des quelque 2 000 copies que nous avons, environ 400 ont été codées. Ces textes ont été choisis au hasard dans l'ensemble des paquets et des sujets, de façon à constituer un échantillon représentatif de la population. Sur cet échantillon ont été réalisés des index :

- index des lettres par occurrences
- index des lettres par catégories grammaticales
- index des mots par occurrences
- index des mots par catégories grammaticales
- index des types de phrase

Toute autre liste ou index portant sur des éléments du code pourraient être fournis à la demande.

Ces index sont des résultats intermédiaires qui constituent des instruments de travail à la fois pour les équipes de l'I.H.R.D.P. et pour notre propre recherche autonome.

(1) Le détail de la codification a été publié dans les cahiers du C.R.A.L. 3^e série, n° 1 "Productions écrites d'élèves de l'école élémentaire".

Cette dernière n'est pour le moment qu'une ébauche.

Nous présentons ici les premiers résultats d'une analyse provisoire, c'est-à-dire sommaire et incomplète, non tant de point de vue de la méthode - qui nous paraît efficace et adaptée à son objet - que du point de vue du corpus (trop) petit et insuffisamment diversifié sans doute) et des données analysées.

Tels qu'ils sont, ces résultats doivent servir d'éléments de réflexion et de discussion pour établir les bases d'une analyse définitive dont l'objet serait la description linguistique d'un corpus écrit et la comparaison de plusieurs corpus supposés différents.

LE CORPUS sur lequel a porté l'analyse est encore réduit par rapport à l'ensemble des copies codées : il n'a été possible d'analyser dans ce premier train de traitement que 270 copies (90 élèves traitant chacun les trois sujets).

LA METHODE appliquée est l'analyse factorielle des correspondances. Nous l'avons choisi pour deux raisons d'abord parce qu'elle est très utilisée dans n'importe quelle discipline pour des études de comparaison, et nous avons en vue la comparaison des trois sujets. Ensuite, dans la mesure où elle s'applique à des variables linguistiques dénombrées dans des textes, elle a l'avantage de permettre de considérer un texte comme un ensemble organique de variables, de déterminer les liaisons éventuelles entre variables linguistiques dans différents textes.

LES DONNEES LINGUISTIQUES. Il est évident que l'intérêt des résultats dépend de la façon judicieuse dont les données linguistiques auront été choisies et recueillies à partir des textes.

Les variables que nous soumettons à l'analyse seront toutes des dénombrements de faits linguistiques : qu'appelons-nous "fait linguistique" et quels sont ceux qui méritent d'être retenus ?

entre les variables : après un retour aux textes, on pourrait établir celles qui sont dues au thème et celles qui résultent de différences dans le maniement de la langue. Enfin, il doit être possible (cf. supra) de vérifier que les enfants n'emploient pas la même langue suivant le sujet traité, et de préciser les différences.

A chacune de ces questions nous essaierons de répondre en comparant les résultats des différentes analyses que nous avons effectuées.

Nous redisons ici que ces résultats peuvent seulement servir d'indications : ils ne seront définitifs que lorsque les données analysées auront été systématiquement recueillies à partir d'une grille d'analyse complète et raisonnée.

Dans une première analyse, chacune des variables "grammaticales" et "syntaxiques" énumérées ci-dessus a été dénombrée pour chacun des élèves dans l'ensemble des trois textes. On a 90 observations et 56 variables représentant chacune la somme de chaque variable pour les trois sujets.

Les trois premiers facteurs donnent 34.18% de la variabilité totale. Nous bornerons notre analyse à leur examen.

Le trait de différenciation le plus net concerne les personnes du verbe et du pronom personnel : l'emploi du "je" est un caractère assez commun à l'ensemble des textes et faiblement discriminant ; par contre dans le plan du premier facteur, l'emploi de la 3ème personne et de la 2ème s'oppose à celui de la 5ème et de la 6ème, ces deux dernières paraissent encore plus étroitement liées que les deux premières. Le second facteur isole la 4ème personne dont l'emploi, proche de la première, apparaît comme incompatible avec celui des autres personnes.

La projection des individus dans le plan de ces mêmes facteurs 1 et 2 permet de voir que seuls des individus isolés sont caractérisés en particulier par l'emploi de la 3ème ou de la 2ème personne ; ils s'opposent à quelques rares individus caractérisés par la 5ème ou par la 6ème personne. Par contre l'utilisation de la 4ème personne comme de la 1ère apparaît comme un fait de groupe, une tendance plus générale, et ceci d'autant plus que les points correspondants à ces variables sont assez voisins de l'origine.

Variables dites "grammaticales"

V0	nombre de mots par copie
V1	nombre de <u>verbes</u> en ER
V2	IR (type grandir)
V3	IR (type partir)
V4	RE et OIR
V5	au mode Indicatif
V6	Subjonctif
V7	Participe
V8	Infinitif
V9	au temps Présent
VA	Imparfait
VB	Passé simple
VC	à la personne 1
VD	2
VE	3
VF	4
VG	5
VH	6
VI	nombre de verbes autres
VJ	nombre de substantifs singulier
VK	pluriel
VL	nombre d'adjectifs singulier
VH	nombre d'adjectifs pluriel
VN	nombre de pronoms 1,2,3,4,5,6
VO	pluriel
VP	nombre de pronoms relatifs
VQ	nombre de pronoms personnels 0
VR	1
VS	2
VT	3
VU	4
VV	5
VW	6
VX	nombre d'adverbes
VY	nombre d'adverbes discontinus
VZ	nombre de conjonctions de coordination
W0	nombre de conjonctions de subordination

Remarques : 1) Certains temps simples n'apparaissent pas dans cette liste parce qu'ils étaient très peu, voire pas, représentés dans le corpus.

2) Les temps composés étant analysés dans la codification des textes n'ont pu être relevés dans cette première analyse.

Variables dites "syntaxiques"

V0	W1	nombre total de groupes
V1	W2	nombre de groupes Sujet
V2	W3	nombre de groupes Complément d'objet direct
V3	W4	Complément prépositionnel(indirect)
V4	W5	Attribut
V5	W6	Apposition
V6	W7	Circonstanciel "direct"
V7	W8	Circonstanciel prépositionnel
V8	W9	Apostrophe
V9	WA	Prépositionnel (comp. de non)
VA	WB	Adverbe
VB	WC	Comparaison
VC	WD	nombre de propositions Affirmatives
VD	WE	nombre de propositions Négatives
VE	WF	nombre de propositions Interrogatives
VF	WG	Impersonnelles
VG	WH	nombre de phrases par copie
VH	WI	nombre de subordinées
VI	WJ	nombre d'autres propositions

Ont été supprimés de même dans cette liste tous les types de groupes et de propositions qui n'étaient pas représentés.

Cette affirmation se trouve confirmée si l'on examine le plan du 3ème facteur qui fait éclater les couples 2ème et 3ème personne d'une part et 5e, 6e personnes d'autre part tout en faisant apparaître de nouvelles différenciations entre les individus. L'emploi des personnes 2, 3, 5, 6 dans ce corpus est un trait d'originalité.

La première personne pose un problème particulier. En effet, alors que par ailleurs l'emploi du verbe à une personne donnée est étroitement lié à celui du pronom à la personne correspondante, les pronoms de 1ère personne sont opposés aux verbes de 1ère personne, ce qui, à première vue est inattendu. Donc, personne des verbes et personne des pronoms ne sont pas nécessairement redondantes dans un corpus.

Les caractères les moins discriminants (voisins de l'origine) et par conséquent les plus communs de cet ensemble de textes sont ceux que l'on pouvait attendre a priori :

nombre de groupes

nombre de groupes complément d'objet

nombre de groupes sujet

phrases affirmatives

verbes en - ER

verbes en - IR (type partir)

indicatif.

- ... dominance de la phrase simple affirmative avec structure sujet-verbe - C.O.D., le verbe étant à l'indicatif et appartenant au 1er groupe ou 3ème groupe en - IR
- ... si la variable "participe" apparaît elle aussi proche de l'origine et donc générale, c'est que les temps composés ont été analysés et que l'emploi des temps composés est, lui aussi, une tendance générale. Dans le plan du premier facteur, le présent s'oppose au groupe passé simple imparfait, participe, dans la mesure où le présent s'oppose au passé.
- ... en ce qui concerne le groupe verbal, nous avons déjà noté la structure verbe - C.O.D. ; il faut ajouter la structure verbe + groupe adverbe, les deux étant voisines, compatibles et vraisemblablement cumulables.

Autres structures communes, mais opposées aux précédentes dans le plan du premier comme du second facteur : le groupe verbal avec attribut, avec complément d'objet indirect.

... enfin la coordination apparaît aussi comme un trait général ; il serait important ici de nuancer et de distinguer dans les données la coordination entre les groupes, entre les propositions, entre les phrases : la variable coordination représente en effet pour le moment le nombre total de ~~conjonctions de coordina-~~tion sans distinction ; il est difficile, dans ces conditions, de tirer sur ce point des conclusions sur la complexité syntaxique des phrases analysées.

... Par contre, on peut affirmer que la coordination est la seule transformation généralement appliquée par les élèves, avec sans doute la transformation impersonnelle, les deux étant d'ailleurs opposées dans le plan du second facteur.

Une relecture des textes permettrait de vérifier que le point impersonnel est plus éloigné de l'origine parce qu'il est attiré par le caractère présent, lui-même lié à l'interrogation. En effet, dans le plan du premier facteur, impersonnel est voisin du présent et s'oppose au participe, c'est-à-dire est peu compatible avec les temps composés ; il s'oppose à l'imparfait et au passé simple dans les plans du premier et du second facteur. On peut donc considérer que l'application de la transformation impersonnelle est une caractéristique commune avec un verbe au présent de l'indicatif, mais qu'elle est inconnue avec d'autres temps et d'autres modes.

... Avec les deux transformations simples d'interrogation et négation nous abordons les caractères nettement discriminants. En effet, l'interrogation comme la négation apparaissent comme des caractéristiques individuelles, l'interrogation étant liée plus étroitement que la négation à l'apostrophe, au présent et à la troisième personne.

Les deux sont incompatibles avec la 6^o et la 5^o personne des verbes (opposées dans le plan du premier facteur).

La négation n'est pas opposée à la 4^o personne sans en être voisine, mais est opposée à la 2^o personne. Dans le plan du 2^o et du 3^o facteur. C'est dire que l'interrogation et la négation non seulement ne sont pas généralement utilisées et sans doute sont utilisées de façon stéréotypée, sans grande variété, toutes les possibilités d'emploi n'étant pas exploitées.

... autre transformation : celle d'enchâssement. Il apparaît que la subordination ne peut pas être considérée comme une caractéristique générale de ce corpus. Mais il y a beaucoup plus.

Tout d'abord dans le plan du premier facteur, interrogation, subordination et négation varient dans le même sens : c'est dire que les élèves qui utiliseraient l'une auraient aussi tendance à utiliser l'autre, s'opposant à ceux qui n'en utiliseraient aucune. On pourrait conclure que la complexité syntaxique définie à ce niveau par l'application des transformations d'interrogation, de subordination, de négation et à la rigueur, impersonnelle, est un facteur important de différenciation des individus.

- 1 - que c'est une aptitude inégalement acquise
- 2 - que leur position le long de l'axe du premier facteur pourrait fournir une classification des individus sur ce point.

Toutefois deux autres oppositions permettent de nuancer cette première conclusion.

En premier lieu, ce que nous avons appelé subordination jusqu'ici est la ~~variable~~ "nombre de conjonctions de subordination". : il ne s'agit donc pas de la transformation d'enchâssement dans son ensemble, mais de certaines de ses réalisations, d'autres étant représentées par les relatives. Or les deux caractères "nombre de conjonctions de subordination". et "nombre de relatifs" sont très nettement opposés dans le plan du premier facteur, la relativisation étant un trait encore plus individuel que la subordination, et étroitement lié à la 6e personne ; c'est dire qu'il y a peu de variété sur le verbe de la proposition relative, mais surtout c'est suggérer qu'il y a deux types différents de subordination utilisés concurremment par les élèves et que l'emploi des propositions subordonnées relatives exclut celui des subordonnées conjonctives.

En second lieu, le second facteur montre que ce que l'on peut appeler complexité syntaxique au niveau de la phrase varie en sens inverse de la complexité à l'intérieur de la proposition : le facteur 2 oppose les transf. d'interrogation, de subordination, de ~~relativisation~~ qui vont de pair avec la phrase sujet-~~verbe~~complément, et les compléments circonstanciels et groupes prépositionnels, l'emploi des groupes prépositionnels et des circonstanciels prépositionnels étant des caractères discriminants.

On peut donc conclure que, les quelques caractéristiques de complexité syntactique présentes dans le corpus analysé

- 1 - sont des caractéristiques discriminantes et ne sont pas communes
- 2 - s'opposent et s'excluent entre elles, les élèves employant soit des subordinées conjonctives, soit des compléments circonstanciels, soit des relatives, une structure de phrase complexe excluant la complexité à l'intérieur de la proposition.

DEUXIÈME ANALYSE DÉTAIL. La même analyse a été faite ensuite en tenant compte de la diversité des textes, c'est-à-dire en effectuant les dénombrements à part sur chacun des trois sujets. Chaque élève figure donc trois fois dans la liste des individus, qui sont donc 270 au total, les variables restant les mêmes. Nous examinerons les résultats comparativement à ceux de l'analyse précédente.

Des cinq facteurs extraits, les trois premiers rendent compte de 41,75 % de la variabilité totale, un pourcentage qui n'était pas atteint avec quatre facteurs dans l'analyse précédente. Nous pouvons en conclure que les différences du corpus sont plus fidèlement et plus exactement décalées et analysées à partir de données non regroupées à priori.

Nous sommes aussi tenté d'en déduire qu'il existe des variations entre les textes correspondant aux différents sujets, ce qu'il faudra vérifier avec de nouvelles analyses.

... Naturellement, le trait de différenciation le plus net concerne les personnes du verbe et du pronom personnel, mais il est possible de nuancer les conclusions de l'analyse précédente. Dans le plan du premier facteur l'emploi des 2^o et 3^o pers. s'oppose à celui de la 5^o et de la 6^o, la 1^o et la 4^o restant des tendances communes et des faits de groupe. Toutefois, il faut noter que la 5^o et la 6^o pers. sont moins étroitement liées que l'on aurait tenté de l'affirmer d'après l'analyse précédente ; en particulier verbes et pronoms 5^o pers. d'une part, pronoms de 6^o pers. d'autre part sont assez éloignés dans les plans du 1^o et du 2^o facteur, et opposés dans le

04 15 DF

IFA 3

OK

21

1P

1K

1N

1R

1T

1U

1V

1W

1X

1Y

1Z

00 02

01 03

02 04

03 05

04 06

05 07

06 08

07 09

08 10

09 11

10 12

11 13

12 14

13 15

14 16

15 17

16 18

17 19

00 02

01 03

02 04

03 05

04 06

05 07

06 08

07 09

08 10

09 11

10 12

11 13

12 14

13 15

14 16

15 17

16 18

17 19

20

1

18

16

12

20

09

1B

1D

12

08

10

12

14

12

20

01

1F

13

15

12

2F

1E

13

15

12

2D

01

1E

13

15

12

1D

01

1E

13

15

12

1D

01

1E

13

15

12

1D

01

1E

13

15

12

1D

01

1E

13

15

12

1D

01

1E

13

15

12

1D

01

1E

13

15

12

1A

13

15

12

21

13

15

12

"SOMMES" FACTEURS 1 AND 2

Vb. 1^{re} pers.

13

Groupes prépositionnels
Vb. 6^{es} pers.

P. 6^{es} pers.

P. 12345
Autres Vb

G. Apposition

Attribut

Nb. phrases

P. pers.

Adj. sing.

P. relatif

Passé simple

Imparfait.

Circ. Prépositionnel

Interjections

P. 5^{es} pers.

Vb. 5^{es} pers.

Comparison

Interrogation

@1

Apostrophe

Négation

Adj. pluriel

P. 1^{re} pers.

Infinitif

Présent

P. 4^{es} pers.

Vb. 4^{es} pers.

Conj. subordination

Subjonctif

1

P. 2^{es} pers.

Vb. 2^{es} pers.

Impers.

Vb. "grandis"



plan du 3° ; l'ensemble de ces caractères s'oppose aux verbes de 5° pers. dans le plan du 2° facteur. Ce qui a priori peut paraître étrange, mais qui ne l'est plus si l'on considère que le point verbe à la 6° pers. est proche du point substantif pluriel et adjectif pluriel. On peut donc dire que la 5° et la 6° personne sont concurrentes surtout dans le cas où l'élève emploie pour la 6° pers. non point un pronom pers. sujet mais un substantif pluriel, un GN pluriel.

Pratiquement, on peut préciser que la personne du pronom et celle du verbe ne sont pas nécessairement redondantes.

... Enfin, en ce qui concerne les transformations d'enclassement, on voit se confirmer qu'il y a bien 2 types de subordination utilisés par les élèves de façon concurrente : l'emploi des relatives qui va de pair avec l'emploi des adjectifs, est un trait plus nettement individuel et opposé non seulement à l'emploi des subordonnées conjonctives, mais à toutes les autres transformations appliquées dans le corpus analysé. On peut donc conclure qu'une structure complexe du groupe nominal est incompatible avec une structure complexe de la phrase.

... En ce qui concerne les tendances communes, deux séries de conclusions s'imposent.

Tout d'abord on peut toujours affirmer la prédominance de la phrase simple affirmative avec sujet-verbe-complément d'objet direct, le verbe étant à l'indicatif et appartenant au 1er groupe, au 3ème en IR Mais

- a) les relations entre les temps des verbes peuvent être précisés
- b) les autres transformations semblent ne pas avoir a priori le même pouvoir discriminant.

Les temps de verbes. Alors que dans l'analyse précédente l'opposition présent-passé (sans doute d'ordre thématique) l'emportant sur toutes les autres, et que par ailleurs l'emploi des temps composés (variable "participe") apparaissait comme une tendance commune, ici au contraire, il apparaît que l'opposition principale (plan du 1° facteur) est entre le passé simple et l'imparfait d'une part, et le présent et les temps composés d'autre part.

La variable participe est en outre moins proche de l'origine et opposée à "passé simple" dans le plan du 1° et du 2° facteur, ce qui signifie que passé simple et passé composé s'excluent. (On ne dit pas plus-que-parfait car la proximité des points présent et participe montre que, dans ce cas surtout les verbes au participe ont des auxiliaires au présent).

Conclusion en apparence décevante puisqu'il s'agit d'une Lapalissade linguistique. Pratiquement, elle n'est pas dépourvue d'intérêt ; elle montre en effet que cette opposition aspectuelle est présente dans le corpus mais les conclusions deviennent des questions auxquelles devront répondre d'autres analyses sur des données plus précises : passé simple et passé composé s'excluent ; cette opposition est un critère de classement, mais classement des individus ou classement des textes ? Est-ce une opposition d'ordre thématique (imposée par le genre du texte, son contenu) ou d'ordre personnel (choix ou tendance individuelle) ? En d'autres termes, est-ce que l'emploi du passé simple peut être considéré comme caractéristique de certains individus, les autres employant de préférence le passé composé ? Est-ce que le choix de l'un ou de l'autre de ces temps est au contraire réellement régi par leur valeur ? Est-ce qu'il ne faut pas tout simplement y voir une opposition "morphologique", les élèves conjuguant de préférence le passé simple à la 4° et à la 6° pers., et le passé composé aux 2°, 3°, et 5° pers. ?

Pour répondre à ces questions, il faudrait déjà plus de précision dans l'identification des observations.

Quoiqu'il en soit, cette analyse sur des données "détaillées" a permis de faire apparaître que l'opposition présent-passé-sans grand intérêt linguistique puisque liée au thème - n'est que secondaire : elle n'apparaît que dans le plan du 3° facteur.

... Quant aux autres transformations appliquées, l'interrogation, la négation, l'impersonnel, leur caractère discriminant est très nettement moins marqué. Elles apparaissent non comme des tendances générales, mais comme des tendances de groupe. C'est un point sur lequel des données analytiques peuvent donner des résultats moins nets que des données synthétiques, surtout si l'on songe que l'on travaille sur des textes très courts et sur des données très rares. S'il s'agit de tendances de groupe, encore reste-t-il à déterminer s'il s'agit en priorité de groupes d'individus ou de groupes de textes.

Que conclure de la comparaison de ces analyses ?

Si les conclusions sont les mêmes dans les grandes lignes, l'analyse à partir de données analytiques est plus riche et plus fine : elle nuance sans la contredire jamais, l'analyse sur données synthétiques. Mais cette dernière a aussi son intérêt : dans un corpus somme toute assez homogène (voir image des points observations) le "grossissement qu'elle présente pour certaines oppositions (ici l'emploi des transf. d'interrogation, de négation, impersonnel, par exemple) peut fournir sinon un guide, du moins une aide dans le décryptage des résultats de "détail".
Nous ne trouvons pas inutile d'examiner les 2 séries de résultats en parallèle.

TROISIEME ANALYSE

Si l'on examine à part les variables dites grammaticales et les variables dites syntaxiques, de façon à neutraliser les éventuelles interférences des unes avec les autres, on constate :

- pour les variables grammaticales, les résultats sont exactement les mêmes que dans l'analyse précédente (avec 270 observations)
- pour les syntaxiques, les caractères nettement discriminants sont interrogation et apostrophe, qui s'opposent dans le plan du 1^o facteur aux compléments circonstanciels (prépositionnels et directs), compléments d'objet indirect et groupes apposition.
- l'emploi des groupes prépositionnels est aussi un caractère fortement discriminant opposé aux précédents.
- l'impersonnel est très proche de l'origine et donc bien une ton-dance commune.

L'intérêt limité - de cette analyse dissociée des variables "grammaticales" et "syntaxiques" est de confirmer les conclusions des premières.

Les limites viennent surtout de l'insuffisance des données et de leur répartition arbitraire entre les deux groupes de variables en effet, certains dénombrements de catégories grammaticales (conj. de subordination, de coordination, pronoms relatifs en particulier) devraient être traités plutôt comme des variables syntaxiques.

2 VG
5 pers.
VV

VT
2° of 30
pers.
VD VS
VE

V7 Parakeets
V3 St. og.

H8 line. Pup.

VP Relatip

Dr. Resp. 67
VW
VE

Purmit
V9/12

W6
Imped

VC Vb. 10 pers.
WA G. pup. at.

HO
any sub.

W7

HJ
advise

V19 VX
V2 HDM HV 14 4
V5 HAWB V2 any coord.

V6

Subjunct
WCO
confusion

H9
epitaphic

H5
WEX
Wegon
Interrogation

VQ
VA Impul
VM Adv. Plural

V6 Apposition

VF
4 pers
VU

VH vbs. 6. pers.

VB, Sample
Dates Sample

"DETAIL" FACTORS 1 of 2

Br. 63 pers. VN

VI

AS

VE 1631 pers

V6

VS 2131 pers

VD 1631 pers

MF

Parent V9

M9

Conj. Sub. W0

MG

VD

HBVX

V8

VI

VI

VA

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VH

VN

W0 Apention

VL

VP 2131 pers

W0 67 pers.

VI

VC

W0 67 pers.

WB

VB 2131 pers

Poste V7H7

VA 3131 pers

W0

V9 1631 pers

"DETAIL" FACTEURS 1443

QUATRIEME ANALYSE

Les analyses précédentes ont fourni une réponse au premier groupe de questions que nous nous posions : parmi toutes les variables dénombrées, c'est-à-dire les caractéristiques linguistiques envisagées, il a été possible de distinguer

- les caractéristiques communes du corpus
- les caractéristiques discriminantes, c'est-à-dire celles qui permettent de préciser l'originalité linguistique de certains individus ou groupes de textes.
- les caractéristiques qui, dans ce corpus, sont incompatibles.

Une question reste posée : pour les mêmes caractéristiques linguistiques, y a-t-il des différences selon le sujet traité ?

Il y aurait deux façons de chercher la réponse :

1) dans l'analyse détail, il suffirait de projeter sur le graphique des variables, le graphique des observations où chacun des 270 points représente à la fois un élève et un sujet différent. On pourrait voir ainsi si les groupes qui s'opposent sont constitués par des individus différents des textes différents, ou éventuellement les deux. Dans ce dernier cas, il serait vraiment difficile d'en tirer des conclusions pratiques et de déboucher sur des applications pédagogiques.

2) Une autre solution consisterait à traiter séparément les trois sujets différents et à mener en parallèle trois analyses avec les mêmes variables dénombrées chaque fois sur un sujet différent. C'est cette dernière méthode que nous avons choisi d'employer ici. Le sujet trois présente la plus grande dispersion des individus, c'est-à-dire que c'est avec le sujet libre que l'on obtient dans les textes la plus grande diversité linguistique.

Paradoxalement, en apparence du moins, la plupart des variables sont très groupées autour de l'origine. Ceci ne signifie pas qu'elles soient sans intérêt, mais qu'elles constituent des caractéristiques assez communes, leur pouvoir discriminant étant faible par rapport à d'autres traits qui permettent d'isoler du lot certains individus et sont donc responsables de la dispersion. La diversité linguistique repose sur la diversité d'emploi des personnes du verbe : la 5° exclut la 3°, à un moindre degré la 2°, à un moindre degré la 4°.

En ce qui concerne la 6° un fait reste à expliquer : l'emploi des pronoms personnels de 6° personne est fortement discriminant, et paraît un fait individuel. Il faut noter que sous l'étiquette "pron. pers.", quelle que soit la personne, on regroupe les pronoms sujets et les pronoms compléments. Il n'est donc pas étrange que le trait pronom personnel 6° pers. soit opposé au trait "verbe 6° pers." Mais ce qui surprend à première vue, c'est que le trait verbe à la 6° pers. soit opposé au trait substantif pluriel, comme d'ailleurs au trait "substantif singulier". L'explication est simple : les substantifs se rencontrent en fonction de sujet, mais aussi en fonction de complément. Si l'on examine le graphique, on constate non seulement que la structure sujet-verbe-complément d'objet direct est commune, plus que les structures avec d'autres types de compléments, mais surtout on remarque que les traits-emploi de la 3° pers., interro., verbes 6° pers.,

groupes sujets et C.O.D.

sont opposés dans le plan du 1° facteur aux traits

- emploi de la 5° pers. pron. pers 6° pers.
- substantifs plur. subst. sing. groupe
- prép., circons. prép. circ : direct C.O.I.

ce qui oppose une structure que l'on peut considérer comme simple à une structure plus recherchée surtout dans l'utilisation du groupe nominal.

Par ailleurs, présent et subordination sont dans le même sens que les structures simples, tous les temps du passé étant opposés, ce qui tend à montrer que l'opposition présent-passé n'est pas uniquement thématique mais correspond sans doute à une plus grande souplesse dans le maniement de la langue. Que simple s'oppose à recherche est une évidence ; la détermination des caractères qui permettent de définir la simplicité et la recherche n'est pas dépourvue d'intérêt, enfin la dispersion des individus prouve que le sujet libre est un bon terrain d'expérience pour juger des possibilités de chacun.

Ayant fourni des clés pour la lecture des graphiques, nous ne nous attarderons pas sur l'interprétation des résultats de l'analyse des sujets 1 et 2. On peut voir tout de suite que les groupements des variables et les traits discriminants ne sont pas les mêmes.

Quelles sont donc les conclusions que nous pouvons tirer de cette comparaison ? - Dispersion des individus beaucoup plus grande avec le sujet libre

- Avec le sujet 1, outre le fait que les individus sont peu dispersés et par conséquent qu'il n'est pas possible de déterminer leur originalité linguistique, il faut noter que seul le premier facteur est important, il y a au fond une seule ligne de variabilité.

Avec le sujet 2, les individus sont aussi moins dispersés qu'avec le sujet trois, mais la variabilité dont on peut rendre compte est plus importante dans la mesure où il y a une opposition suivant le premier facteur, mais aussi suivant le second, prouve qu'un assez grand nombre de caractères sont discriminants.

Il reste maintenant à en tirer des conclusions et des applications pédagogiques, et pour ce faire en parfaite connaissance de cause sans doute serait-il précieux, comme nous l'avons dit plus haut, de compléter les analyses à la lumière de ces premières conclusions.

V6: Vb. 5th pers.

VV Pr. 5th pers.

v6 Subjunctif

Subj V1 2nd Vb 6th pers

MO Conj. sub. V1 2nd Vb 6th pers

V9 Preterit (ME Nipalun V1)

V6 5th pers

VT Pr. 3rd pers

Participle V7

VP Relative

Anterior part VJ

Impt VA

Adj Sing V1

V1 Sub. V3 ND V3

V1 V3 V4 V5 V6 V7 V8 V9

V1 V2 V3 V4 V5 V6 V7 V8 V9

V1 V2 V3 V4 V5 V6 V7 V8 V9

V1 V2 V3 V4 V5 V6 V7 V8 V9

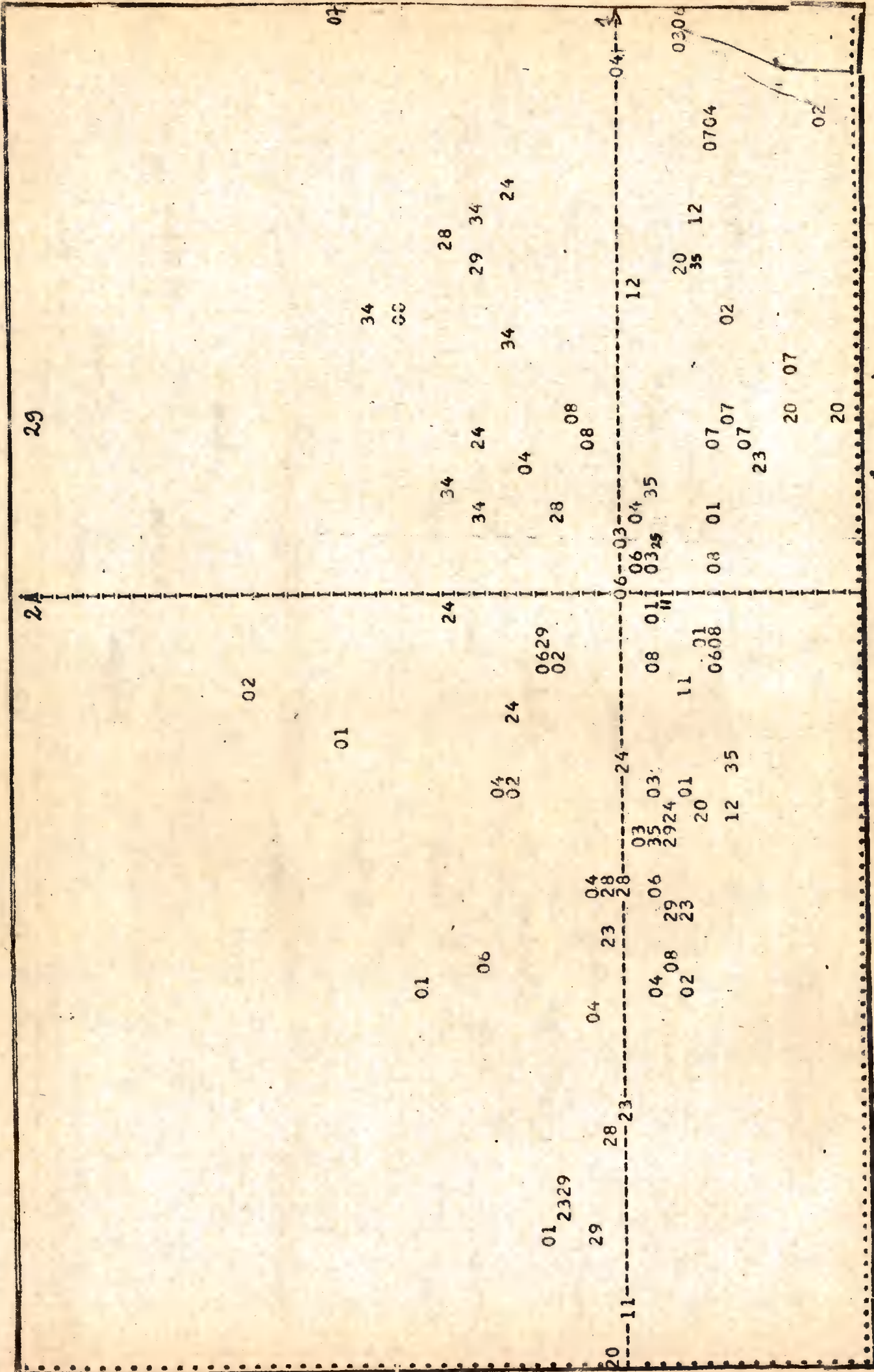
V1 V2 V3 V4 V5 V6 V7 V8 V9

V1 V2 V3 V4 V5 V6 V7 V8 V9

V1 V2 V3 V4 V5 V6 V7 V8 V9

VC

VU 2nd 3rd pers



SUJET 1 FACTEURS AER2 (Observations)

In. pno. 63p VW
VN D-12345
VI Quotas Vb.

VB 28 pers
VE MF investigation
V6 VT 28 28 pers
Subj mech

add sy
VL VC WA G 600p
WB Lic. pref.

VP R
VW
Vb. 55 pers

Subj mech
V9 VXC
WB
V3 28
V2 V5
V4
V8
V10
V11
V12
V13
V14
V15
V16
V17
V18
V19
V20
V21
V22
V23
V24
V25
V26
V27
V28
V29
V30
V31
V32
V33
V34
V35
V36
V37
V38
V39
V40
V41
V42
V43
V44
V45
V46
V47
V48
V49
V50
V51
V52
V53
V54
V55
V56
V57
V58
V59
V60
V61
V62
V63
V64
V65
V66
V67
V68
V69
V70
V71
V72
V73
V74
V75
V76
V77
V78
V79
V80
V81
V82
V83
V84
V85
V86
V87
V88
V89
V90
V91
V92
V93
V94
V95
V96
V97
V98
V99
V100

W9 6*
MC φ

WJ Dep.
WE Neg.

WC Conf. V0 Motis
VS
VJ 18 28 28

V3 =
V6 4 pers
VF
V3 +

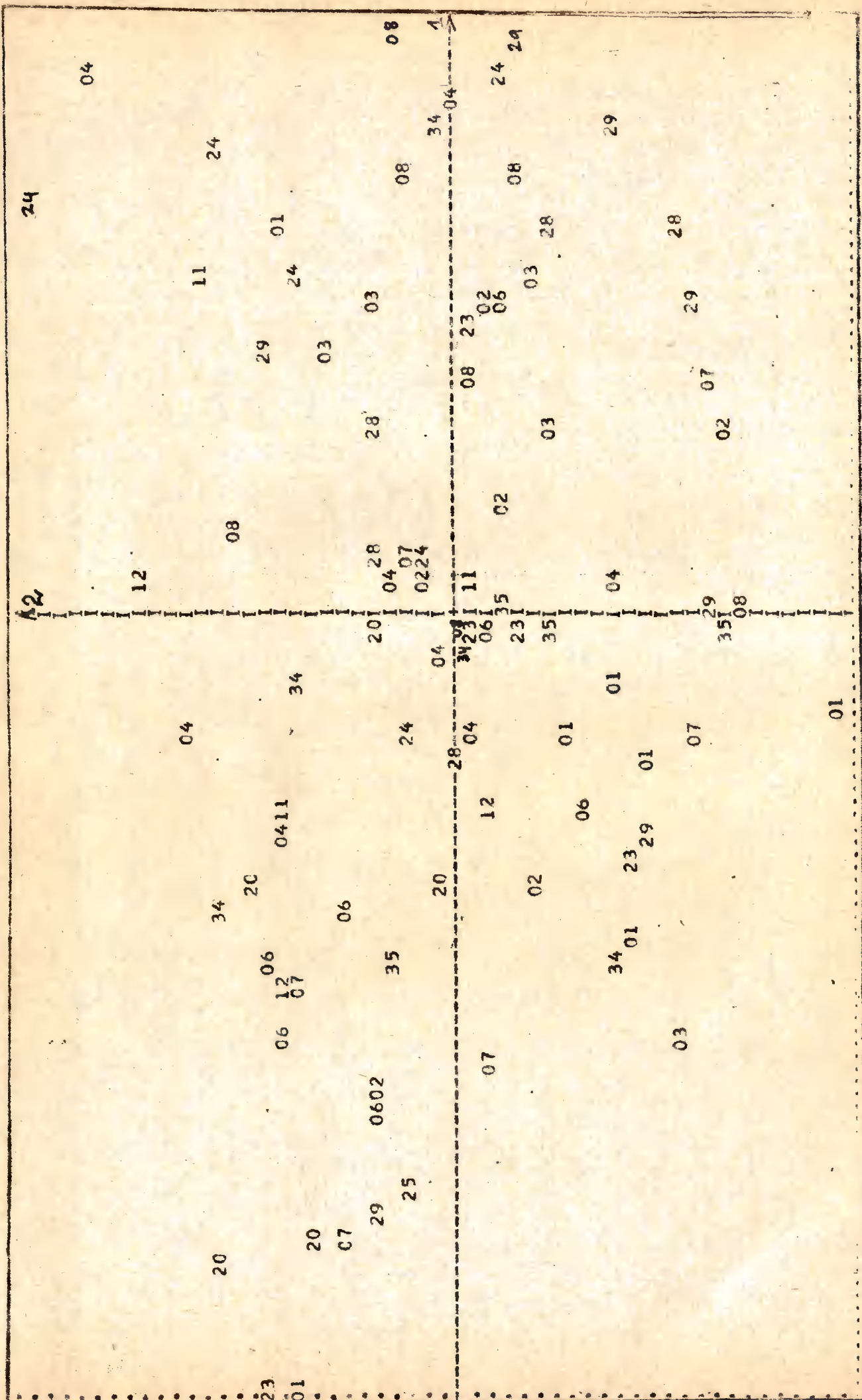
W7 Lic.
V7 28 28 28
VJ 28 28 28

VU
28 28 28

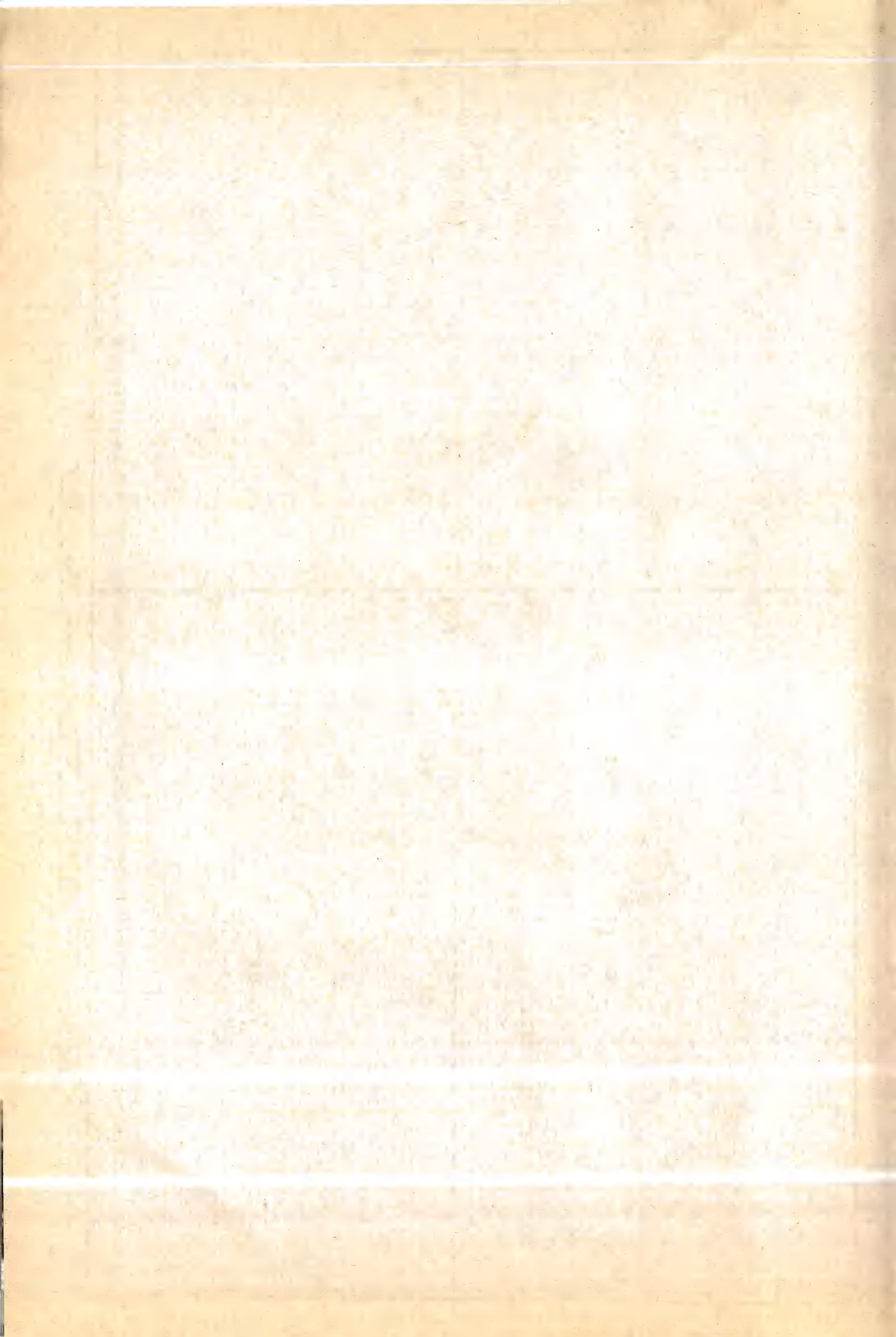
VA 38 28

VB 28 28 28

SUBJECTS - INVESTORS 6001 V0 28 28



SUJET 3 : EACREURS 1er2 Observations



REMARQUES SUR LA SPECIFICITE DU LANGAGE PARLE

D. WEIL - Université Louis Pasteur Strasbourg
U.E.R. des Sciences du Comportement et de l'Environnement
Section de Psycho-linguistique.

Je suis un petit peu gênée parce que le travail dont je vais parler n'est pas comparable aux travaux dont on a entendu parler ce matin, ceci en raison de leur ampleur et du premier état d'achèvement dont ils témoignent. De plus mon travail porte sur un corpus de langage parlé.

Je ne suis actuellement pas en mesure de vous présenter l'analyse du corpus à laquelle je travaille encore, mais je peux simplement faire quelques remarques qui pourraient contribuer à cerner certaines caractéristiques du langage parlé - langage parlé par des enfants pour ce qui nous intéresse ici, mais aussi peut-être langage parlé en général.

L'objet de mon travail a différents aspects :

D'une part, il s'agit de décrire du langage parlé, du point de vue syntaxique et organisationnel ; ce qui implique de résoudre les problèmes d'analyse que pose le langage parlé à l'observateur qui dispose des instruments conceptuels et méthodologiques de la linguistique, plus adéquats à l'approche de la langue écrite qu'à celle de la langue orale.

D'autre part, il se propose de dégager l'éventuelle spécificité du langage parlé tel qu'il apparaît dans le corpus réuni, par opposition à la fois aux modèles de l'écrit réalisés dans des textes, et à un corpus de langue parlée par des adultes.

Autrement dit, j'essaie de distinguer, à partir d'une description linguistique, ce qui, dans le langage parlé par des enfants d'âge scolaire peut être rapporté à l'opposition enfant/adulte, et ce qui doit y être rapporté à l'opposition oral/écrit.

Ceci pose problème car les catégories enfants/adultes et oral/écrit, pourtant couramment utilisées, attendent encore une claire définition. En effet, il n'est pas exclu que lorsqu'on oppose le "langage de l'enfant", au "langage de l'adulte",

on ne se représente ce "langage de l'adulte" comme une entité homogène à partir d'un modèle de la langue écrite, actualisée par exemple dans des textes d'auteur. On compare ainsi implicitement des faits de langage, isolés comme "langage de l'enfant", à des faits d'écriture.

Il faut faire ici plusieurs remarques :

- 1) le langage de l'enfant se situe pendant très longtemps (jusqu'à l'école élémentaire) dans le champ de l'oralité, mais d'une oralité partagée avec des adultes.
- 2) les manifestations linguistiques des adultes ne sont pas homogènes et varient selon qu'ils parlent ou qu'ils écrivent (c'est-à-dire qu'ils élaborent un texte).
- 3) le langage parlé des adultes est déterminé par le caractère (formel, familier, etc...) de la situation où il s'insère.
- 4) il intrique des formes spécifiques de la langue parlée et des formes héritées des modèles de l'écrit, dans des propositions fortement déterminées par des facteurs sociaux-culturels et de scolarisation.

Il est donc possible que ce qui apparaît comme une opposition enfant/adulte à l'intérieur des faits de langage, recouvre plus ou moins largement une opposition oral/écrit, elle-même obscurcie par tous les faits d'éducation qui marquent l'histoire d'un sujet, et par des importantes variations individuelles.

Caractériser linguistiquement l'oral et l'écrit ne semble peut-être le plus fécond pour débiter l'étude du langage de l'enfant, étant donné que le langage de l'enfant est d'abord "langage parlé" avant d'être "langage d'enfant", et que l'opposition enfant/adulte renvoie à trop d'a priori théoriques autres que linguistiques pour que l'analyse du langage n'en soit pas altérée.

Avant d'énoncer quelques caractéristiques syntaxiques et organisationnelles les séquences de langage parlé par des écoliers qui constituent mon corpus, je vais rapidement souligner quelques points de méthode.

- 1) J'ai rencontré beaucoup de difficultés dans le découpage du corpus en "segments", en référence à ce que l'on trouve

défini comme "phrases" dans les grammaires : à la fois des unités de discours, de l'ordre du concret, et unités grammaticales, de l'ordre de la description formelle. Isoler ses segments dans une séquence de langage parlé implique le recours à une diversité de critères dont aucun n'est en soi pleinement satisfaisant : syntaxique, sémantique, prosodique.

On pourrait penser que dans le champ de l'oralité, les critères relevant de la prosodie sont prépondérants. Cependant les pauses, et les courbes mélodiques ne recouvrent pas forcément les unités syntaxiques ou sémantiques qu'on définirait comme "phrases". Pour énoncer, en langage parlé, une unité de sens ou une unité syntaxique, il faut souvent plusieurs énonciations successives entrecoupées de pauses plus ou moins longues.

Le critère sémantique ne paraît pas pertinent non plus à lui tout seul : parce que l'unité de sens peut être portée par un segment qui d'un point de vue syntaxique n'est qu'un fragment de phrase.

De même un critère syntaxique exclusif obligerait à éliminer nombre de segments faisant pourtant partie intégrante de la séquence.

Il a donc été décidé (et ceci constitue un choix qui peut être discuté) de fragmenter la chaîne parlée en segments dont la forme et les limites seraient réglés par la relation de dépendance de leurs constituants au prédicat :

- un segment est appelé phrase si tous les compléments nécessaires au prédicat y apparaissent ;
- un segment est appelé fragment si des termes dépendants nécessaires au prédicat n'y sont pas présents.

2) La nature du corpus (voir annexe) soulève elle aussi des problèmes.

Il s'agit d'un extrait d'un corpus assez étendu qui avait été réuni pour mettre en évidence la diversité des types de discours selon les situations de parole dans une même population d'écoliers.

- Dans cet extrait, la particularité énonciative des quelques locuteurs concernés s'y déploie, et l'on ne peut pas en faire abstraction lorsqu'on tente de caractériser le discours. Elle interdit par ailleurs les généralisations hâtives.

Ainsi les comptages que l'on peut effectuer doivent être rapportés non seulement au nombre total de phrases du corpus, mais encore au nombre de locuteurs concernés.

- On y découvre aussi l'importance de la situation de parole comme cadre déterminant les formes du discours. Ainsi les trois premières séquences citées en annexe sont tirées d'un ensemble d'interviews que j'ai réalisés dans des écoles en dehors des classes, et où ma demande a été perçue par les enfants comme semblable à la demande du maître. Ce qui s'est trouvé reproduit dans ces interviews de groupe, c'est la "mise en route" du lundi matin, désignée dans de nombreuses classes comme "conversation libre" où les enfants racontent leur dimanche et où le récit débouche sur un jeu plus ou moins formel de questions-réponses.

Les trois extraits suivants proviennent d'enregistrements d'enfants effectués dans leur famille, dans une situation de jeu, de toilette ou de repas, en l'absence de tout adulte, et sans qu'on leur ait demandé quoi que ce soit quant au contenu ou à la forme de leur parole. L'enjeu des conversations enregistrées est donc très varié :

- dans le premier extrait on assiste à un échange linguistique assez insolite entre deux enfants au cours de leur repas, qui consiste en une élaboration autour du regard, de ce qui voit et de ce qui se voit ;
- dans le deuxième extrait, deux garçons font le bilan de leur journée de ski, et produisent à deux un "récit" fort différent de leurs productions enregistrées à l'école ;
- le troisième extrait met en scène deux enfants qui tentent d'expliquer au troisième ce jeu de questions/réponses où l'on doit découvrir une anomalie, chaque joueur répondant à la question qui leur est posée en donnant la réponse adéquate à la question précédemment énoncée.

Ce corpus doit être comparé à un corpus de langage parlé par des adultes, décrit par Mme Simone Monsonego du CIRS à Nancy, lui aussi extrait d'une transcription intégrale des séances d'un groupe de diagnostic auquel participèrent autrefois de jeunes cadres d'entreprise. C'est une situation très particulière de discussion mais qui a l'avantage de laisser les participants libres de leurs contenus et de la forme de leurs échanges.

Il nous avait semblé dans un premier temps trouver là un corpus de langage parlé d'adultes susceptible de soutenir la comparaison avec celui des enfants, du point de vue des situations de parole. Cela s'avèrera peut-être à discuter.

Je vais tenter de vous citer quelques caractéristiques syntaxiques et organisationnelles des séquences de langage parlé par les enfants, celles qui m'ont semblé pouvoir alimenter un débat sur les caractéristiques de l'oral, et par différence, de l'écrit.

1. Les types de phrases

On note une réelle préférence pour les phrases simples sur les phrases complexes (leur forme étant très diverse : verbales, non verbales, à présentatif...) aussi bien chez les enfants que chez les adultes, pourvu que l'on veuille bien compter à part chez les adultes les phrases verbales commençant par "il me semble que", "je crois que", "je pense que" etc... qui ne sont pas vraiment des phrases complexes dans la mesure où le syntagme de discours qui les inaugure fonctionne comme une modalité stéréotypée de prise de parole.

Il faudrait voir si cette préférence est un trait caractéristique du langage parlé. Mais il faut remarquer que si la catégorisation des phrases en phrases simples et complexes nous renseigne sur l'articulation des propositions, elle ne constitue guère un bon indice de complexité de la phrase. Pour cela il faudrait analyser les constituants grammaticaux et lexicaux de ces phrases simples, ce qui est en cours.

Ainsi nombre de phrases simples sont des phrases segmentées du type "mon père il a fait ceci ou cela" ou encore "ce que je peux te dire, c'est que..." Il semble que les phrases segmentées répondent à un projet rhétorique et il faut souligner que leur répartition n'est pas homogène dans le corpus et qu'elles caractérisent certains locuteurs.

2. Le statut des phrases

Dans la conversation, il semble que la reconnaissance du statut des phrases énoncées par l'interlocuteur, soit intuitivement assumée sans problèmes même pour de jeunes enfants.

Cependant la description du statut des phrases (simplement répertoriées par la grammaire en phrases déclaratives, interrogatives, impératives) en langage parlé est extrêmement problématique, car les critères de reconnaissance semblent se situer à plusieurs niveaux :

- des critères linguistiques internes à la phrase (y compris des critères prosodiques)
- des critères discursifs dépassant le niveau de la phrase
- des critères extralinguistiques qui font référence au type de communication, au nombre des locuteurs en jeu etc...

Par ailleurs le statut de nombre de phrases est ambigu, chevauchant plusieurs catégories ; et une quatrième catégorie - celle de l'exclamation - est fortement représentée en langage parlé, mettant en jeu des critères tout à fait spécifiques.

Prenons l'exemple suivant, tiré du premier extrait en annexe : "enfin, mais puisque c'était démolé, tu ne l'as pas redéfait". Un certain nombre d'indices sont en faveur d'une phrase exclamative :

- "enfin" et "mais", éléments lexicaux, marquent en tête de phrase, la position de Pierre par rapport à ce qui vient d'être dit (= référence au discours) et indiquent qu'il apporte la contradiction.
- "puisque" (et non seulement "comme" ou "parce que"), conjonction donnant à la subordonnée un caractère de preuve plutôt que de cause.
- l'intonation de la subordonnée (courbe descendante, accentuation de "puisque").
- forme déclarative de la principale "tu ne l'as pas redéfait".

Mais en fin de phrase, la courbe mélodique se modifie et prend un contenu interrogatif, et le ton d'abord intempestif devient de moins en moins assuré. Ces indices vont dans le sens d'un statut interrogatif intriqué au statut exclamatif.

La diversité des indices ici en concurrence, et la variété des niveaux auxquels ils se situent (situation - discours phrase -) me semblent spécifiques de la langue parlée, l'écrit visant justement à privilégier certains types d'indices.

On pourrait faire le même travail d'analyse sur l'exemple tiré du 3^e extrait : "mais les cochons d'Inde, c'est pas la même chose que les hamsters" qui oscille entre le statut déclaratif-exclamatif et le statut interrogatif : du point de vue syntaxique cette phrase présente les caractéristiques de la phrase déclarative mais l'intonation est celle d'une interrogation au sens non de demande d'information (cf le discours auquel renvoie le "mais" introductif) mais de demande d'approbation.

De même dans l'exemple n° 6 "alors tu poses des questions, t'as pas le droit de continuer, mais alors tu ne dis pas autre chose hein..." où l'on a deux parties, la première, déclarative voire impérative (forme linguistique déclarative prenant un statut impératif dans le contexte situationnel) et la seconde, où l'adverbe "hein" et le changement de la courbe mélodique introduisent un élément interrogatif portant effet rétroactivement sur l'ensemble de la phrase.

3. L'organisation interne de la phrase

Divers de ses aspects m'ont semblé caractériser le discours parlé : - la reprise ou la répétition d'un groupe fonctionnel donné, à l'intérieur d'une même phrase.

Par exemple (extrait n° 1) : "J'ai trouvé ça beaucoup mieux que en France /comme ça/ aux autres endroits" ou encore (extrait n° 3) "on les mettait tout le temps dans un parc/ une espèce de parc en bois avec un grillage dessus/dans l'herbe..." ou encore (extrait n° 1) "et après en pleine fête y avait ma soeur qui/Corine/qui voulait jouer..."

On assiste là à une élaboration de l'énoncé, au fur et à mesure de l'énonciation, élaboration qui disparaîtrait derrière l'énoncé achevé, à l'écrit. L'après coup de l'élaboration se manifeste forcément, à l'oral, dans cette sinuosité de ces fragments intégrés successivement à la phrase.

- l'ordre des unités de discours, pour les mêmes raisons d'élaboration hic et nunc sans doute, se trouve bousculé par rapport

à l'ordre logique de la pensée rationnelle, ordre logique qui serait reconstitué à l'écrit.

En voici un exemple (extrait n° 1) : "et après il y avait aussi le monsieur/le propriétaire du mas/et il a/il a aussi participé à ce dernier jeu/il s'appelait Denis Colon". Ce retour, en fin de phrase, sur l'identité de l'actant est très bien admis à l'oral, mais serait apposé au groupe "le propriétaire du mas" à l'écrit.

- un autre aspect est la punctuation du discours par les expressions idiolectales et les modalités de prise et de maintien de parole.

La punctuation est donnée par des expressions comme "tu vois", "comme ça", "hein" etc...

La prise de parole est souvent marquée par des morphèmes adverbiaux ou des adverbes de discours : "et puis", "alors", "et ben", "bon", "alors voilà", "bon alors voilà", "euh"...

Ces éléments ont fréquemment en plus une valeur rhétorique : ils marquent l'attitude du locuteur devant la demande de parler ou expriment son désir propre d'intervenir, ainsi que son désir à maintenir l'attention de l'autre.

On retrouve cette punctuation et des modalités de prise et de maintien de parole dans le langage parlé des adultes, portées par des syntagmes différents : "je pense que", "il me semble que", pour la prise de parole, "n'est-ce-pas", "hein vous voyez", "vous pensez?", "vous êtes sûrs", pour la punctuation.

Tous ces traits, rapidement répertoriés ici, me semblent marquer l'organisation linéaire du discours parlé, en une succession de segments le plus souvent syntaxiquement indépendants, mais non autonomes les uns par rapport aux autres dans le discours.

À l'écrit au contraire, la dépendance des segments dans le discours est marquée par une dépendance syntaxique explicite des segments, à l'intérieur de la phrase, et d'une phrase à l'autre.

Ainsi l'écrit ne peut être conçu comme une transcription de l'oral. Il n'est pas soumis aux mêmes règles d'organisation que l'oral.

Il est évident que ces traits de l'organisation du discours parlé sont très variables et qu'ils peuvent varier d'un individu à l'autre, d'une région à l'autre, d'une époque à l'autre.

Et cette différence découle de la position du locuteur par rapport à ses énoncés, à l'oral et à l'écrit : l'organisation du texte écrit est rendue possible par la distance du sujet à son texte comme produit, et par l'élaboration après coup d'une première énonciation, à travers la mobilisation de structures de phrases propres aux textes littéraires, aux textes journalistiques, aux exposés didactiques. Cette élaboration recouvre en fait la première énonciation d'un vêtement de pensée rationnelle et particulièrement soumise aux conventions sociales, masquant de ce fait l'irruption du sujet-parlant dans ses énoncés.

Ainsi les conditions propres de l'oralité et de l'écriture définissent des modes d'organisation des énoncés radicalement différentes. Le passage de l'oral à l'écrit et de l'écrit à l'oral ne peut donc être direct, immédiat : il en passe par la soumission à ces contraintes et à ces règles spécifiques.

La démarche pédagogique dont le double but est l'approfondissement de la langue maternelle, et l'initiation à la lecture et à l'écriture de textes dans cette langue, ne peut donc s'établir sur l'idée d'une simple transcription d'un registre dans l'autre.

Il faudrait étudier particulièrement les voies par lesquelles l'enfant reconnaît le langage parlé dans les phrases les plus élémentaires où il apprend à lire : en effet, ce sont souvent des phrases qui, à l'instar des phrases des traités de linguistique, ne sont pas susceptibles d'être jamais dites par quiconque. Au moment de l'apprentissage de la lecture, l'univers linguistique de l'enfant est le langage parlé, avec ses structures syntaxiques spécifiques et ses modes d'organisation du discours. C'est donc à une "écriture" et non à une transcription d'oral que l'enfant est d'emblée confronté lors de ses premiers déchiffrages, et continuellement par la suite. Et lorsque l'écolier fait ses premiers essais d'écriture de phrases, ce sont des énoncés qui bien qu'ayant d'abord été dits, sont des énoncés "aménagés" par rapport au langage habituellement parlé, avant d'être portés au tableau pour constituer des modèles. L'enfant a donc toujours à assumer une coupure entre deux ordres du langage, et il serait extrêmement fécond pour la pédagogie de

pouvoir la caractériser d'un point de vue linguistique d'une part, et d'étudier le cheminement psychique par lequel le jeune sujet-parlant prend conscience de ce hiatus et l'assume d'autre part. Car cette démarche conditionne sans doute son propre accès à l'écriture et par ailleurs le développement de sa connaissance des pratiques auxquelles donne lieu sa langue maternelle dans le champ social dont il participe, ces deux aspects s'interpénétrant continuellement, évidemment.

INDEX

Les signes / // désignent les pauses de plus en plus longues, repérées sur la bande d'enregistrement.
Le signe * ... donne des repères de donnée indiqués par ailleurs en marge (extraits 4, 5, 5).

1 - B 77 - Pierre (8 ; 9)

Alors ces vacances d'été / je suis allé en Camargue et j'ai fait du cheval / là-bas / et j'ai trouvé ça beaucoup mieux que en ... en France comme ça aux bons endroits... les selles là-bas sont un peu spéciales / ce sont des sortes de chaises / un peu ... et on a ... on a fait des promenades d'une journée / comme ça / on est allé sur des plages se baigner ... on a vu / moi j'ai vu des cormorans / des aigrettes / des hérons / des flamands roses / de tout... i'z étaient tous en liberté (un canarade : où ils étaient ?).

- i'z étaient dans des marais avec des roseaux / des marais avec 10 cm d'eau

(J. Daniel : est-ce que tu as vu des taureaux ?)

- oui j'ai vu des taureaux sauvages / en liberté ...

(J. Daniel : ... est-ce qu'y avait des ... est-ce qui'z étaient à cheval pour les attraper ?)

- non / on les laissait pour des corridas / les choses comme ça ... j'ai vu un p'tit bout de rodéo moi ... mon père a vu tout un rodéo / mais nous on était parti pour une promenade d'une journée alors / on est arrivé et on vu seulement la fin ...

(raconte nous ça ...)

- la fin? / eh bien c'étaient deux équipes de monsieurs / qui étaient à cheval / et alors ils devaient aller chez une dame / et la dame lui donnait un bouquet et ... alors ils devaient se promener dans l'arène / et puis l'autre monsieur de l'équipe ennemie / il devait essayer de lui attraper le bouquet ... et celui qui avait l bouquet / il devait avoir le bouquet à la hauteur de la tête / alors on les voyait / i'f'saient des virages serrés et tout... et après / y avait aussi le monsieur / le propriétaire du mas / et il a / il a aussi participé à ce dernier jeu ... i' s'appelaient Denis Colon...

(Pierre 0. : est-ce que c'était sur de l'herbe ou en pleine Camargue / avec de ...).

- ben / c'était sur de la terre ...

(Pierre 0. : molle ?)

- non / pas très molle / non / plutôt dure / y avait des arbres...

(J. Daniel : les taureaux / ils les chassaient le samedi ou le dimanche ?)

- ah ça / j's ... j'peux pas t'dire ... mais c'que j'peux te dire / c'est qu'à la fin du rodéo j'ai vu trois taureaux se promener à 10 m du nas / tu vois / comme ça ...

(J. Daniel : i't' fonçaient pas d'essus ?)

- ben justement / le rodéo / mon père m'a dit y avait une ficelle autour du poitrail du taureau / alors on devait s' tenir avec une main sur le taureau et on devait y rester / 10 secondes ...

(Pierre 9. : ton père / il est allé d'essus ?)

- oh / t'es fou eh / t'es fou quand même ...

D 136 - Raoul

Et ... pendant les vacances nous avons construit un golf dans la cour ... et ... avec des bouts d'bois... et on s'est amusé pendant toutes les vacances à faire du golf... Et ... après / on avait fait une fête et ...

Laurent : pour l'inauguration ?

Raoul : mais oui... et ... après ... en pleine fête / y avait ma soeur qui / Corine / qui voulait jouer et elle avait pris une pierre ... alors elle a tiré dans le golf et tout est cassé...

Et après la rentrée j'ai tout redéfait ...

Pierre : fin / mais puisque c'était démolé / tu n'l'as pas redéfait ?

Raoul : enfin / ce qui en restait ...

Pierre : ah oui tu l'as ...

Thierry : les déchets ... pour pouvoir en refaire l'année prochaine.

Raoul : ah oui / mais j'ai un grand tas de bois... oh mais y en a dans la cour aussi / y en a beaucoup...

Pierre : eh bien ...

Pendant les vacances j'étais dans les Cévennes / près d'Alès / à trente km à peu près... et ... nous avions deux cochons d'Inde... alors on en avait acheté un y avait à peu près deux ans / d'puis deux ans on l'avait / et puis on en avait / on nous en avait donné un / et pis y avait un (femelle et un mâle... et alors euh... on les mettait tout l'temps dans un parc / une espèce de parc en bois avec un grillage dessus / dans l'herbe / et pis une fois / alors on n'avait pas vu qu'y avait un trou sous le... sous le parc et alors d'abord y a le mâle qui est parti... il est parti dans les ronces et pis après la femelle est partie... et ils se sont réfugiés dans un / dans un tas d'fumier qui était au sec quoi / qui avait des orties dessus / et pis y avait des ronces comme ça / qui tombaient en cheveu / et i's étaient réfugiés là... alors on était tristes parc'qu'on les avait plus / on les trouvait plus / et pis on croyait qu' c'étaient les chats qui les avaient mangés... et a'ors euh... un jour on s'prom'nait comme ça / pour ramasser des poires / et on a vu une grosse boule brune qui qui défilait / alors on a demandé c'que c'était et puis on a vu après / qu'y avait deux boules près du tas d' fumier / alors on a dit / ça y est / c'est eux... et alors on voulait les chercher... alors i' z'étaient déjà habitués euh... à la brousaille... alors dès qu'on voulait les attraper i' nous mordaient ... Alor'on était obligé de prendre des gants et puis i'z'étaient dev'nus sauvages... et comme y avait des ronces c'était pas très facile... et alors euh... finalement on les a retrouvés... alors après on n'a plus pu les mettre dans l'parc... on a fait un autre parc / plus grand et plus... on les a rapportés après / à la maison...

Dominique : et vot' parc / qu'est-ce que vous en avez fait ?

Bruno : ben / on l'a détruit / on a gardé les planches...

Dominique : et avec les planches qu'est-ce que vous avez fait ?

Thierry : ben / l'aut'parc

Bruno : ben oui / on a mis des planches supplémentaires pour faire plus haut et plus grand...

on les achète normalement ... y a des camions qui viennent chez nous pa'c'que c'est des maisons en ruine qu'on a ...

et alors euh... mes parents les r'construisent et des ouvriers...

et les toits . c'est pas des toits en tuile / c'est des toits en lauze

Thierry : c'est quoi la lauze ?

Bruno : c'est des espèces de pierres plates / de grandes pierres plates / et qu'i' faut planter avec des clous... alors on fait un trou d'dans et on plante le clou pour qu'elles tiennent...

alors on met d'abord les poutres / et puis les planches dessus / on les cloue / alors les planches il en restait beaucoup pa'c'qu'on a beaucoup d'autres toits à faire... et après on mettait du ... un espèce de goudron / une espèce de feuille goudronnée qu'était sec / pour pas qu'la pluie rentre après ... on mettait les lauzes dessus alors i'nous restait déjà des ...

Vincent : et il avait au moins 1 m2 de plus ?

Bruno : 2 m, 50 à peu près

Vincent : et il était de combien de hauteur à peu près ?

Bruno : oh de hauteur / ça à peu près / euh ... 25 cm

Thierry : moi à ta place j'aurais fait haut comme ça / pa'c'que tu sais / le cochon d'Inde...

Bruno : oh mais ça grimpe pas les cochons d'Inde ...

Vincent : ah si ça grimpe à des... mais pas à des choses lisses...

Thierry : oh mais les cochon d'Inde / ça trouve des p'tits trucs pour surélever...

Bruno : oui / mais les cochons d'Inde c'est pas la même chose que les hamsters

Thierry : oui / mais p'tit à p'tit les cochons d'Inde / ça fait des trucs surélevés et puis / zou / ça passe par-dessus ...

T'as vu un Gaspard sauter ?

Vincent : mais Gaspard c'est un hamster

Thierry : les cochons d'Inde c'est presque comme ...

Vincent : c'est beaucoup plus lourd d'abord / un cochon d'Inde / et j'vois qu'c'est moins intelligent qu'un hamster... Un cochon d'Inde, c'est pas très intelligent ...

Bruno : mais z'i'zont d'lair pourquoi veux tu qu'i' s'en aillent ? I'sont contents...

Thierry : Et s'i' sont dans la nature / i' sont encore plus contents.

6'45" Gilles : qui c'est un homme invisible...
 Sylvain : mais comment * puis-je voir puisque je suis aveugle...
 Gilles : oh arrêtez (rires...)
 Sylvain : tu es un homme invisible et tu n'es pas un aveugle...
 7' Gilles : ah ...j'suis un aveugle bon *... j'suis un aveugle /
 alors je te vois...
 Sylvain : tu n'es pas un aveugle...
 7'15" Gilles : ah ...* fallait l'dire plus tôt / fallait l'dire plus
 tôt... je suis un aveugle...
 Sylvain : tic et tic et... ah oh ... (rires des deux)
 7'30" et pourtant je suis * un aveugle... je suis pas
 un aveugle humain... (chante) : je suis pas un
 7'45" aveugle * je suis pas un aveugle... je sais pour-
 tant / je ne suis pas un aveugle...
 8' Gilles : vois *
 Sylvain : et pourtant... je ... suis pas un aveugle... et
 8'15" pourtant je suis en train de ma*...
 Gilles : et ben / machin
 Sylvain : ouais (bruits)
 Gilles : tu t' rappelles Donald au pays des matelots ?
 Gilles ; hein ? ...
 Sylvain : ah non non / c'est pas Donald au pays des mat'lots
 ...
 8'30" Gilles : enfin tu t'rappelles ?*... Donald .. euh ...
 Sylvain : amoureux .. vach'ment bien eh...
 Gilles : avec le euh.. comment ça s'appelle... le prof
 8'45" Donald dingue *
 Sylvain : et la machine-là / la machine...
 Gilles : ah oui
 9' Sylvain, Gilles : ch ch s... * pouah ma gl... eh ou ou ouh...
 9'15" Gilles : et Donald qu'arrive pas à ouvrir la f'nêtre *
 nia nia gr f ch uit ... glein glein ... cr...

1' Olivier : Thierry i' i' s'sentait * pas dans son assiette/
 hein c'matin... c'est-à-dire que Thierry / i'
 fait son malin... i' il est pas du tout moniteur
 de ski / i' sait pas bien faire du ski...

Benoit : il a jamais dit qu'i' savait en faire... hein
 maman ?

1'15" Olivier : ouais i' fait son malin... i' dit * i' faut faire
 comme ça / et comme ça / et comme ça et i' sait
 pas l'faire... ..
 et quand même on a déjà vu des gens faire du ski ...
 des gens qu'... des gens qui savent mieux faire du
 ski qu'Thierry... et pourtant ..*... pas maman quand
 même...

1'30" Benoit : et moi j'en connais un qui sait moins en faire
 que lui / i' s'appelle Olivier Weil...

Olivier : mais moi j'vous dis pas qu'j'en fais mieux qu'lui /
 j'dis * qu'i' sait pas faire du ski..... je
 dis pas qu' je l'bats... hein maman ?

1'45" Mère : hum...

Benoit : mais en vérité Thierry sait faire du ski...
 c'est un bon skieur... .. tu sais pourquoi i'
 nous emmène faire ça ?...

Olivier : hein hein...

2' Benoit : pa' c' qu'i' veut * nous apprendre... (brouillé)

Olivier : mais si / i' savait en faire / ce s'rait p'êtr'
 mieux...

2'15" Mère : mais il sait le faire...

Olivier : non *

5'45" Gilles : mais c'est pas l' jeu des adverbes * ...

Pascal : non mais regarde...

Gilles : faut lui apprendre / hum / i' sait pas ...

Sylvain : ben voilà ... alors tu poses des questions...
 t'as pas l'droit de continuer ... alor' i't'
 répond ce qu'i' veut / et moi aussi ... mais
 alors tu ... tu lui dis pas autre chose * hein /
 tu comprends ? ... i' faut toujours poser des
 questions / et i' faut trouver ce qu'on ... ce

6'

qu'on... comment on fait ça hein... tu comprends ?

6'15" Pascal : non

Gilles : non / même moi j'comprends vraiment plus c'que tu m'avais dit en classe * alors...

6'30" Sylvain : bon alors voilà... y a un gars.. bon alors le jeu / c'est que moi je euh... pour la première question je dis euh... n'importe quoi / hein... Alors lui / i' retient * la réponse euh pas ma réponse / la vraie réponse / i' prend la vraie réponse / de lui ... et tu lui poses une question

6'45" Gilles : et je réponds de sa réponse*...

Sylvain : oui c'est ça la maladie qui

Pascal : ah...

Sylvain : et c'est toi qui dois trouver cette maladie...
... c'est pas mal hein ? ...

Pascal : mais c'est quoi la maladie ?

7' Sylvain : et ben la maladie / c'est que ... c'est que... toi tu réponds à ma question / et moi * je reponds à ta question...

Pascal : j'ai compris / j'ai compris

Gilles : on y va de nouveau ?

Sylvain : allez / alors allons-y ...

Pascal : alors euh ... quand estce que vous vous êtes levés ?

7'15" Sylvain : quand je me suis levé / tu veux dire / parc' que c'est pas à nous deux qu'on pose les questions* / c'est à un seul .. bon moi / je me suis levé ... tiens j'me rappelle plus l'heure... bon ça fait rien... continue avec lui...

F. FRANCOIS - Paris V°

Voilà à peu près dans notre groupe d'étude comment nous nous sommes posés le problème : nous sommes partis de cette remarque bien connue : c'est avant tout au niveau du maniement de la langue que se conditionne l'échec scolaire et en particulier l'échec scolaire des enfants appartenant à l'ensemble de ce que l'on appelle "milieux sociaux non favorisés". Et nous nous sommes posé la question, qui va peut-être vous sembler très prétentieuse et très mal posée : pouvons-nous dire que cet échec est lié à une incapacité réelle à manier la langue dans ses fonctions ou au fait qu'il y avait en quelque sorte une langue des classes favorisées qui était par ailleurs celle à laquelle on demandait de réussir en classe et aux examens. L'hypothèse que nous avons voulu développer c'est qu'il était nécessaire de distinguer entre la fonction "normative" de l'école et sa fonction "supernormative". Qu'est-ce qu'on entend par fonction normative de l'école ? C'est ce qui reflète la nécessité de passer d'un langage en situation à un langage hors situation. Je prends un exemple simple : lorsque je parle à quelqu'un qui est en face de moi ou lorsque je parle à quelqu'un que je connais, avec qui j'ai l'habitude de communiquer au sujet d'un certain nombre d'objets qui nous sont familiers, en règle générale, je n'ai pas à utiliser de moyens linguistiques spécifiques pour "actualiser" ce dont je parle. Tout les mots de la langue sont abstraits en ceci qu'ils peuvent se référer à une infinité d'objets ou d'aspects de la réalité. En situation, nous savons à quoi se réfèrent les expressions le chien et en haut dans la phrase le chien est en haut. C'est pourquoi la "langue" en situation n'est pas organisée comme la "langue hors situation". Par exemple, les phrases à sujet pronominal y sont beaucoup plus fréquentes que les phrases à sujet nominal. De la même façon la continuité de ce que j'exprime est généralement inutile à manifester dans mon discours, parce qu'elle est rendue présente par la situation. La fonction normative, inévitable, de l'Ecole, c'est de permettre à l'enfant de communiquer sur des objets absents et des interlocuteurs variés, aussi bien, quoique avec des formes différentes, à l'oral qu'à l'écrit. Certes l'enfant ne peut commencer à apprendre à parler qu'en fonction de situations relativement stables et récurrentes : où ailleurs que dans cette situation acquerrait-il le sens des mots ? Mais la culture de l'enfant de trois ans c'est en partie d'échapper à ce champ limité de la "communication situationnelle".

On ne peut se contenter de dire que l'enfant apprend à parler en parlant, "que chacun parle comme il en a l'habitude et tout ira bien", ce n'est pas vrai. Il doit nécessairement changer de situation et de modalités de communication.

Mais d'autre part, notre pratique pédagogique actuelle et pour des raisons qui sont liées à la société de classe dans laquelle nous vivons et pour des raisons liées peut-être aussi à une surnorme propre au milieu scolaire, confond cette nécessité avec des modèles culturels particuliers qui ne sont nullement nécessaires à l'exercice même de ce changement de fonction du langage. Je pense par exemple à la nature extrêmement particulière des textes littéraires qui sont proposés aux enfants au niveau de l'enseignement primaire, et je pense également à ce que nous appelons la réussite scolaire au niveau linguistique qui est fondée sur des critères extrêmement bizarres, sur lesquels je pense, la plupart des gens présents ici, sont d'accord. Le problème que nous nous sommes posé, c'était, étant donné que bien entendu, je ne suis pas tantôt en face de la norme, tantôt en face de la surnorme, de savoir si nous pouvions avoir des indices qui nous permettaient de différencier l'une de l'autre.

Nous en sommes actuellement au stade de la pré-enquête. Il faudra peut-être modifier ce que nous demandons aux enfants ou la façon dont nous dépouillons les textes. Actuellement, nous distinguons tout d'abord trois types d'utilisation du langage. C'est-à-dire que nous mettons les enfants en face de trois tâches différentes. Première tâche : la tâche (on a discuté longuement ce matin sur la notion de "texte libre" je n'y reviens pas) la tâche donc de réutilisation personnelle du langage : "raconte-moi une histoire dont tu aimerais qu'elle t'arrive". Deuxième tâche, celle qui consiste à raconter une bande dessinée. Nous supposons qu'elle est celle qui fait le plus appel à la fonction de norme réelle de l'école. En effet la possibilité pour l'enfant de coder une réalité extra-linguistique, est effectivement ce qu'il faut, quelle que soit la structure sociale dans laquelle nous vivons ou nous vivrons, que l'école donne. Enfin la troisième tâche, c'est de répéter dans ses propres mots, un texte littéraire que le maître a dit plusieurs fois, texte qui est volontairement assez chargé en structures lexicales relativement compliquées. La première hypothèse que nous faisons c'est que le "handicap socio-culturel" des enfants des classes défavorisées, ne

se présentera pas de la même façon en fonction de ces trois tâches, et en particulier que les différences entre les enfants seront moindres dans la tâche "récit de bande dessinée", que dans le récit libre, qui est au contraire une des tâches les plus contraignantes qui soit.

Deuxième hypothèse : les différents indices que nous allons pouvoir utiliser ne donneront pas les mêmes résultats. C'est à dire que par exemple, un indice de richesse lexicale, fondé en particulier sur ce qui est le plus variable chez les enfants, par exemple le nombre d'adjectifs, différenciera très fortement les classes sociales, et là je pense que les pré-enquêtes que l'on a pu faire donnent déjà une indication claire. En revanche, si je me demande, face à une bande dessinée, combien de circonstanciés d'action seront rendus par les enfants, il n'est pas du tout certain qu'il en sera de même et il est même probable que les résultats seront différents. La question que nous nous sommes posé ensuite était alors : que mesurent effectivement ces indices eux-mêmes ? Premièrement nous savons tous que lorsque l'on cherche à compter, on est toujours amené à compter ce qui est le plus facile à compter, sans que ce soit forcément le plus intéressant. Je prends un problème très important, celui de la continuité du récit, du lien entre les phrases successives. Or, aussi bien à l'oral qu'à l'écrit, ce qui fait qu'un récit se continue, est quelque chose qui n'est lié, ni à l'existence de signes spécifiques, de signes du type conjonction de subordination ou autres, ni non plus à la reprise du même mot. Nous comprenons tous par exemple, la continuité des phrases : est-ce que tu sors ? non il pleut, et il est bien entendu que je n'ai pas à répéter non je ne sors pas, il pleut. Autrement dit c'est au contraire le récit dans lequel on reprendrait tous les mots de la phrase précédente qui ne serait pas continu du tout. Mais s'il n'y a pas nécessairement d'outils spécifiques pour marquer la continuité des énoncés quel indice utiliser pour mesurer cette continuité ? C'est évidemment plus facile de compter les fautes d'orthographe... D'autre part nous voudrions écarter les valorisations implicites dans le choix des indices, valorisations qui nous semblent sous-jacentes dans la plupart des enquêtes socio-linguistiques et en particulier dans celles de Bernstein. Prenons trois exemples sur lesquels je pense on peut se mettre d'accord, Premièrement le problème de la richesse lexicale. Il est bien entendu, que lorsque je mesure les progrès d'un enfant de 1 an 1/2 à trois ans, il est

souhaitable qu'il acquière un certain lexique. Il est mieux de savoir deux mots que d'en savoir un seul. Bon, mais l'ennui, comme l'a dit Anne-Marie Houdebine est de savoir : est-il mieux et absolument mieux d'en savoir 10 000 que d'en savoir 5 000 ? Le problème ne se pose évidemment plus du tout de la même façon. Deuxièmement, le problème de la complexité syntaxique. Au début de notre étude, j'avais pris en quelque sorte l'attitude extrêmement iconoclaste, qui consiste à dire : puisque tout peut être dit par des énoncés simples, la complexité syntaxique relève uniquement du domaine de la surnorme, de la volonté des pédagogues qui tiennent absolument à habituer les enfants à ne pas dire : il faisait beau, nous sommes sortis, mais : puisqu'il faisait beau nous sommes sortis. ce qui ne change pas grand chose à la nature de ce qui est communiqué. En fait les choses sont plus compliquées que cela. En effet on peut dire que lorsque je divise mon information en deux phrases successives, ce qui est toujours possible, j'ai naturellement à ajouter un anaphorique dans la seconde phrase, pour indiquer par exemple, que les participants de la seconde phrase sont les mêmes que ceux de la première.

C'est pourquoi on rencontre dans une analyse du discours comme celle de Harris, qui veut réduire les phrases réelles en énoncés plus simples, de grandes complications pour indiquer que le participant de la seconde phrase dont on dit qu'il est gentil est le même que celui dont on a dit qu'il arrivait dans la phrase unique : le gentil garçon arrive. Autrement dit il faut bien voir que dans un certain nombre de cas, la phrase complexe, c'est-à-dire la phrase comportant les éléments subordonnés, est plus "simple" que la juxtaposition de deux phrases qui doit, en tout cas si l'interlocuteur est de mauvaise volonté, ou ce qui revient au même, si j'ai à recréer pour lui la situation, introduire un élément explicite de continuité.

Troisième problème de choix d'indice : le problème de l'indice de la longueur du texte. Il serait bien entendu caricatural de dire plus c'est long, plus c'est beau. Mais est-ce que nous n'avons pas en particulier dans les textes libres tendance à dire que le sujet qui parle le plus longuement, investit davantage le langage, fait quelque chose de mieux que celui qui parle peu. Il ne suffit donc pas de mesurer richesse lexicale ou complexité syntaxique. Le sens de ces termes ne pas de soi. Considérons tout d'abord le lexique. Il faut rappeler ici un certain nombre de points qui conditionnent la problématique. Premièrement, comme vous le savez, il y a une structure hiérarchique du lexique en ce qui concerne les fréquences, puisque les 100 premiers mots les plus fréquents d'un texte constituent en moyenne, 50 à 60 % de ce texte autrement dit, il y a des mots extrêmement fréquents qui ./. - 72 -

reviendront dans n'importe quel texte. Deuxième point, qui me semble extrêmement important, vous savez que lorsqu'il s'agit des mots les plus fréquents, établir un vocabulaire fondamental a un sens précis, mais que lorsque ensuite on passe aux mots les moins fréquents, leur présence ou leur absence sont complètement conditionnées par la nature de ce dont je parle, et par la nature des habitudes de communication de tel ou tel sous-groupe. Il n'est donc pas possible de caractériser absolument des mots en disant ils ont tel ou tel indice, de fréquence : ça n'a plus aucun sens de comparer la relative rareté du 3 000 ème ou du 10 000 ème mot d'un lexique, le 10 000 ème mot d'un lexique peut être un mot tout à fait en fait fréquent dans par exemple telle ou telle profession, telle ou telle activité, caractéristique de tel âge et non pas de tel autre, etc. On nous a parlé ce matin de ce qui se passait à Limoges et de la richesse du lexique servant à désigner les types de repas, Il est bien entendu que là nous avons une diversité nécessaire et qu'une des caractéristiques les plus évidemment "sumnormatives" est de faire comme si on savait avec certitude quel vocabulaire particulier il faut absolument acquérir. Enfin, en dehors du problème des mots les plus fréquents et ensuite d'une dispersion de termes dont on ne peut absolument plus évaluer la fréquence comme on le fait d'habitude, le problème qui se pose est de savoir quel est le lien entre justement la structure du lexique et le passage au langage hors situation, c'est-à-dire à la nécessité de coder telle ou telle chose. Et là nous nous trouvons devant une situation extrêmement difficile. Premièrement ce qui caractérise le code linguistique par rapport à tout autre code, c'est que tout peut être paraphrasé, c'est-à-dire que après tout il n'y a jamais quelque chose qui doit absolument être dénommé d'un mot et non pas par un syntagme et non pas en particulier par la combinaison d'unités de fréquence plus élevée. Autrement dit nous ne pouvons pas dire absolument que pour parler de telle ou telle chose, j'ai besoin de tel ou tel mot. Après tout on pourrait même soutenir que la richesse lexicale est indépendante de la capacité à cerner telle réalité, par exemple tel événement qui se passe sur une bande dessinée. La thèse "antinormativiste" absolue ce serait alors de dire que ce que nous avons à mesurer c'est la capacité à se servir de quelques centaines d'unités extrêmement fréquentes, non pas la présence ou l'absence d'un lexique riche qui serait uniquement renvoyé du côté de la surnorme et du côté des mécanismes d'identification à la classe dirigeante ou aux groupes favorisés. ./.

Le problème est en fait plus compliqué que cela. Il y a évidemment lexique facultatif qui peut être utilisé dans une situation mais n'est cependant pas nécessaire, mais il y a d'autre part un certain nombre de "pré-codages" dont nous pouvons extrêmement difficilement nous passer, l'ensemble des systèmes d'oppositions par lesquels il faut que nous passions dans une langue donnée pour organiser tel type d'expérience. Par exemple il y a des dichotomies extrêmement fréquentes, en particulier des circonstants, des prépositions ou adverbess de lieu sur, sous ou dessus, dessous ou à gauche, à droite, dont je vois très mal comment on pourrait les paraphraser. Prenez d'autre part le problème du lexique technique, par exemple les noms de nombre, on voit très mal un enfant réinventant la numération, indépendamment de l'acquisition des noms de nombres, autrement dit-il est bien vrai qu'à la limite je peux tout paraphraser, je peux toujours remplacer un mot par un autre, il y a en quelque sorte un certain nombre de préorganisations lexicales dont il est extrêmement coûteux de chercher à faire l'économie. Autrement dit le problème de l'analyse lexicale est plus difficile que ce que l'on aurait pu penser, si on voulait simplement renvoyer le lexique peu fréquent du côté des codages relativement inutiles.

Prenons un second problème, celui de la complexité syntaxique. En fait quand on construit des indices de complexité syntaxique, on ne mesure pas une complexité réelle pour le sujet, mais la complexité en fonction de l'exposé dans une grammaire donnée. Je prends un exemple très gros : on peut vouloir présenter une grammaire générative qui par exemple dérivera les fonctions secondaires des fonctions primaires. Ça ne veut évidemment pas dire que l'adjectif épithète est une structure plus compliquée, en quelque sens que ce soit du mot compliqué, du point de vue psychologique que l'adjectif attribut. A la limite on aboutit à des caricatures qu'on a trouvés dans certains ouvrages, dans lesquels par exemple on engendre par suppression du relatif et du verbe l'adjectif épithète à partir d'une relative, elle-même engendrée à partir d'une indépendante, et on arrive à dire que l'adjectif épithète est une des structures les plus compliquées qui existe en français. C'est absolument de l'idéalisme, c'est-à-dire une façon de plaquer un cadre d'exposition sur la réalité. Bon, autre exemple tout aussi comique, est-il plus facile, plus difficile ou d'égale difficulté de dire : je veux aller, ou je veux que tu ailles ? Je dois dire pour montrer que ces faux problèmes nous menacent tous que quand je me suis aperçu que des enfants extrêmement jeunes utilisaient des formes du subjonctif, cela n'a étonné. Ça montre simplement que c'est moi, qui plaquait ma grammaire déductive, qui part des formes de l'indicatif pour aller aux formes du subjonctif sur les conditions de langage de l'enfant. Il est bien entendu

que l'enfant entend extrêmement souvent des formes du type : il faut que t'ailles, je veux que t'ailles, etc... et que pour lui la forme "ailles" n'est pas en elle-même une forme plus compliquée que la forme "irait" ou la forme "vais". Autrement dit il est bien entendu qu'il y a effectivement des complexités syntaxiques, mais que ce ne sont nullement nécessairement les mêmes que celles qui découlent de l'ordre d'exposition dans telle ou telle grammaire générative ou autre.

Quelles sont donc à partir de ces remarques négatives les hypothèses que l'on peut proposer. La première hypothèse que je voudrais proposer, c'est l'hypothèse fonctionnelle, hypothèse fonctionnelle que l'on pourrait opposer à deux autres types d'hypothèses, l'hypothèse que je pourrais appeler normativiste, et l'hypothèse qui chercherait au contraire à trouver indépendamment de la forme des discours une analyse purement fondée sur le contenu. Problème de l'hypothèse normativiste : encore maintenant en milieu pédagogique on utilise extrêmement souvent la notion de faute. Or cette notion de faute concerne environ 1 % des formes émises par les enfants. On focalise son attention sur tel ou tel point et on oublie tout le reste. La notion de faute, de ce point de vue relève uniquement de "l'hyper-norme". J'ai été élevé en tant qu'appartenant aux classes linguistiquement les plus favorisées, dans la connaissance de quelques secrets, du type il ne faut pas dire par contre, il faut dire en revanche, il ne faut pas dire pallier à, etc... Ceci est quelque chose qui est le bien commun d'un certain nombre d'universitaires. Dans ces exemples nous avons l'aspect le plus ridicule de la notion de faute qui relève alors du code des bienséances et non de la langue. Le même aspect

Deuxième aspect tout à fait différent de la notion de faute : la notion de faute comme liée à la probabilité d'une mise en relation d'un plus grand nombre d'unités. Autrement dit, plus nous avons à mettre en relation explicitement dans la phrase des unités distantes, ou des unités nombreuses, plus les mécanismes qui manifestent cette mise en relation, sont des mécanismes difficiles et sur lesquels nous risquons de faire des fautes. Je dis que là, de ce point de vue, nous avons un sens objectif de la notion de faute, comme déterminée par la nécessité de coder ce qui dans la communication quotidienne passe par la situation ou reste implicite.

./.

Cette complexité syntaxique réelle est complètement différente de celle qu'on pourrait déduire d'un ~~doit d'at~~ ~~analyse~~ ~~de type~~ ~~transformationnel~~, soit de certaines analyses de type fonctionnel. Je prends un exemple qui n'a semblé extrêmement intéressant ; qui est l'analyse récente que Tissot ~~Moulin~~, et Lhermitte, ont fait du phénomène de l'agrammatisme. Il est bien entendu qu'analyser la langue de l'enfant ce n'est pas du tout analyser la pathologie du langage. Mais ce dont Tissot, ~~Moulin~~ et Lhermitte se sont aperçus, c'est que contrairement à ce que l'on croyait, l'adulte, agrammatique n'était nullement quelqu'un qui n'avait plus de syntaxe ou qui n'avait plus de petits mots. Ils se sont aperçu également qu'on ne pouvait pas distinguer la complexité des modalités et des fonctionnels. Autrement dit, on ne peut pour simplement : modalités centripètes intégrées au syntagme : faciles, fonctionnels centrifuges sortant du syntagme : difficiles. Pour les malades en questions, des, parce que, des si des alors, ou des conjonctions de subordination c'était très facile, des articles ou des prépositions aussi mais il y avait deux types d'unités qui étaient objectivement difficiles, d'une part les relatifs d'autre part les adjectifs possessifs, qui sont les deux classes syntaxiques, l'une constituée de fonctionnels, l'autre de modalités, qui ont effectivement une double référence, le relatif dans deux propositions et également l'adjectif possessif qui renvoie au "possesseur" et à la "chose possédée". Je ne dis pas que là nous avons une vraie mesure de la complexité syntaxique mais nous avons malgré tout quelque chose d'extrêmement intéressant d'autant plus que nous savons très bien que c'est sans doute un universel linguistique, qu'aucun enfant ne commence à utiliser correctement les adjectifs possessifs, que tout enfant commence à dire, c'est à bébé, ensuite c'est à moi, ensuite c'est non, et de même que tout enfant commence par 'utiliser les "faux" relatifs c'est à dire les relatifs qui sont en fait des expansions d'un prédicat nominal : c'est ou y a un garçon que tu connais. avant de faire des phrases du type, c'est le garçon que j'ai rencontré hier qui a fait ceci ou cela. Autrement dit, je ne suis pas du tout hostile à une grammaire qui chercherait à trouver des correspondances entre complexité dans la théorie et complexité dans le comportement. Je dis que cela vaut dans les cas de genèse réelle, et non pas dans les cas de genèse idéal comme on le fait dans les grammaires ~~transformationnelles~~, qui on ne sait trop pourquoi, ne veulent pas par exemple considérer les phrases à prédicat nominal comme une donnée de base.

Autrement dit, au niveau d'une hypothèse que j'appellerai fonctionnaliste, je voudrais opposer à la notion de correction, le problème d'une fonction réelle de la syntaxe c'est-à-dire d'une fonction de mise en relation d'un certain nombre d'éléments. Je voudrais opposer cette notion de fonction syntaxique à la notion de norme. Je voudrais aussi l'opposer aux tentatives qui ont été faites, qui ne semblent actuellement prématurées, et peut-être contraires à la façon même dont nous apprenons à parler pour faire une analyse du contenu, c'est-à-dire pour retraduire les textes des enfants dans une espèce de langue neutre, de langue artificielle, comportant uniquement par exemple des termes et des relations et pour ensuite essayer de voir s'il y a une distance entre les structures syntaxiques de l'enfant et la quantité d'informations qu'ils apportent. Or en fait, là je crois que cette tentative est extrêmement intéressante, mais que nous n'avons pas les moyens de les mener à bien, car en fait lorsque l'enfant acquiert des structures syntaxiques, il acquiert non seulement là des formes correctes, il acquiert aussi des moyens d'organiser son expérience, et c'est une chose légitime de se demander, est-ce que l'enfant est capable de passer d'une organisation de cette expérience à l'autre, par exemple est-ce que l'enfant est capable de faire des paraphrases. Mais essayer d'évaluer, indépendamment de toute forme linguistique, quelle est la quantité d'informations que l'enfant apporte dans son discours, cela présente tout d'abord une difficulté technique. Certes, on peut isoler des termes, des procès, des relations, des circonstants, traduire ainsi une partie notable de l'information. Toute l'information c'est une autre question. D'autre part, encore une fois certains contenus sont très liés aux formes particulières de codage par lesquels on les communique. D'autres sont plus indépendants. Là encore l'étude linguistique doit précéder la fabrication des tests.

La seconde idée que je voudrais développer, c'est qu'il n'y a pas d'un côté une compétence linguistique, de l'autre une utilisation de cette compétence. C'est une idée qui est ressortie ici un certain nombre de fois, il n'y a pas la compétence linguistique et les utilisations du langage, il y a des structures linguistiques qui ne sont actualisées qu'en fonction de telle ou telle situation de communication. Ce pourquoi la notion de compétence linguistique coupée des conditions de la communication est largement un mythe. Je pense que c'est un des points sur lesquels les successeurs de Bernstein ont fait avancer l'analyse. ./.

En particulier des auteurs comme Lawton ou Robinson ont très bien montré qu'il n'y avait pas une fois pour toutes un code restreint, et un code élaboré, mais que par exemple lorsqu'on demandait à des enfants des milieux sociaux dits défavorisés, de faire un discours abstrait, par exemple un discours sur le suicide à un interlocuteur, supposé être un interlocuteur savant, avec qui il fallait faire un discours savant, il y avait très peu d'écart entre les enfants des classes favorisées et les enfants des classes défavorisées, et que en revanche, c'est lorsqu'on demandait aux enfants de faire un discours en quelque sorte moins contraint à leurs camarades, que l'écart entre richesse, longueur, complexité quelle que soit la valeur de ces indices etc. apparaissait extrêmement grand. Autrement dit nous n'avons aucun moyen de tester la compétence linguistique comme telle, ce qui fait que pour ma part, du point de vue des expériences à mener, il me semble qu'il faut nous tenir à distance de deux extrêmes. Le premier extrême c'est celui dans lequel on se placerait si on se tenait dans une perspective purement expérimentale, c'est à dire si par exemple on proposait une tâche de complétement de phrases où en fait on introduit un très grand nombre de biais : la demande peut être plus ou moins bien reçue par les enfants. Elle n'a pas forcément pour l'enfant la même signification que pour l'adulte. Elle est peut être pure, mais loin de toute communication. L'autre extrême conduit aux difficultés auxquelles s'est heurté notre collègue ce matin : on a des textes dits libres, et que ces textes sont tellement différents dans leur contenu, évidemment comparer le nombre de pronoms personnels ou qui que ce soit d'autre dans des textes de contenu entièrement différent n'a plus aucune signification. C'est pour ça qu'il me semble que si nous voulons avancer nous avons intérêt à comparer les productions des enfants dans des tâches relativement précises, mais qui néanmoins n'entraînent automatiquement un certain type de réponse. Je peux dire que par exemple, pour ma part, d'après tous les essais qu'on a fait ici ou là, la tâche qui consiste à décrire une bande dessinée est une bonne tâche, c'est-à-dire c'est une tâche dans laquelle les enfants peuvent effectivement avoir des comportements linguistiques qualitativement différents mais comparables. Voilà où nous en sommes actuellement. Je voudrais vous dire les points sur lesquels nous sommes à peu près sûrs de ce que nous faisons, et les points sur lesquels actuellement nous nous interrogeons et nous voudrions essayer que vous nous aidiez à trouver des solutions.

Premièrement le problème de la faute. Faut-il ou ne faut-il pas compter des fautes. Eh bien dans la mesure premièrement où la notion de faute actuellement conditionne l'échec ou le succès scolaire il serait aberrant de ne pas en tenir compte. Notre idée c'est que naturellement il ne faut pas en soi compter les fautes selon les milieux sociaux mais nous poser la question qui ne semble extrêmement importante : est-ce que la faute varie corrélativement aux autres indices ou ne varie pas corrélativement avec les autres indices. Alors la difficulté à laquelle nous nous sommes heurtés est double : premièrement il est bien entendu que plus un sujet utilisera des structures relativement compliquées plus il aura de probabilités de faire de fautes. Je prends un indice tout simple : rien n'empêche jamais de faire un récit uniquement au présent, il est bien entendu que si je fais un récit en utilisant des formes d'imparfait, de plus-que-parfait ou de futur, la probabilité que je fasse un plus grand nombre de fautes est plus grande. Autrement dit, une des difficultés mathématiques devant laquelle nous nous trouvons c'est que je ne peux comparer le nombre de fautes par texte que à un degré de complexité syntaxique ou lexical relativement égal. Vous voyez la grosse difficulté que ça présente et il est bien entendu qu'un compte absolu des fautes n'a aucune signification.

Deuxième point : doit-on compter toutes les "fautes" ou peut-on constituer un indice représentatif ? Le premier indice que nous avons essayé d'utiliser c'était au niveau de la langue écrite, les fautes portant sur la forme des modalités verbales. Dans la mesure où ce sont des données qui ne peuvent être acquises que coup par coup et non pas activement par le sujet on pouvait penser qu'elles reflétaient le pôle passif de l'acquisition du langage. L'ennui c'est que sur ce point là actuellement nous n'avons pas tout à fait assez de fautes et que donc nos indices ne sont pas suffisamment différenciateurs.

Dans un autre relevé sur 40 élèves de CM1, on a pu noter :

- a) il n'y a pratiquement pas de fautes à l'oral
- b) à l'écrit la quasi-totalité des fautes prennent la forme de fautes d'orthographe
- c) il n'y a pas de corrélation entre fautes d'orthographe d'usage et grammaticales.
- d) il n'y a pas non plus de corrélation entre le nombre de ces fautes et les autres indices linguistiques.

En sommes surtout si on se réfère à des corpus réels d'adultes et non à ce qu'ils croient dire la notion de faute n'a peut-être pas grand sens à l'oral chez des enfants de plus de 6 ans. Elle en a un à l'écrit, celui de fautes d'orthographe mais à peu près complètement dissocié du reste du comportement linguistique. Au moins comme résultat négatif, c'est là quelque chose à tester sur une population plus étendue. Considérons un second indice : la longueur

Là je pense, qu'on le veuille ou non, qu'il est absolument nécessaire de calculer un indice de longueur de l'énoncé, non pas encore une fois comme un peu naïvement nous avons commencé à faire au début en disant : longueur = mesure de l'incitation à communiquer, ce qui était faire une hypothèse psychologique d'un coût absolument prohibitif. Il est bien entendu qu'il y a des modèles sociaux différents et une fois que j'ai mesuré la longueur, je n'ai rien mesuré d'autre que la longueur. En revanche il faut que je mesure cette longueur puisque par exemple pour ensuite évaluer correctement la richesse lexicale, je dois forcément comparer la richesse lexicale au nombre de mots utilisés. Bien sûr, isolée cette mesure de la longueur n'a aucune signification. Elle n'a de valeur que par comparaison avec d'autres indices ou en fonction de sa variation dans des conditions différentes. Ainsi dans les trois tâches différentes, tâche de description d'après image, tâche de description avec tous les guillemets "libre", et tâche de répétition ou de paraphrase d'un texte complexe donné oralement : il peut être très intéressant d'évaluer la longueur relative dans les trois tâches en question, l'hypothèse qui semble se confirmer étant que la longueur relative ne varie pas de la même façon chez tous les enfants dans les trois tâches. Alors à ce moment là, la mesure de longueur sert à quelque chose : au moins à indiquer un problème.

Troisième et quatrième mesures possibles : mesures contrastées, alors la l'hypothèse est très grosse et elle peut être discutée, de ce que je vous appellerai pré-codage linguistique et néo-codage linguistique. C'est du jargon pour dire simplement que d'un côté des enfants apprennent des mots et les structures et que par ailleurs les enfants apprennent à appliquer ces structures lexicales et syntaxiques dans des situations nouvelles. L'hypothèse que nous faisons c'est que nous n'aurons pas une variation concomitante si nous comptons le nombre de mots différents, la complexité et les structures syntaxique évaluées tout simplement à partir du nombre de syntagmes nominaux non

nécessaires, et de propositions subordonnées, et enfin troisième indice que je vais appeler nombre de circonstants, c'est-à-dire capacité à verbaliser des données qui restent ordinairement implicites en situation. Je veux dire par exemple à indiquer devant ceci il y a cela, derrière ceci il y a cela et ainsi de suite. Or, l'hypothèse qui semble se confirmer c'est que si on fait une comparaison entre enfants de milieux sociaux favorisés et défavorisés, la complexité syntaxique, c'est-à-dire l'habitude de faire des phrases compliquées sera effectivement plus grande chez les enfants des classes favorisées, la richesse lexicale sera également plus grande, mais que en revanche dans la tâche d'arriver à donner un certain nombre de circonstances donc à coder une donnée expérimentale visuelle, les deux groupes sociaux seront à égalité. J'ajoute qu'en relation avec l'analyse des correspondances qui a été faite ce matin, qu'on avait fait une pré-expérience qui consistait à comparer dans deux grands types de complexité syntaxique, à savoir la complexité liée à des syntagmes nominaux prépositionnels facultatifs et d'autre part à des propositions subordonnées. On avait trouvé quelque chose qui était assez intéressant, c'est-à-dire que la complexité syntaxique était relativement homogène chez les enfants des classes favorisées, c'est-à-dire que les deux types d'introducteurs de complexité étaient relativement également répartis chez un même sujet et que chez les enfants des classes défavorisées, en tout cas ceux que l'on avait testé, il y avait dissociation complète des deux types de complexité syntaxique c'est-à-dire que l'on avait l'une ou l'autre.

Autrement dit, j'y reviendrai ensuite, notre but ici ne va pas être uniquement de trouver des moyennes et des écarts-types. Notre but aussi, et c'est en cela que des problèmes mathématiques vont se reposer, va être de voir s'il y a des typologies caractéristiques de tel ou tel groupe de sujets. Donc on cherchera dans trois tâches différentes, à comparer les types de fautes, la longueur, la complexité syntaxique conçue comme phénomène de codage facultatif, c'est-à-dire conçue comme phénomène non nécessaire, à l'expression ~~à un certain contenu~~, et enfin la complexité syntaxique conçue comme phénomène de codage nécessaire. Sur ces deux derniers points je prends un exemple qui soit, j'espère, clair : prenez les syntagmes, de type syntagme pré-positionnel sur la table, sous la table, on peut dire que ces syntagmes prépositionnels ou les autonomes correspondants : la table, il est dessous, mais on retrouve ici la sur-norme !! sont le moyen préférentiel d'exprimer des relations spatiales. En revanche prenez l'existence de propositions subordonnées du type : pendant qu'on se

L'absence de propositions en ordre des du genre : parce qu'on se
promenait, il est tombé de l'eau, à la place de : on se promenait,
il est tombé de l'eau, je dis que c'est un codage syntaxique facul-
 tatif. Et notre hypothèse c'est que des deux indices, le second sera
 plus sensible à la variation des classes sociales que le premier.
 Dernier problème, le problème des parties du discours auquel on a fait
 des allusions précédemment. Alors il ne semble que, mesurer des par-
 ties du discours en soi, naturellement est une absurdité. Il y a eu
 des énormités écrites sur le compte relatif des noms par rapport au
 verbe. Il est bien entendu qu'on peut dire alternativement qu'un
 langage qui comporte plus de verbes est un langage plus concret ou
 plus abstrait qu'un langage qui comporte plus de noms, ça n'a stric-
 tement aucune signification. On peut avoir des énoncés nominaux à
 actualisateurs : C'est Pierre, voilà Pierre, il y a un chien, dans
 lesquels il n'y a pas de verbe du tout.

Autrement dit un nombre élevé de noms par rapport aux ver-
 bes peut manifester un comportement essentiellement désignatif. Mais
 les prédicats nominaux peuvent s'accompagner de verbes sans sujet
 nominal. C'est Pierre il ... ou qui. D'autre part en français, le nom
 est le polyfonctionnel par excellence et la croissance du nombre des
 noms pourra (mais non nécessairement) signifier augmentation du nombre
 des circonstants, des relations etc... De plus il est possible de
 remplacer un énoncé à sujet nominal par un énoncé à sujet pronominal.
 Autrement dit, l'indice nom/verbe, devrait tenir compte de l'indice
 nom/pronon. Mais ce dernier indice est à son tour très difficile à
 bien mettre au point, puisque premièrement il existe des pronoms
 "personnels" dont chacun sait qu'ils sont plus naturels à utiliser
 que les autres désignations, et qu'il n'y a que César qui s'appelle César,
 quand il parle de lui même, mais que nominativement nous disons je ou nous.
 Deuxièmement nous savons aussi que si on considère des pronoms de
 type il, ils ont deux utilisations entièrement différentes, l'utili-
 sation deictique ou l'utilisation automatique, (puisque ailleurs qu'à
 l'impératif on ne peut utiliser le verbe tout seul), il fait ça, il
fait ça, et au contraire l'utilisation anaphorique, du type : le garçon
est venu, il était content et non pas le garçon est venu, le garçon
était content. Autrement dit s'il est bien vrai que l'indice nom/verbe
 il ne faut pas oublier qu'il n'est pas, lui non plus, univoque. L'en-
 fant a tendance à utiliser d'abord soit des phrases à prédicat nominal
 soit des phrases à pronon, + verbes ensuite des phrases à sujet nomi-
 nal, et qu'il fait partie de la norme scolaire, (je vous signale ./.)

d'ailleurs que c'est un usage français qui ne se retrouvera absolument pas en Allemagne ou beaucoup moins en Allemagne ou en Angleterre) de dire que ce n'est pas beau de répéter les mêmes mots et qu'il est extrêmement nécessaire d'utiliser des pronoms anaphoriques. Sans doute, avant d'utiliser des indices chiffrés est-il nécessaire de se demander comment effectivement peut se faire dans un texte le rappel d'un participant déjà nommé. Mais ceci nous conduit à des questions encore plus générales et sur lesquelles nous ne sommes pas du tout certains d'arriver à trouver des indices chiffrés corrects. Premier problème extrêmement difficile, il est bien entendu qu'intuitivement, même si nous ne pensons pas qu'un devoir a pour but de respecter les trois unités, d'espace, de temps et d'action, nous évaluons la continuité dans le récit, continuité que non seulement nous valorisons parce que nous respectons une certaine norme scolaire, mais qui effectivement est une nécessité de la communication. Il fait partie de la norme et non pas de la surnorme d'être capable d'intégrer en soi un certain nombre d'exigences de l'interlocuteur, et donc de lui rendre présent ce qui est évident pour moi, mais qui ne l'est pas pour lui. Le problème c'est que toutes les expériences que l'on a pu faire, nous ont montré qu'il n'y avait aucun parallélisme entre les moyens lexicaux et syntaxiques et la réelle continuité de ce discours. Autrement dit, que la continuité en question peut être perçue d'une façon intuitive mais qu'on ne voit absolument pas une façon objective de la mesurer. Il y a certes des éléments de continuité ou de discontinuité que je peux mesurer dans un texte, par exemple je peux mesurer le passage d'un type de phrase à un autre type de phrase. Je peux mesurer par exemple le passage de phrases de type sujet/verbe à des phrases à questions ou à des phrases à expansions adjectivales. Mais de telles interruptions, de même que les brusques apparitions du je, ou les brusques apparitions d'un autre thème, ne sont absolument pas des faits qui signifieraient pour autant qu'il y a une rupture de la continuité signifiée. Autrement dit nous sommes obligés de constater ceci : il fait partie du moyen de communication linguistique qu'il y ait une continuité thématique qui se manifeste, quels que soient les moyens formels utilisés ou leur absence. C'est pourquoi on ne peut pas trouver de bon indice formel de la continuité du texte

Deuxième problème, celui de l'originalité. Certes c'est commode, on peut toujours commencer par critiquer cette notion. A ce sujet chacun d'entre nous a une attitude extrêmement ambivalente qui consiste au fond à admettre chez les enfants un certain degré d'originalité mais pas trop. C'est-à-dire que nous souhaitons qu'ils ne répètent pas les stéréotypes mais nous n'admettons pas néanmoins, en règle générale, qu'ils se mettent face à la question "qu'est-ce que tu as fait en famille dimanche dernier" ? à parler par exemple de la vie sexuelle de leurs parents. Autrement dit la notion d'originalité n'est sans doute pas un concept opératoire dans la mesure où nous ne pouvons pas trouver de moyens de la mesurer à proprement parler. Néanmoins, (mais les psychologues sont ici nombreux et ils vont peut-être pousser des cris) il y a quelque chose qui n'est pas une mesure de l'originalité, mais qui est un indice dont j'aimerais voir si il varie en fonction des autres ou pas. C'est l'indice dont j'ai parlé précédemment, à savoir l'existence de modèles de phrases différents. On peut malgré tout faire l'hypothèse qu'un sujet qui a en quelque sorte intégré la faculté de coder, est celui qui est capable de passer d'un modèle de phrase à un autre. Et que un sujet qui en quelque sorte n'a pas intégré le modèle linguistique, c'est celui qui tend lorsqu'un premier modèle linguistique, c'est celui qui tend lorsqu'un premier modèle de phrase a été utilisé à réutiliser le même modèle de phrase. Autrement dit je pense qu'il serait absurde d'appeler la variété des modèles de phrases, mesure de l'originalité, de même qu'il serait absurde de considérer que le processus de subjectivité dans le langage se mesure au nombre de fois où on dit je, ce qui est à proprement parler démentiel. Il reste en revanche qu'il y a un moment où le sujet est prisonnier du pré-codage, c'est-à-dire qu'un certain type de phrase s'impose à lui de toute façon et un moment où le sujet est relativement autonome par rapport au pré-codage, c'est-à-dire le moment où le poids de ce qu'il a à dire l'arène ou lui permet, ou lui laisse la possibilité de changer de modèle de phrase. Autrement dit, l'indice, non pas d'originalité, mais de maîtrise de la langue que je vous propose serait l'indice de possibilité de changer de modèle de phrase. Il est bien entendu, chacun d'entre nous a lu Prévert, que la répétition des mêmes modèles de phrases peut constituer un effet d'originalité absolument remarquable et, peut être le comble de la maîtrise du langage. Reste qu'il est sans doute plus intéressant de mesurer la diversité des structures utilisées que la complexité comme telle.

Pour revenir sur les facteurs précédemment isolés, je voudrais signaler que dans une préenquête, on a pris d'une façon un peu agressive au niveau de la classe de sixième des enfants de **classés** sociales favorisées et défavorisées, **classés** mauvais élèves par les enseignants et **afin de voir si les différents indices ainsi établis** variaient ou non de la même façon. Et sur des effectifs relativement très faibles, c'est-à-dire 30 enfants de chaque côté, j'ai eu la satisfaction ou l'inquiétude, comme vous voudrez, de voir que les hypothèses que je faisais étaient vérifiées, mais ça m'inquiète un peu parce que je trouve que c'est un peu trop beau, à savoir qu'effectivement les enfants des classes sociales favorisées avaient des énoncés plus compliqués, avaient des messages plus longs, avaient une richesse lexicale plus grande et faisaient moins de "fautes" mais que les enfants des classes sociales défavorisées, en revanche, codaient mieux l'expérience qu'on leur donnait, c'est-à-dire étaient mieux capables de transmettre, sous forme en particulier de récits et de circonstants la bande dessinée. Je ne pense pas du tout que cette toute petite expérience soit une vérification de ce que j'ai dit, un des points dont j'aimerais que nous discutions, c'est celui-ci : je pense que des études très extensives comme les deux qui ont été présentées ce matin, présentent beaucoup d'intérêt pour centrer un certain nombre de points particuliers. Je veux dire par exemple, pour analyser les formes générales du récit ou pour résoudre le problème très intéressant y-a-t-il un dénominateur commun au code lexical dans les différentes écoles. En revanche il ne semble que si l'on veut essayer de faire des études qui viseraient à étudier ce problème des relations entre classe sociale et langage, on a intérêt à faire actuellement des études de type semi-clinique, comme celle que je propose, c'est-à-dire d'étudier mettons, 20 enfants d'un groupe, 20 enfants d'un autre et de voir si les différents facteurs en question varient corrélativement ou non, car actuellement, si nous faisons des études trop extensives nous risquons de ne pas savoir extrêmement bien ce que nous allons mesurer.

C'est une chose de mesurer la longueur de 3 000 devoirs, une autre de déterminer ce que signifie cet indice, ce qu'on ne peut déterminer que par son lien, variable, avec les autres.

Il est bien entendu que aucune étude ne saurait dans l'état actuel des choses tout mesurer à la fois et en ce qui concerne la structure linguistique et encore plus en ce qui concerne les facteurs qui conditionnent l'attitude des sujets à l'égard du langage. Ainsi, nous n'allons pas pouvoir balayer d'un trait de plume le problème de la conscience normative, c'est-à-dire le problème de la conscience qu'ont l'ensemble des gens qui appartiennent en particulier à la classe ouvrière, que les détenteurs de la belle langue, ce sont les autres. Nous avons fait une expérience à ce sujet, en essayant de nous adresser à des militants syndicaux qui cherchaient justement à améliorer leur langue orale et écrite et en leur disant "la belle langue n'est pas la propriété privée de M. Giscard d'Estaing, mais à partir du moment où l'ensemble de vos syntagmes s'enchaînent les uns aux autres et où vos interlocuteurs savent ce que vous voulez dire, vous parlez tout aussi bien ou mieux que M. Giscard d'Estaing". Et bien ce qui est absolument certain c'est que notre discours a été parfaitement mal reçu, c'est-à-dire que personne n'y a cru. C'est-à-dire que l'idée selon laquelle il y a une belle langue qui est la propriété privée de la classe dominante, est une idée profondément introduite dans la classe ouvrière dans l'état actuel des choses. Et je pense que pour une part l'attitude des enseignants y est pour quelque chose, en effet, je pense que dans l'état actuel des choses, lorsque nous montrons des modèles littéraires aux enfants, nous ne distinguons pas assez entre la compétence qu'ils doivent acquérir en tant que récepteurs et la compétence qu'ils doivent acquérir en tant qu'émetteurs. Après tout jamais nous ne disons clairement aux enfants, le but n'est pas que chacun d'entre nous s'exprime comme Chateaubriand. Or tout se passe comme si, ce but était naturellement impossible à atteindre, mais s'il restait plus ou moins le but. Or, tant que l'on aura une pédagogie du modèle, on sera amené à confondre notre compétence en tant que récepteur et notre compétence en tant qu'émetteur. Or il est bien vrai par exemple qu'il y a un travail considérable à faire pour que chaque francophone soit en état de démêler les fausses complexités et soit en état de comprendre n'importe quel message ; mais par bonheur le but n'est pas que chaque français s'exprime comme Chateaubriand, Giscard d'Estaing, ou tel ou tel autre. La distinction entre norme et sur-norme se fonde sur l'opposition entre éléments nécessaires et facultatifs du codage et corrélativement sur la différence nécessaire entre notre savoir comme récepteur et comme émetteur. L'idéologie de la norme agit en confondant tout cela.

directement eux-mêmes dans le cadre scolaire faire passer les expériences en question ce qui naturellement rendait plus difficile de leur demander d'avoir parallèlement un enregistrement oral. Bien entendu il s'agit ici d'une pré-enquête. Dans une enquête définitive il faudrait multiplier les types de noniement linguistique, en introduisant en particulier, comme l'ont fait des auteurs comme Robinson, les cas où l'enfant est amené à expliquer à un autre enfant le manie-
ment d'un appareil compliqué et où on s'aperçoit que les résultats sont très différents de ce qu'on obtient par exemple dans un récit.

- On retrouve l'utilité des méthodes comme celle qu'on a présenté ce matin, c'est-à-dire que nous ne devons plus uniquement chercher à comparer des moyennes entre deux groupes, ni non plus, c'est déjà un progrès, des écarts-types, des dispersions autour de la moyenne, nous devons nous demander ; y-a-t-il un certain nombre de stratégies linguistiques possibles, et ces stratégies linguistiques sont-elles ou non associées et cohérentes, dans des groupes sociaux différents.

Il est bien entendu qu'on ne va pas trouver un déterminisme simple entre classe sociale et type de stratégie linguistique. En revanche on peut trouver, sans doute une image beaucoup plus stricte dans les classes défavorisées de la norme linguistique qui fera que en cas de déficit lexical, le "droit" d'utiliser une périphrase syntaxique ne sera pas utilisé.

- Alors très très rapidement sur le premier point, la critique exacte que je fais à Bernstein c'est de dire que la dichotomie code restreint, code élaboré, est une mauvaise dichotomie, il faut une trichotomie, code en situation, code hors situation, sur-norme, et donc sa notion de langage élaboré n'est pas une notion univoque. Et que d'autre part les indices qu'il utilise, du type par exemple, complexité syntaxique, on et pour soi, mélange des phénomènes fonctionnels et des phénomènes non fonctionnels. Du point de vue linguistique, sont des indices trop gros. On peut même aller plus loin que ce que vous dites, car une des hypothèses que l'on avait faites, c'est que c'est justement dans les milieux sociaux favorisés, comme l'enquête de Simon l'a montré d'ailleurs, qu'on laisse l'enfant utiliser le langage à vide, - ou plutôt avec une fonction essentiellement fantasmatique et absolument pas en vue de l'organisation du réel, alors que dans les milieux sociaux défavorisés, la fonction de communication en situation est beaucoup //

plus forte. Or on peut penser que c'est dans la mesure où on laisse en quelque sorte l'enfant fantasmer et utiliser le langage d'une manière irrationnelle que peu à peu il acquiert des comportements extrêmement compliqués du type arriver à parler de ce qui n'est pas présent, etc. et c'est en quelque sorte en mentant, en fantasmant, en discutant, etc... que l'enfant arrive peu à peu à faire le détour qui fait que le langage ensuite devient éventuellement un instrument de codage rationnel. Autrement dit que, là je suis bien d'accord avec vous, je ne voudrais pas représenter un fonctionnalisme au sens étroit du type pour lequel la seule chose bonne, vraie, bien etc, c'est d'arriver à coder l'expérience au sens de la mettre dans la boîte bien carrée des mots.

- Là encore c'est un lièvre ~~extrêmement important~~ qui vient d'être soulevé, c'est celui de la maîtrise des données déterminantes d'une situation expérimentale. Ce qui vient d'être dit ne fait penser à un livre connu "Pygmalion à l'école". Il est certain que l'attente de l'observateur et de l'observé retentit sur les résultats. C'est absolument capital, et il est sûr que dans beaucoup d'expériences célèbres et considérées comme très rigoureuses, en réalité on n'a pas maîtrisé ces variables là qui sont extrêmement importantes. Je voudrais revenir en arrière. On posait la question d'une hiérarchie des situations : qu'est-ce que cela apporte aux enfants de constater une certaine égalité socio-culturelle devant la situation racontée, verbalisée ? Il me semble là que c'est plus un problème à poser aux pédagogues qu'un problème expérimental. Et là je pense à des difficultés rencontrées dans le cadre d'expériences de validation IERDP. Les enseignants souvent n'ont pas distingué situation expérimentale et situation pédagogique. Et ils ont tendance à penser qu'à partir du moment où l'expérimentateur choisit tel type de tâches à proposer aux enfants, pour des raisons expérimentales et parce qu'on peut contrôler ainsi tel ou tel type de variable, cela signifie qu'il faut valoriser ces tâches sur le plan pédagogique. Or il n'en est rien. Il me semble que nous aurions intérêt quand nous présentons nos travaux à des pédagogues, à bien distinguer ces deux types de situation. Les situations pédagogiques, la hiérarchie des situations pédagogiques, c'est une chose. Les contraintes d'une situation expérimentale, c'en est une autre. Et je crois que le problème posé était plutôt un problème pédagogique qu'un problème de situation expérimentale.

R. JOLIVETUniversité de Lausanne

La recherche dont les grandes lignes et les principaux résultats actuels vont vous être présentés, est en cours au sein du groupe de linguistique fonctionnelle de l'I.F.R.D.P. C'est à Horteza Mahmoudian que nous en devons l'idée et la première impulsion. L'analyse syntaxique et le codage de cette analyse ont été accomplis par le signataire de ces lignes et par Danièle Manesse, qui a dû depuis malheureusement pour nous, interrompre sa collaboration.

Pour les participants à ce colloque, plutôt centré sur l'analyse des textes d'enfants, notre intervention n'aura sans doute qu'une valeur informative. En effet si, comme pour les recherches animées par François Charpin à Limoges ou l'équipe du C.R.A.L. de Nancy, nous utilisons également un ordinateur, il n'est pas dans nos objectifs, pour le moment, de pousser l'analyse au dela du cadre de la phrase. Par ailleurs l'échantillon analysé provient d'un corpus constitué ~~d'enregistrements de langue~~ orale exclusivement.

I. Le corpus

Ce corpus a été recueilli, il y a maintenant quatre ou cinq ans déjà, dans des classes de sixième et de cinquième d'un C.E.S. de la région parisienne, par les soins du professeur de français de ces classes, Madame Baudrillard, qui a également effectué la transcription phonologique et orthographique.

Il est difficile de caractériser d'un mot ces productions. La procédure pour les obtenir est restée constante ce qui assure, de ce point de vue, une certaine homogénéité à l'ensemble. Un thème est proposé et une discussion s'engage entre les élèves réunis dans la salle de classe. Ces discussions (histoires d'animaux, récits de vacances, la mode des cheveux longs etc). ont été enregistrées au vu et au su des participants. Il s'agit donc d'une sorte de "discours" (j'emploierai ce terme dans la suite, faute

d'en trouver un meilleur) des enfants, mais discours relativement libre seulement. Les moments de vrai dialogue semblent avoir été relativement rares.

Homogénéité donc du corpus. Mais il est bien évident que ce corpus est bien loin de prendre en compte toutes les variables sociales, géographiques et de situation. Pour nous, il ne s'agit pas d'une limitation critiquable puisque nous n'avons pas eu d'autre objectif que de mener à bien une pré-enquête, c'est-à-dire, surtout, d'évaluer l'intérêt, les possibilités et les limites du traitement automatique dans le domaine syntaxique. On fera bien, toutefois, de garder en mémoire ces limitations considérables, pour ne pas chercher à en faire trop dire aux résultats qui suivront. Ils ne reflètent qu'un usage de la langue et on aurait tort de les généraliser sans précautions, même à un usage analogue d'enfants de même niveau scolaire, à plus forte raison à une notion aussi "floue" que la "langue de l'enfant" (quel enfant ? quelle langue ?).

De ce corpus, un échantillon, que certains sondages nous incitent à considérer comme représentatif, a été extrait de façon aléatoire. Il s'agit d'un échantillon de discours et non d'un simple échantillon de phrases. Ce ne sont pas des phrases qui ont été tirées au sort dans le corpus pour figurer dans l'échantillon mais des discours. Disposant d'un ensemble de discours d'enfants nous avons laissé au hasard le soin de désigner ceux qui figureraient dans l'échantillon analysé. Il pourrait être intéressant de procéder à la même analyse sur un échantillon de phrases et non un échantillon de discours. Si l'on mettait en évidence des différences statistiquement significatives il faudrait sans doute les attribuer à l'existence de contraintes combinatoires, au moins à l'état de tendance, entre les constructions syntaxiques des phrases d'un même discours (répétitions évitées, souci de variété, cohérence...). Cette étude est un des prolongements possibles de notre enquête, qui rejoindrait sans doute certaines des préoccupations des participants à ce colloque.

./.

L'échantillon extrait du corpus a été ensuite "nettoyé", pour reprendre l'expression de Denise François. C'est-à-dire qu'on en a éliminé les mots-phrases (non ; oh si ; ah oui ; etc.) et deux ou trois phrases inanalysables... au moins dans le cadre linguistique qui est le nôtre. Ces mots-phrases et ces phrases inanalysables ne sont pas perdus pour autant mais notés à part et non soumis au traitement automatique.

Un exemple : dans la phrase,

"Castrer, ça veut dire l'empêcher de voir une chatte,
quoi, enfin, pas l'empêcher mais..."

la partie ^{soulignée} n'a pas été analysée.

De même, à l'intérieur d'une phrase analysable, il peut y avoir un ou plusieurs éléments qui traduisent une hésitation, amorcent une construction finalement abandonnée etc... Ces éléments ont également été extraits. Ainsi pour le que de la phrase suivants :
"Et pis quand la télé est mise, dès qu'il voit que ... l'image, il s'allonge sur les bras (du fauteuil R.J.) "

En doute que ces ratés, ces renords, mériteraient une analyse poussée, dont des applications pédagogiques sont loin d'être exclues. Mais tel n'était pas notre objet. D'autres méthodes d'ailleurs que le traitement automatique sont sans doute plus indiquées pour traiter ce type de faits.

De toute façon ce ne sont pas des considérations de "bon usage" qui ont guidé ce "nettoyage" de l'échantillon, mais uniquement, si l'on peut risquer ce néologisme, d'"analysabilité". On s'est par ailleurs conformé au principe classique des philologues pour l'édition des textes : celui de la correction minimum.

L'échantillon analysé comporte exactement 254 phrases. La définition de la phrase, pratiquement, n'est évidemment pas chose facile. Le point de départ a été la définition d'André Martinet : "énoncé dont tous les éléments se rattachent à un prédicat unique ou à plusieurs éléments coordonnés" (1), mais il est resté, bien sûr, des cas difficiles. Un exemple : "Alors en s'amusant il met la patte, il essaie de la retirer et pis il tire le rideau".

Une, deux ou trois phrases ? Il serait imprudent de s'appuyer sur la présence de "et pis" pour affirmer qu'on a ici des prédicats coordonnés. En effet cet élément, semble, dans l'ensemble du corpus, avoir une fonction plus démarcative (ou de cohérence, d'enchaînement peut-être) que proprement coordinative. Certains

Certains enfants commencent presque toutes les phrases de leur discours par "et pis" (ou encore : "alors"). Pour trancher les cas de ce genre il fallait soit introduire un indice spécial (d'"hésitation" en quelque sorte) dans le codage de l'analyse soit s'en tenir à une décision arbitraire. C'est la seconde solution qui, à l'époque, a été retenue. Le principe adopté a été celui de la segmentation maximum en cas d'hésitation. Donc trois phrases dans l'exemple cité plus haut.

264 phrases c'est évidemment bien peu... mais s'est aussi beaucoup car ces 264 phrases ont livré plus de 1 700 unités syntaxiques (unités de base de l'analyse, cf. ci-dessous) et que chacune de ces unités a exigé la perforation d'au moins deux cartes IBM, perforations effectuées... à la main au moyen d'une pointe métallique sur cartes pré-perforées ! Nous n'évoquerions pas ce "folklore" qu'il vaut mieux oublier s'il n'avait pour conséquence un nombre d'erreurs de codage d'ailleurs faible semble-t-il (les erreurs éventuelles d'analyse mises à part) mais non nul. De toute façon, de quelque manière qu'on s'y prenne, avant qu'un ordinateur puisse fournir des résultats il y a un grand nombre d'opérations "humaines" qui interviennent qui sont autant de sources possibles d'erreurs. Ceci dit pour ceux qui seraient encore tentés d'accorder un brevet d'infaillibilité à tout ce qui sort d'une imprimante d'ordinateur.

L'unité de base de l'analyse est donc ce que nous appelons l'unité syntaxique. C'est le syntagme minimum qui dans une phrase indique sa relation aux autres éléments soit principalement par sa présence seule (cas des "autonomes" comme la nuit ou rapidement dans "le poète écrit ses vers la nuit/rapidement"), par sa position dans la phrase (sujet, objet, etc.) ou encore par l'utilisation d'un élément spécialisé dans l'indication des relations (motème fonctionnel dans la terminologie d'André Martinet, c'est-à-dire prépositions et conjonctions de subordination principalement). Les modalités, déterminants grammaticaux du nom (défini, indéfini, pluriel, tout, etc.) et du verbe (temps, modes, voix, etc.) sont traitées comme des caractéristiques du noyau nominal ou verbal qu'elles déterminent et non comme des unités syntaxiques indépendantes.

II. La grille d'analyse

Chaque unité syntaxique est décrite par des informations concernant sa position dans la phrase, sa fonction, la classe de l'élément remplissant cette fonction et les modalités qui l'accompagnent éventuellement, et enfin la place de cette fonction dans la hiérarchie fonctionnelle d'ensemble de la phrase.

La position de chaque unité est codée par un chiffre correspondant à son rang d'apparition dans la phrase. Cette information permet d'effectuer certaines études, sinon directement sur l'ordre des "mots", du moins sur les positions préférentielles des différentes fonctions de la phrase, soit en considérant l'ensemble des fonctions de la phrase, soit en ne tenant compte que des fonctions primaires, en rapport direct avec le prédicat de la phrase, comme l'a fait Jean-Claude Corbeil (2). Du point de vue de leur position, les unités syntaxiques de la phrase,

"Ce chiot il est très marrant"

seront numérotées de 1 à 5 comme suit :

ce chiot / il / est / très / marrant

1 2 3 4 5

Pour la fonction de chaque unité syntaxique, un inventaire de neuf possibilités a été établi. Chaque fonction est d'autre part caractérisée par rapport à la nature directe (fonction primaire) ou indirecte (fonction non primaire) de sa relation au prédicat de la phrase. On précise en outre s'il y a ou non coordination entre cette fonction et une autre fonction dans la phrase.

La classe de l'élément qui remplit la fonction considérée est ensuite indiquée. La grille d'analyse comporte ainsi plus de trente classes. On a codé également le caractère de syntème de certains éléments (essentiellement des composés) mais sur ce point nous sommes peu satisfait des résultats, l'analyse a été trop flottante faute d'une réflexion théorique suffisante. Enfin, lorsqu'un élément introduisait une expansion par que ("il voit que..." ; "la crainte que...") le fait était également codé.

A tous ces renseignements s'ajoute l'indication des modalités éventuelles d'un nom ou d'un verbe.

Toutes ces informations s'inscrivent en code sur une première carte perforée.

Une deuxième carte (et éventuellement une troisième), consacrée à la même unité syntaxique, décrit les rapports qu'elle entretient avec les autres unités de la phrase auxquelles elle est directement liée, soit qu'elle les détermine (rapport en amont) soit qu'elle soit déterminée par elles (rapport en aval). Les unités en amont (en fait, dans notre codage, l'unité en amont, car il ne peut y en avoir qu'une seule) sont codées par le rappel de leur fonction, primaire ou non, et de la classe de l'élément qui la remplit.

Les unités en aval sont codées simplement par leur fonction et par leur classe. On possède donc toutes les informations nécessaires pour rendre compte des relations entre les unités syntaxiques d'une phrase, aussi bien en ce qui concerne la nature de ces relations que leurs points d'incidence. Ces informations permettraient la mise au point d'un programme de visualisation de la structure syntaxique de chaque phrase, sous forme d'arborescence par exemple. Mais la mise au point d'un tel programme reste délicate et les résultats seraient moins intéressants que les comptages variés qu'il est possible de faire sur chaque variable codée.

Pour illustrer ce qui précède voici quelles sont les informations codées pour la courte phrase de l'échantillon :

"Ce chiot il est très marrant"

Outre les informations sur la position des différentes unités syntaxiques, on codera donc :

- pour "ce chiot" : sujet-primaire-coordonné par reprise

nom commun animé

non synthématique

modalité : démonstratif

en amont : prédicat-ni primaire ni non primaire

verbe d'état

en aval : rien

- pour "il" : sujet-primaire-coordonné par reprise

pronom personnel 3^o personne (du singulier) ./.

- non synthématique
 en amont : prédicat-ni primaire ni non primaire
 verbe d'état
 en aval : rien
- pour "est" : prédicat-ni primaire ni non primaire-non coordonné
 verbe d'état
 non synthématique
 modalités : présent-indicatif
 en amont : rien
 en aval : sujet-nom commun animé
 sujet-pronom personnel 3^o personne
 attribut-adjectif
- pour "très" : autonome-non primaire-non coordonné
 adverbe
 non synthématique
 en amont : attribut-primaire
 adjectif
 en aval : rien
- pour "marrant" : attribut-primaire-non coordonné
 adjectif
 non synthématique
 en amont : prédicat-ni primaire ni non primaire
 verbe d'état
 en aval : autonome-adverbe

La grille d'analyse complète qui a été utilisée pour la pré-enquête figure ici en annexe. A son sujet d'importants problèmes, théoriques et pratiques, se posent.

Il y a d'abord des insuffisances évidentes. Par exemple, on n'a pas retenu la fonction "apposition"... il est vrai que dans notre échantillon elle ne se rencontre jamais. Mais il faudrait bien sûr la prévoir pour l'analyse d'autres corpus, relevant de niveaux de langue différents. De la même façon, distinguer la fonction épithète antéposée de l'épithète postposée était sans intérêt pour notre échantillon. En ce qui concerne la fonction ou plutôt l'ensemble de fonctions, baptisé "syntagme à indicateur de fonction", il est certain que nous avons privilégié à l'excès le signifiant de cet ensemble de fonctions aux dépens des signifiés de ces relations. Sans retomber dans l'excès d'analyse sémantique de la grammaire traditionnelle, il existe des critères fonctionnels permettant de distinguer la fonction "si" de la fonction "quand" ou "après que" par exemple.

./.

Du point de vue des classes retenues il y aurait aussi plusieurs remarques à faire. Ce que nous avons regroupé sous l'appellation traditionnelle d'"adverbe" comporte en fait des éléments au comportement fonctionnel très différent.

D'une façon générale cette grille d'analyse, pour des raisons en partie matérielles, a certainement été faite trop vite ; elle est insuffisamment mûrie, insuffisamment fondée du point de vue théorique, fonctionnaliste en particulier, et il est certain que nous lui apporterons d'assez notables modifications s'il nous est donné la possibilité de passer de la pré-enquête à une enquête plus étendue.

Mais, de toute façon, un problème général se posera toujours. Celui de la "finesse" de toute grille d'analyse. Il est bien évident que l'on peut faire varier dans des limites assez larges le degré de finesse d'une grille d'analyse.

Où s'arrêter ? Il n'y a sans doute pas de réponse définitive à cette question. Tout dépend en fait des conditions de l'enquête et de la taille du corpus en particulier. Ce qu'une étude sur corpus met bien en évidence ce sont les faits fréquents. Dès que l'on s'intéresse aux phénomènes rares ou que l'on désire affiner l'analyse, auquel cas les critères se multiplient, il faut soit traiter un corpus considérable, soit avoir recours à d'autres méthodes (tests, questionnaires, etc.).

L'utilisation d'un ordinateur est sans doute un puissant secours pour l'étude de la combinatoire des unités, mais elle suppose toujours l'analyse préalable de ces unités et le codage de cette analyse, opérations pour lesquelles c'est la vitesse de travail de l'homme, non celle de la machine, qui intervient. L'ordinateur ne permet donc pas d'augmenter de façon considérable les corpus traités, au moins pour l'analyse syntaxique, il permet simplement d'exploiter plus "à fond" un corpus. Dès lors, il est permis de se demander s'il y a vraiment intérêt à multiplier les variables prévues par une grille d'analyse. En ce qui concerne la pré-enquête que nous avons menée, la grille utilisée était, d'une certaine façon, trop riche, puisqu'il faut le plus souvent opérer sur des regroupements de variables (par exemple, toutes les classes de noms) pour obtenir des résultats statistiquement significatifs.

Du point de vue fonctionnaliste on pourrait également critiquer la construction d'une grille de classes de monèmes... Les classes ne sont pas un donné à priori mais la classification doit découler de l'examen des latitudes fonctionnelles. En toute rigueur on n'aurait donc pas le droit de prévoir des classes avant l'examen des latitudes fonctionnelles de monèmes dont rien ne nous dit, au départ, qu'ils se comportent effectivement comme les descriptions d'autres niveaux du "français" nous l'indiquent. Mais, en réalité, l'hypothèse que nous formons est la suivante. Le système linguistique étudié n'est pas irréductible au système partagé par les francophones adultes entre lesquels il y a intercompréhension. Hypothèse contestable, sans doute, pour les partisans d'un structuralisme "pur et dur", strictement formaliste. Mais hypothèse admissible - recevable en tout cas - par des "réalistes" qui, sans nier l'existence de spécificités caractérisant chaque comportement linguistique individuel, subordonnent néanmoins ces différences à l'existence effective de processus de communication.

Nous n'étudions pas la langue du corpus comme un linguiste pourrait le faire d'une langue jamais décrite et dont il ne sait rien, mais par comparaison, en partant du cadre fourni par l'analyse fonctionnelle du français "normal". Ce n'est qu'à partir de là qu'une exploitation pédagogique ultérieure sera possible. Pour simplifier, disons que si un écart se marque entre le comportement de classes du corpus et le comportement fonctionnel de ces mêmes classes en français "normal", c'est qu'il ne s'agit pas, rigoureusement parlant, des mêmes classes, mais, du même coup, est mis en évidence un champ possible pour l'action pédagogique. Il reste, bien entendu, que ce français "normal", base de la comparaison, serait par ailleurs à définir plus précisément. Il n'est pas plus simple, plus "fixé", moins soumis aux variations socio-culturelles ou situationnelles que les variétés de français que l'on pourrait recueillir dans différents établissements scolaires et selon des procédures variées. L'image que nous en donnent les grammaires est assurément déformée et sans doute déformante pour les pédagogues et pour nous également. Le dilemme dans lequel, finalement, nous nous trouvons plus ou moins enfermé pourrait peut-être être résumé ainsi.

La construction d'une grille d'analyse suppose des analyses préalables, sur d'autres matériaux. Rien ne nous assure, a priori, que la grille ainsi construite sera satisfaisante pour les matériaux auxquels on décide ensuite de l'appliquer. En somme, toute grille d'analyse constitue en un certain sens une norme, et le corpus auquel on l'applique n'est jamais vraiment analysé complètement "en lui-même et pour lui-même", mais toujours de façon plus ou moins comparative. Ce cercle n'est pas très satisfaisant du point de vue linguistique. Mais il n'est pas très facile d'en sortir et peut-être, d'un point de vue pédagogique alors, n'est-il pas très intéressant d'essayer même d'en sortir... Pour mener une analyse linguistique plus rigoureuse il faut, de toute façon, un corpus très considérable.

D'autres questions encore peuvent être soulevées, qui intéressent la grille d'analyse. Que faisons-nous des cas que Lucien Tesnière (3) a regroupé sous l'appellation de "translation" : substantivation d'un adjectif par exemple. Va-t-on coder "le rouge" dans "le rouge lui va bien" comme un substantif ou comme un adjectif ? En fait, on se fonde sur le comportement statistiquement le plus fréquent du monème considéré pour l'affecter à une classe. Les faits de ce genre n'échappent pas pour autant puisqu'un adjectif que l'ordinateur annoncera comme "sujet" et accompagné d'une modalité nominale (ici : "défini") sera évidemment un adjectif substantivé. Mais d'autres situations pourraient s'avérer moins simples.

Autre question. Si nous définissons une classe de noms communs "abstrait" sur la base de la possibilité d'introduire une expansion par que ("la crainte qu'il vienne..."), va-t-on considérer le substantif crainte comme un nom commun abstrait même dans "Sa crainte était visible". ? C'est la solution que nous avons pu retenir grâce au grand nombre d'informations codé pour chaque unité syntaxique. Un traitement approprié permet en effet de faire le départ entre ceux des noms communs abstraits qui sont effectivement accompagnés d'une expansion de ce type et ceux qui ne profitent pas de cette latitude.

Un dernier point à signaler, et qui est très important. On a noté que dans le cas de 100%, et au moment de la lecture de la norme de chaque unité syntaxique, d'une façon plus générale, les

Un dernier point enfin, et non des moindres. On aura noté que nous ne codons pas, pour le moment, la forme lexicale du noyau de chaque unité syntaxique. D'une façon plus générale, nous ne nous sommes pas intéressé aux faits de morphologie, au sens qu'André Martinet donne à ce mot, sauf lorsque ces faits sont fréquemment susceptibles de devenir pertinents du point de vue syntaxique (cf. les types de "pluriel" dans la grille d'analyse, en annexe). Cette insuffisance de notre grille est volontaire. Parce que ce codage est relativement simple et ne pose pas de problème trop délicat, il nous a semblé sans inconvénient d'en faire abstraction pour ce qui n'est, rappelons-le, qu'une pré-enquête... compte-tenu notamment des conditions matérielles dans lesquelles s'effectuait par ailleurs le codage des variables syntaxiques. Bien entendu, nous n'escamoterions pas cet aspect dans une enquête à plus grande échelle.

C'est donc sur l'ensemble des informations que livre l'application de la grille d'analyse à l'échantillon du corpus que l'ordinateur a ensuite travaillé pour livrer un certain nombre de données numériques.

III. Quelques résultats

On a obtenu jusqu'à présent les réponses à une quarantaine de questions de nature quantitative, certaines très simples, d'autres plus élaborées, qui font intervenir des croisements de variables (par exemple : combien y a-t-il de sujets, et de quelles classes, repris par un pronom, sur le modèle de "ce chiot il ..."). Il ne peut être question de présenter ici tous ces résultats (plus de 1 500 pages d'ordinateur...) dont l'analyse et l'interprétation sont loin d'être achevées. On ne fera donc état ici que de quelques-uns d'entre eux et des plus nets.

Avant toute analyse linguistique plus précise, l'une des premières questions que l'on peut poser est la suivante : quelle est la "longueur" moyenne des phrases dans l'échantillon ? Le nombre d'unités syntaxiques par phrase est une bonne mesure de leur longueur. Plus intéressante en tout cas, car plus proche de la réalité linguistique fonctionnelle, que le nombre de "mots" de la phrase, car une définition du "mot" ne peut être qu'arbitraire (mot graphique) ou très difficile à manier (mot phonique)./.

Si l'on trace la courbe du nombre d'occurrences de phrases de n unités syntaxiques (n croissant de un en un à partir de la valeur deux puisque les mots-phrases ont été exclus de l'échantillon) on obtient pour les discours de chaque élèves des courbes assez divergentes et très irrégulières. Mais la courbe obtenue en faisant la somme de ces courbes est beaucoup plus régulière et se rapproche d'une courbe "en cloche" dissymétrique : montée rapide de 2 à 5 unités syntaxiques par phrase, descente rapide de 5 à 8 unités syntaxiques par phrase, puis plus lente et plus "accidentée" jusqu'à 19 unités syntaxiques, longueur maximum observée dans l'échantillon. La moyenne s'établit à 6, 5 unités syntaxiques par phrase, le mode de la courbe (longueur la plus fréquente) à 5 unités syntaxiques par phrase.

Cette courbe est intéressante en elle-même, comme une caractéristique de l'échantillon (qu'il faudra bien sûr compléter d'une étude statistique), mais une partie de sa signification éventuelle nous échappe, faute de points de comparaison. Y-t-il réellement là une caractéristique de la langue du corpus ?

On peut néanmoins partir de cette constatation : le nombre d'unités syntaxiques par phrase le plus probable est de 5. Quelles sont les cinq fonctions les plus fréquentes dans l'échantillon ? Il s'agit, par ordre de fréquence décroissante, des fonctions suivantes : sujet, prédicat, syntagme à indicateur de fonction, autonome et complément d'objet direct. Attirons l'attention sur le fait qu'aucune des fonctions plutôt spécifiques des adjectifs (épithète, attribut du sujet ou de l'objet) ne figure parmi les cinq fonctions les plus fréquentes. La fonction épithète, la plus fréquente des trois (plus de la moitié des adjectifs sont épithètes), représente moins de 3 % de l'ensemble des fonctions, contre plus de 13 % à la fonction de complément d'objet et près de 24 % à la fonction sujet. Il y a de toute façon peu d'adjectifs dans l'échantillon : 68 occurrences, soit 4 % de l'ensemble des classes.

Si, partant de ces cinq fonctions, nous examinons ensuite les classes qui constituent le plus souvent le noyau de ces fonctions, nous obtenons les résultats suivants.

Plus de 80 % des fonctions sujet ont pour noyau des pronoms, notamment il, on et je qui, à eux trois, représentent près de 60 % du total des éléments en fonction sujet.

Plus de 80 % des fonctions prédicatives sont remplies par les classes verbales. Parmi celles-ci on compte en moyenne deux fois plus de verbes objectifs que de verbes non objectifs.

En fonction autonome, on rencontre, dans plus de 80 % des cas encore une fois, des unités spécialisées dans cette fonction : les adverbes.

Les verbes ne représentent que moins de 20 % des éléments en fonction d'objet direct, dans plus de 80 % des cas on trouve dans cette fonction des noms ou des pronoms.

C'est la fonction de syntagme à indicateur de fonction qui manifeste la plus grande variété quant aux classes susceptibles d'en constituer le noyau. Verbes d'une part, noms et pronoms de l'autre, se partagent presque également : 40 % aux premiers et 60 % aux seconds.

Par ailleurs, les modalités nominales les plus fréquentes sont le défini, le pluriel et l'indéfini : pour les modalités verbales : l'indicatif, le présent et également un bon nombre d'infinitifs.

Si l'on tient compte enfin des indications fournies quant aux positions des fonctions dans la phrase, nous obtenons les deux séquences suivantes, à peu près aussi fréquentes l'une que l'autre :

Autonome+Sujet+Prédicat+Objet+Syntagme à indicateur de fonction
Sujet+Prédicat+Objet+Autonome/Syntagme à indicateur de fonction

Cet ensemble de données permet de construire une sorte de "phrase-type" pour le corpus. En voici deux exemples, d'ailleurs approximatifs et fictifs mais qui ne paraîtront sans doute pas trop invraisemblables : "Souvent il griffe les rideaux avec sa patte".

"L'arbitre arrête le match des fois après l'heure".

Il va de soi que la probabilité d'obtenir des phrases exactement sur le modèle de la phrase-type reste très faible, puisque l'on fait intervenir un grand nombre de variables.

Si, partant de cette "phrase-type", on examine les possibilités d'enrichissement syntaxique qu'elle comporte, on fait intervenir l'analyse des liens fonctionnels entre unités syntaxiques. Dans quelle mesure chacune des cinq fonctions de la "phrase-type" est-elle susceptible d'expansions et quelles sont ces expansions ?

Pour la fonction prédicative, cette recherche est sans objet. Les expansions les plus fréquentes sont évidemment les quatre autres fonctions de la "phrase-type".

L'expansion d'un syntagme à indicateur de fonction est très fréquente, un verbe est en effet assez souvent le noyau de cette fonction.

Un complément d'objet sur trois, en moyenne, constitue le noyau d'une expansion, celle-ci étant, le plus souvent, un syntagme à indicateur de fonction.

Les expansions d'un autonome sont bien plus rares, une fois sur sept environ. Il s'agit alors principalement d'adverbes ("bien souvent", "très longtemps").

Le cas le plus intéressant est celui de la fonction sujet, qui constitue très rarement le noyau d'une expansion. Certes, ce résultat était partiellement prévisible puisque ce sont les pronoms, et notamment les pronoms personnels, qui apparaissent le plus souvent en fonction sujet. Mais cette fonction est aussi la mieux représentée dans l'échantillon : on compte près de 350 sujets et, malgré tout, plus de 50 noms dans cette fonction. Or il n'y a pas 10 expansions (faits de coordination mis à part) pour cette fonction. Sans doute y a-t-il là une caractéristique de la langue parlée mais les linguistes ont depuis longtemps noté le caractère assez spécifique de la fonction qui unit le sujet au prédicat, liaison constitutive d'une part importante des énoncés minimaux en français, et il est permis de se demander si une limitation de l'expansion des éléments en fonction sujet ne constitue pas une tendance relativement constante de cette fonction en français. Seules d'autres études sur ce point particulier, portant sur des usages et des niveaux de langue différents permettraient de vérifier cette hypothèse. ./.

Plus généralement, l'étude des liens hiérarchiques entre les unités syntaxiques de chaque phrase permet également de chiffrer l'aptitude de chaque classe à remplir la fonction prédicative de même que l'aptitude de chaque classe et de chaque fonction à constituer le noyau d'une expansion. Nous n'exposerons pas les calculs, d'ailleurs très simples, qui permettent d'obtenir ces indices, nous nous en tiendrons aux résultats (4).

L'étude de l'aptitude des classes à remplir la fonction prédicative fournit des résultats absolument triviaux dans le cadre d'une seule pré-enquête, mais la banalité des résultats obtenus ne doit pas masquer l'intérêt que peut présenter, à notre avis, cet indice pour une étude comparative avec d'autres corpus. Trois grandes catégories se dégagent : les verbes, très souvent en fonction prédicative, les noms et les pronoms non personnels ainsi que le pronom de troisième personne du singulier qui ne sont qu'assez rarement prédicats et enfin les autres classes d'unités qui ne sont que très rarement ou jamais en fonction prédicative.

L'étude de l'aptitude des unités syntaxiques à constituer le noyau d'expansions peut partir soit des fonctions soit des classes.

Du point de vue des fonctions, c'est évidemment la fonction prédicative qui est la plus riche en expansions, viennent ensuite la syntagme à indicateur de fonction (dont le noyau est souvent verbal), l'attribut du sujet, l'objet etc. le sujet vient tout à fait en dernier lieu après l'autonome et l'épithète.

Du point de vue des classes, on observe une forte corrélation entre l'aptitude à fonctionner comme prédicat et l'aptitude à constituer le noyau d'expansions. Cela n'a rien d'étonnant puisque bon nombre des unités syntaxiques d'une phrase sont des expansions du prédicat : les classes qui sont souvent en fonction prédicative ont donc naturellement un nombre d'expansions élevé. Néanmoins, la comparaison détaillée de ces deux indices, comparaison que nous ne pouvons pas mener ici, devrait être intéressante, car, à côté d'une remarquable analogie d'ensemble, il y a, pour certaines classes, de fortes divergences.

Une autre présentation possible des résultats les plus généraux consiste à partir des classes et à étudier leurs latitudes fonctionnelles.

Certaines classes regroupent des unités qui ne peuvent remplir qu'une seule fonction : actualisateurs d'un prédicat nominal, prépositions et conjonctions, le pronom on, les coordonnants. D'autres classes ont des latitudes fonctionnelles très limitées, ainsi les adverbes mais aussi, et plus curieusement, les pronoms personnels. Ceux d'entre eux qui sont bien attestés (je, tu, il/elle et ils/elles) apparaissent dans la grande majorité des cas (de plus de 71 % pour il/elle à plus de 87 % pour ils/elles) en fonction sujet. Autant dire que les formes me/moi, te/toi, le/lui/la/elle, les/eux/elles sont relativement rares dans l'échantillon, sauf pour celles d'entre elles qui peuvent apparaître comme sujets en cas de reprise (cf. "lui, il..."). On en compte en effet 57 occurrences en tout pour les 256 occurrences de ces pronoms toutes fonctions confondues, soit un peu plus de 22 % donc près de 76 % pour la fonction sujet. Or, pour l'ensemble du corpus, si l'on additionne les fonctions déjà complètement d'objet, de syntagme à indicateur de fonction et de sujet, cette dernière ne représente qu'un peu moins de 44 % du total. Il semble qu'il puisse donc y avoir une tendance, qu'il faudra vérifier par des méthodes statistiques plus rigoureuses, à limiter l'utilisation des pronoms personnels à la fonction sujet. En ce qui concerne les verbes, la fonction prédicative est bien entendue majoritaire, mais d'autres fonctions sont bien attestées, notamment celle de syntagme à indicateur de fonction et celle de complément d'objet. Ce sont en définitive les noms qui manifestent, comme on pouvait s'y attendre, les plus grandes latitudes fonctionnelles.

Nous avons présenté jusqu'ici les résultats concernant les faits les plus fréquemment attestés dans l'échantillon. On peut aussi s'intéresser aux catégories prévues par la grille d'analyse ou aux schèmes syntaxiques attestés dans d'autres usages de la langue et qui n'apparaissent jamais ou très rarement dans l'échantillon. Ces raretés ou ces absences sont intéressantes mais difficiles à interpréter. Certaines s'expliquent sûrement par la nature des enregistrements qui constituent le corpus. De toute façon, l'échantillon est tout à fait insuffisant pour que l'on puisse conclure de l'absence d'attestation à la né-connaissance par exemple, ou à des difficultés quant au maniement.

Sans entrer dans le détail des schèmes syntaxiques nous nous en tiendrons aux catégories prévues par la grille d'analyse qui n'apparaissent jamais ou seulement très rarement. Parmi les fonctions on a déjà signalé la rareté des fonctions spécifiques des adjectifs : l'attribut de l'objet est très rare (9 occurrences) de même que l'attribut du sujet (31) et l'épithète (43).

Pour les classes, nous ne rencontrons pas d'actualisateur non verbal, un seul numéral qui ne soit pas une modalité nominale (cf. "Trois sont venus") et deux occurrences seulement pour les pronoms personnels nous et vous.

Certaines modalités nominales n'apparaissent jamais, entre autres les possessifs ton/ta et votre/vos et encore : aucun, propre, différents, divers, nul, quel, tout ; d'autres modalités sont rares, ainsi les suivantes qui ne sont pas attestées plus de trois fois : démonstratif, notre/nos, leur/leurs, autre, même, seul, certain, chaque. Parmi les modalités verbales le passé antérieur et le futur antérieur ainsi que les formes surcomposées n'apparaissent jamais, le passé simple est rare mais pas totalement absent (6 occurrences... sur 436 modalités temporelles) de même que l'impératif, le subjonctif, les participes et le conditionnel. Très peu de passifs enfin (7) et encore moins d'interrogatives (5).

Tels sont, brièvement résumés, quelques-uns des principaux résultats obtenus jusqu'à ce jour. Nous pourrions en présenter bien d'autres mais les discussions qu'il faudrait alors mener pour les interpréter nous entraîneraient trop loin du sujet de ce colloque. D'ailleurs, tous les traitements ne sont pas terminés, d'autres résultats sont attendus qui porteront en particulier sur les modalités nominales et verbales. La présentation de cette sélection de résultats avait de toute façon essentiellement pour but d'illustrer ce que nous pouvons attendre de la méthode utilisée.

Que conclure ? Assurément bien des points resteraient à discuter et nous sommes bien conscient - surtout conscient peut-être - des limites de ce genre de travail. Encore une fois les résultats présentés ici ne sont pas généralisables, ils ne valent probablement que pour le corpus d'où l'échantillon a été extrait.

Plus généralement l'intérêt de l'utilisation de l'ordinateur nous apparaît, répétons-le, surtout pour la mise en évidence des phénomènes fréquents. D'où souvent l'impression que les résultats obtenus correspondent à ce que l'on savait déjà ... intuitivement. Il n'est peut-être pas mauvais, néanmoins, d'avoir des données concrètes et chiffrées pour appuyer ces intuitions ou ces constatations empiriques.

Enfin une part importante de la signification des données obtenues nous échappe par manque de bases de comparaison.

C'est-à-dire par manque d'études quantitatives analogues ou au moins partiellement comparables, portant sur des populations différentes, d'autres niveaux de langue etc. Ce n'est que lorsque nous disposerons de recherches de ce type que l'outil puissant d'analyse offert par la statistique pourra être mis pleinement à contribution. Alors seulement nous pourrions préciser celles des caractéristiques numériques du corpus - s'il y en a - qui le caractérisent en propre. D'où l'intérêt que nous portons aux travaux de François Charpin et à ceux du C.R.A.L. ... d'où aussi notre désir d'obtenir les moyens de passer de la pré-enquête à une enquête plus étendue.

Mars 1975

INTERVENTIONS

Nous avons répondu dans le cours de ce texte à la plupart des questions posées lors de l'exposé oral par les participants au colloque. Nous rappellerons ici l'intervention extrêmement intéressante de Frédéric François, à laquelle nous n'avons pas pu faire un sort car elle pose un problème tout à fait fondamental auquel nous ne voyons pas de solution dans le cadre concret qui est le nôtre dans cette recherche. Partant de la constatation

que le temps de codage est relativement libre, pour l'émetteur au niveau de l'écriture, alors qu'il est relativement lié au temps de la parole dans le discours oral, Frédéric François a dit à peu près ce qui suit (je cite d'après le script) :

"En fait nous avons tous fait l'hypothèse assez idéaliste, car ne tenant pas compte des conditions de la communication, que le discours oral est peut-être moins syntaxique que le discours écrit, mais que le mot "syntaxe" a un sens univoque appliqué à l'oral et à l'écrit. Or il se peut très bien qu'il y ait un grand nombre d'éléments qui servent à introduire la continuité du discours oral, et contribuent donc par là même à la mise en relation, et qui ne sont pas du tout syntaxiques au sens précis de Martinet..."

* Texte rédigé à partir d'une communication faite au colloque sur l'analyse des textes d'enfants, organisé par l'I.N.R.D.P. et le C.R.A.L, à Paris, du 24 au 26 avril 1974.

La plupart des réponses données aux questions posées par les participants ont été incluses dans ce texte, fortement remanié et mis à jour (mars 1975) par rapport à l'exposé oral.

- (1) André Martinet, Éléments de linguistique générale, Armand Colin, Paris, 1957, p. 131
- (2) Jean-Claude Corbeil, Les structures syntaxiques du français moderne, Klincksieck, Paris, 1971, 197 p. (cf p. 99 - 134)
- (3) Lucien Tesnière, Éléments de syntaxe structurale, Klincksieck Paris, 1969, p. 361-410
- (4) le mode de calcul de ces indices est exposé dans un article à paraître dans un prochain numéro de Recherches Pédagogiques

A N N E X E

Grille d'analyse

- Fonctions :
- autonome ("il sort la nuit/souvent")
 - objet direct
 - épithète
 - prédicat ou prédicatif ("je sais qu'il mange")
 - attribut du sujet
 - attribut de l'objet
 - sujet
 - syntagme à indicateur de fonction
 - actualisateur ("voilà le train" ; "c'est mon chien")

Chaque fonction est en outre caractérisée comme :

- primaire (relation directe avec le prédicat)
- non primaire (relation indirecte avec le prédicat)
- ni primaire ni non primaire (prédicat)

- non coordonnée
- coordonnée avec un coordonnant
- + coordonnée sans marque (Pierre, Jean, Paul et Gustave)
- "coordonnée" par reprise (Pierre, il mange)

- Classes :
- actualisateur non verbal (voici, voilà)
 - actualisateur verbal (c'est, il y a, il faut)
 - adjectif
 - adverbe
 - indicateur de fonction (prépositions)
 - indicateur de fonction conjonctif
 - indicateur de fonction relatif (inclus dans les pronoms relatifs)
 - non commun abstrait (expansion par que possible)
 - non commun animé
 - non commun inanimé
 - anthroponyme
 - toponyme
 - pronom non compatible avec des modalités nominales (ce, en, y)
 - pronom compatible avec des modalités nominales (celui-ci)
 - pronom personnel 1 (je, me, moi)

A N N E X E

- pronom personnel 2
- pronom personnel 3
- pronom personnel 4 (nous)
- pronom personnel 5
- pronom personnel 6
- pronom personnel 7 (on)
- verbe d'état
- verbe non objectif
- verbe non objectif avec infinitif à sujet identique
(il vient chercher ses affaires)
- verbe objectif
- verbe objectif à infinitif à sujet différent
(il laisse parler ses enfants)
- verbe objectif à infinitif à sujet identique
(il aime partir en vacances)
- verbe objectif à infinitif à sujet identique ou différent
selon qu'il y a ou non un objet (il sort prendre l'air/il
sort ses enfants prendre l'air)
- vrai pronominal
- syntème verbal (s'anerrecevoir de)
- coordonnant
- pronom réfléchi
- numéral non modalité (trois sont venus)

Modalités :

- négation autonome (jamais)
- négation pronominale (rien)
- collectif (plusieurs)
- défini
- démonstratif
- démonstratif discontinu (ce... ci)
- indéfini
- numéral
- partitif
- pluriel graphique
- pluriel contextuel (beaucoup de journaux et de livres)
- pluriel à liaison facultative (objets inanimés)
- pluriel à liaison facultative (objets animés)
- pluriel à liaison obligatoire ou variation formelle (de petits
hommes ; beaucoup de chevaux)
- possessif 1 (mon)
- possessif 2

ANNEXE 3

- possessif 3
- possessif 4
- possessif 5
- possessif 6
- quantitatif toujours suivi du pluriel (nombre de)
- + quantitatif suivi d'un singulier obligatoire (beaucoup de courage)
- quantitatif admettant l'opposition singulier/pluriel (bien du travail/bien des travaux)
- aucun
- autre
- même
- propre
- seul
- certain
- chaque
- différents
- divers
- nul
- quel
- quelque
- tel
- tout
- présent
- imparfait
- futur
- passé
- passé composé
- plus-que-parfait
- passé antérieur
- surcomposé
- indicatif
- subjonctif
- conditionnel
- impératif
- infinitif
- participe
- passif avec le verbe "être"
- passif avec "se" (le vin se vend bien)
- interrogation par "est-ce que"
- interrogation par inversion

