

UNIVERSITE LOUIS PASTEUR  
I.R.E.M. de Strasbourg  
10, rue du Général Zimmer  
67084 STRASBOURG CEDEX

POUR UNE DESCRIPTION QUANTITATIVE  
DES CARACTERISTIQUES REDACTIONNELLES D'UN TEXTE :  
CONTRIBUTION A L'ETUDE DE LA LECTURE

Par Raymond DUVAL

AVRIL 1981

POUR UNE DESCRIPTION QUANTITATIVE  
DES CARACTERISTIQUES REDACTIONNELLES D'UN TEXTE :  
CONTRIBUTION A L'ETUDE DE LA LECTURE

Par Raymond DUVAL

Tous les textes ne sont pas également faciles à lire. Soit en raison du sujet traité et des connaissances présupposées, soit en raison de leur rédaction : choix du vocabulaire, complexité des phrases, ordre de présentation des informations, degré de redondance, etc... Ces différentes raisons concourent plus ou moins à la difficulté d'un texte pour une catégorie de lecteurs donnée.

Les formules de lisibilité et le test de closure sont des procédures pour tenter d'évaluer le degré de difficulté d'un texte en fonction du public auquel on le destine : les élèves d'un certain niveau scolaire ou adultes d'un niveau d'études déterminé. Ces procédures ne permettent pas cependant de savoir en quoi consistent les difficultés ainsi globalement estimées. On peut, à l'aide de ces procédures, ordonner plusieurs textes sur un même sujet en fonction des scores : chaque score est censé représenter soit un coût plus ou moins grand pour la lecture (estimé a priori par une formule de lisibilité) soit une performance de suppléance aux trous, significative du degré de compréhension du texte par un lecteur. Mais comme nous n'avons aucune description objective précise des différences existant entre ces textes, on ne peut savoir à quels obstacles la lecture peut se heurter, par exemple dans un texte X et qu'elle ne rencontrera pas dans le texte Y. Les relations d'ordre établies avec les formules de lisibilité et surtout avec le test de closure ne permettent aucune interprétation, puisque manque une analyse préalable des différences rédactionnelles. On peut seulement décider qu'au dessus d'un certain seuil, fixé pragmatiquement, le texte devient à la portée de tel public. On ne peut donc attendre du test de closure des informations éclairant la démarche de lecture, c'est-à-dire la

présence ou l'absence de traitements qu'on peut cependant supposer nécessaires à cette démarche.

Le but de ce travail est de définir des variables rédactionnelles qui puissent être contrôlées en termes de performance de lecture. Pour cela il faut considérer un texte, quel qu'il soit, comme la réalisation d'une rédaction parmi d'autres possibles : sans changer les informations données explicitement, il est possible de produire un autre 'texte qui ne soit pas équivalent du point de vue du coût de la lecture. Le passage d'une rédaction à une autre se traduit par une modification de la structure de surface des phrases et éventuellement de leur ordre. Or pour chaque rédaction on peut relever plusieurs proportions d'indices linguistiques choisis en fonction des traitements nécessaires à la compréhension du texte : quels sont les objets dont parle le texte, quelles relations établit-il entre eux, dans quel ordre ces relations sont elles présentées... ?

Pour chaque rédaction on peut obtenir ainsi un vecteur qui la caractérise.

Cette procédure qui permet de comparer entre elles plusieurs rédactions peut être étendue à des textes qui n'ont rien de commun quant à leur contenu : ce sont leurs seules caractéristiques rédactionnelles, les prises offertes au lecteur, qui se trouvent ainsi prises en compte. Certes une vérification de l'effet des différentes variables sur la performance de lecture relève d'une démarche expérimentale plus complexe lorsque les textes présentés n'ont pas le même contenu. Mais il importe d'abord de montrer qu'une évaluation quantitative des caractéristiques rédactionnelles peut constituer une analyse précise des textes et ouvrir la voie à des comparaisons objectives et contrôlables entre plusieurs textes. Ce qui est une étape nécessaire dans l'étude des démarches de lecture.

Qu'il y ait du point de vue rédactionnel, ou de la structure de surface, des différences importantes entre les textes, cela est aisé à mettre en évidence. Prenons par exemple les cinq textes donnés en annexe. On peut souligner tout de suite des différences, ne serait-ce qu'en relevant la fréquence des termes selon leur catégorie grammaticale.

	PROUST	ROGERS	PIAGET	DELEDIC	IREM eng.
Prépositions	9 %	19 %	22 %	8 %	6 %
Adverbes	29 %	12 %	11 %	8 %	14 %
Substantifs	14 %	13 %	22 %	23 %	30 %
Verbes	11 %	21 %	11 %	17 %	10 %
	63 %	65 %	66 %	56 %	60 %

Ainsi la catégorie des termes la plus fréquemment représentée varie selon le texte. Dans celui de Proust ce sont les adverbes, dans celui de Rogers les verbes, dans celui de Piaget les substantifs et les prépositions... On peut noter aussi que la réunion des deux catégories, substantifs et verbes, varie entre 25 % et 40 %, alors que la réunion des quatre catégories oscille autour de 62-63 %.

Bien qu'on observe la stabilité de ces relevés sur différents passages d'un même ouvrage (sauf les manuels de mathématiques), ces simples relevés ne sauraient être considérés comme des caractéristiques rédactionnelles : celles-ci doivent être définies en fonction des traitements indispensables pour une démarche de lecture.

Chacun de ces traitements peut être plus ou moins sollicité, plus ou moins coûteux, selon l'importance des caractéristiques rédactionnelles d'un texte.

I - LECTURE ET CARACTERISTIQUES REDACTIONNELLES D'UN TEXTE.

On peut estimer que le parcours visuel d'une suite de lignes est une lecture lorsque le sujet est en mesure de répondre à différents types de questions : quels sont les objets mis en scène par le texte ? Quel est leur ordre d'entrée ? Quelles relations sont nouées entre ces objets ? Dans quel univers sont-ils situés ?... Cela ne signifie pas que le comportement observable de lecture puisse être décrit comme la recherche d'une réponse à ces différentes questions : ce serait se limiter à la situation d'une lecture explicitement dirigée par de telles questions. Cela signifie seulement que toute performance de lecture peut être analysée en fonction des traitements nécessaires pour répondre éventuellement à telles questions. Mais dans une lecture non dirigée, ces traitements ne débouchent pas toujours sur l'élaboration de réponses. La rédaction d'un texte peut faciliter ou rendre plus difficiles les traitements intrinsèques à la lecture.

1° L'identification des objets mis en scène par le texte et celle des relations dont ils sont l'origine ou le terme.

La compréhension d'un texte fait appel à un certain nombre de connaissances ou d'informations qui sont sollicitées dans le texte sous forme d'objets dont on parle et de relations ou de transformations mises en oeuvre. Deux catégories grammaticales de mots mobilisent principalement ces connaissances de part leur fonction de renvoi extralinguistique à des objets ou à des relations : ce sont les substantifs et les verbes. Ils offrent à la lecture les indices privilégiés pour l'identification.

Celle-ci ne peut cependant être réduite à la seule reconnaissance des substantifs et des verbes. C'est souvent à partir d'une construction nominale de plusieurs substantifs, ou d'une construction verbale, que cette identification peut être effectuée. Il y a dans ce cas une identification préalable des groupements de termes, ou syntagmes, constituant une unité du point de vue de la désignation ou de la construction d'une relation. Cela nous conduit à définir deux caractéristiques rédactionnelles que nous appellerons faute de mieux : la densité lexico-sémantique et la complexité syntagmatique.

Ces deux caractéristiques rédactionnelles recouvrent le phénomène important de la nominalisation, c'est-à-dire de cette transformation généralisée qui convertit un couple de phrases en une phrase unique (Ruwet [11] p 208).

Ainsi la phrase "la construction de la maison progresse" peut être décomposée en deux phrases : "la maison est construite" et "cela progresse".

Cette conception transformationnelle conduit à relativiser la distinction entre substantif et verbe : peut-on mettre dans une même catégorie le nom "construction" qui correspond à un verbe en structure profonde et le nom "maison" qui est déjà un nom en structure profonde ? (Ducrot [4] p 269).

Il y a donc des "substantifs verbaux", des noms issus de verbes et signifiant un procès, une attitude ou un sentiment, et non point un objet (Ruwet [11] p214).

Dans ces conditions prendre comme indices les catégories de substantif et de verbe, est-ce un critère pertinent ?

Avant de répondre à cette question, on peut faire les remarques générales suivantes :

- la lecture est une activité de performance qui porte sur les suites terminales et non sur les structures profondes
- le travail du lecteur ne consiste pas essentiellement à remonter des structures de surface aux structures profondes, c'est-à-dire à faire le travail inverse de la production selon les règles transformationnelles. Si on a pu montrer que certaines transformations singulières (négation, forme passive) entraînaient un coût de traitement dans la vérification de phrases isolées, cela est encore loin d'être observé pour les transformations généralisées des phrases d'un texte.

Pour le lecteur un substantif verbal est, dans la phrase qu'il lit, un substantif c'est-à-dire un terme désignant un objet au sens large, un individu, une classe ou un processus. Et ce substantif verbal dans sa fonction de référence se distingue nettement du verbe, lequel fait de l'objet référé le sujet d'une relation ou d'une transformation. La familiarité ou la non familiarité des substantifs sont pour le lecteur, des différences plus importantes que la nature sémantique formelle du substantif.

L'identification de ce qui est désigné par une nominalisation représente un coût non réductible à la dé-nominalisation en deux ou plusieurs phrases. Ainsi ce qui est désigné par "la construction des nombres entiers (s'effectue chez l'enfant)" n'est pas simplement donné par les deux phrases : "les nombres entiers sont construits" et "cela s'effectue...". Il faudrait d'ailleurs distinguer :

- les nominalisations actives qui ont une fonction anaphorique dans le texte : elles renvoient à une phrase correspondante présente dans le texte avant ou après. "La police portugaise... n'est toujours pas assainie..... L'assainissement de la police...." S. Moirand en a relevé de nombreux exemples dans la presse écrite [(9) p 62...]  
Nous disons ici que la nominalisation est active parce qu'elle est effectivement effectuée dans le texte. Sa fonction est alors anaphorique.
- les nominalisation sédimentées, c'est-à-dire celles qui ne sont pas la reprise d'une phrase du texte. Leur fonction n'est plus anaphorique mais descriptive : elles désignent un objet auquel n'est pas habituellement attribué un terme du lexique ou parfois elles désignent un phénomène, une observation ou une théorie qui ne peuvent être décrites que par un texte. A ce deuxième type appartiennent des nominalisations comme "la construction des nombres entiers..." "la transmission des mouvements de rotation..." ou même "les souvenirs des songes...". Ce type de nominalisation renvoie à des représentations et non plus à des phrases constituantes.

On a souligné que "les nominalisations accroissent la densité d'un texte permettant de hiérarchiser les données de l'expérience et l'argumentation" (R. Lagasse cité par S. Moirand [9] p60 ). Les deux caractéristiques rédactionnelles de densité lexicale et de complexité syntagmatique permettent justement une évaluation de cette densité en prenant, comme on le verra plus loin, les termes qui dans le texte ont une fonction substantive.

## 2° L'organisation des informations au cours de la lecture : la dé-linéarisation du texte.

Comprendre un texte c'est élaborer une représentation de la situation qu'il décrit. Cette représentation résulte d'une réorganisation par le lecteur des informations présentées linéairement par les phrases successives.

Chaque phrase énonçant quelque chose de l'objet que désigne le syntagme nominal, le choix de l'objet sur lequel l'énoncé de la phrase est ancré, est un choix rédactionnel.

Ainsi on peut dire "la construction des nombres entiers n'est pas liée à l'activité du comptage" comme "l'activité du comptage n'entraîne pas la construction des nombres entiers". De ce choix rédactionnel résultent deux phénomènes pour le lecteur.

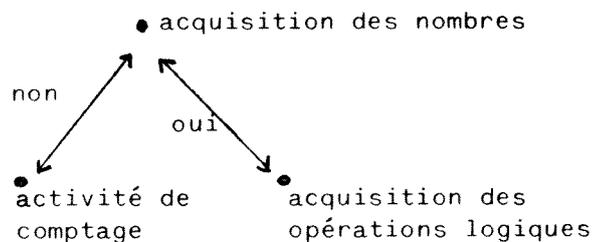
D'une phrase à l'autre, l'ancrage de l'énoncé change : cela oblige le lecteur à une recentration pour appréhender la situation décrite. Ainsi les trois premiers énoncés du texte de Piaget (voir en annexe) sont chacun centrés sur un objet différent : l'acquisition des nombres entiers ("la construction des nombres entiers"), le jeune enfant et l'activité de comptage ("l'évaluation numérique"). Cela donne la première rédaction

- 1) L'acquisition des nombres entiers est liée à l'acquisition des opérations logiques.
- 2) Un jeune enfant n'a pas acquis le nombre entier parce qu'il possède l'activité de comptage.
- 3) L'activité de comptage est d'abord liée à la disposition spatiale des éléments.

On aurait pu avoir une autre rédaction :

- 1) L'acquisition des nombres entiers est liée à l'acquisition des opérations logiques.
- 2) Cette acquisition n'est pas liée à l'activité de comptage.
- 3) L'activité de comptage est d'abord liée à la disposition spatiale des éléments.

Ces deux rédactions correspondent à la même représentation :



Mais dans la deuxième rédaction les deux premiers énoncés ont le même objet d'ancrage, ce qui est plus congru avec le schéma des relations qu'il faut établir entre les trois objets. La deuxième rédaction est plus centrée que la première ; et le second énoncé se prête, dans les deux rédactions, à l'emphase argumentative "Il ne faut pas croire en effet..." "Que cette acquisition soit liée à l'activité de comptage". Mais dans la rédaction de Piaget l'emphase argumentative introduit le second énoncé comme reprise d'une objection qu'un interlocuteur imaginaire opposerait au premier énoncé. D'où peut être la substitution de l'objet "jeune enfant" à l'objet "activité de comptage".

Le deuxième phénomène auquel le lecteur peut être confronté est la reprise du même objet d'ancrage au cours des énoncés successifs : il n'y a dans ce cas aucune décentration de perspective. Un traitement d'une autre nature est alors sollicité. La reprise du même objet d'ancrage ne se fait pas dans les mêmes termes : il peut y avoir pronominalisation, substitution lexicale, référentiation déictique contextuelle. Cette reprise ou répétition est d'ailleurs une condition nécessaire pour assurer la cohérence d'un texte (Charolles [1] p 14-17). A ce titre elle ne comporte aucune contrainte sur la place, dans la phrase, de l'élément répété : on peut concevoir un texte présentant "dans son développement linéaire des éléments à récurrence stricte" et dans lequel il n'y ait pas deux phrases dont les sujets grammaticaux co-réfèrent. Ce phénomène de répétition dans le texte donne lieu à des reports, à une interprétation des termes anaphoriques, d'une autre nature que le changement de perspective que nous évoquions plus haut.

Nous nous limiterons dans ce travail au seul phénomène de la constance ou de la variation des objets d'ancrage dans la suite des énoncés. Nous appellerons "décentration référentielle" cette caractéristique rédactionnelle.

### 3° L'inférence d'énoncés non donnés dans le texte.

Entre les phrases du texte, il doit exister un lien perceptible au lecteur pour que le discours ne paraisse pas incohérent. Les répétitions ne peuvent assurer à elles seules l'unité du texte. D'autres conditions sont aussi nécessaires : la progression sémantique et la dénotation des éléments du texte dans un même monde représentatif (Charolles [1] p 27-31). Cela revient à admettre que le lecteur doit être en mesure de saisir l'implicité d'un texte.

La progression sémantique se fait souvent au prix d'une discontinuité : des intermédiaires sont effacés. Par ailleurs la référence au même monde représentatif suppose l'explication d'un trait commun à deux énoncés ou à deux segments d'énoncés qui, pris isolément, peuvent ne présenter aucun rapport.

La saisie de cet implicite est souvent automatique et non remarquée lorsqu'il s'agit de connaissances pragmatiques communes à un milieu donné ou, plus largement, à une culture. Elle peut exiger au contraire arrêt et réflexion lorsqu'il s'agit de connaissances inconnues du lecteur ou, de relations intrinsèques entre les énoncés, telles les relations logiques, les relations argumentatives lesquelles peuvent s'étendre à des énoncés absents du texte mais postulés ou sous-entendus. La saisie de ces relations, et la saisie de l'unité du texte dans le passage discontinu d'une progression sémantique relèvent d'une activité d'inférence.

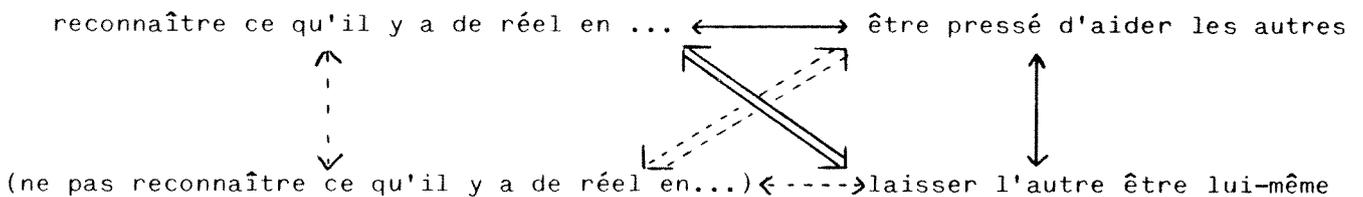
Prenons à titre d'exemple le passage de Rogers visant à décrire le changement qu'apporte une attitude d'écoute (voir annexe). Nous y trouvons la séquence des trois énoncés suivants.

- (1) **Plus** Je suis prêt à reconnaître ce qu'il y a de réel en moi et chez l'autre,  
A
- Moins** J'ai le désir d'essayer à tout prix d'arranger les choses  
B
- (2) **Plus** J'essaie de m'écouter et d'être attentif à mon expérience interne  
et plus j'essaie d'étendre cette attitude d'écoute à un autre  
A'
- Plus** J'éprouve de respect pour les complexités du processus vital  
C
- (3) **C'est pourquoi** Je me sens de moins en moins pressé d'arranger les choses  
B'

Nous avons là un raisonnement par transitivité  $A \rightarrow C \rightarrow B$  présenté de la façon suivante :

(parce que)  $\left\{ \begin{array}{l} \text{énoncé (1) : } A \rightarrow B \\ \text{énoncé (2) : } A' \rightarrow C \\ \text{énoncé (3) : } (C) \rightarrow B' \end{array} \right.$

En outre ce texte de Rogers met en jeu un tableau d'oppositions entre quatre prédicats dont trois seulement sont explicitement donnés.



Les flèches verticales indiquent l'incompatibilité sémantique de deux prédicats. Les flèches horizontales indiquent une liaison habituellement acceptée entre deux prédicats mais que Rogers rejette comme contraires. Les flèches diagonales indiquent les liaisons effectives. Dans le texte la liaison est signifiée par le couple "plus... plus..." et l'opposition par le couple "plus... moins...". Le texte de Rogers est donc construit sur un renversement des liaisons entre prédicats. Mais seule une liaison est affirmée, l'autre étant suggérée. Cette affirmation est justifiée par une transitivité non entièrement explicite.

La lecture de ce texte apparemment descriptif suppose une activité d'inférence. Présenté à des étudiants en même temps que le texte de Piaget et celui de Proust, il avait paru plus facile et des scores en moyenne plus élevés avaient été obtenus pour ce texte : 57 % contre 48 % pour ceux de Proust et de Piaget. Mais les erreurs s'étaient principalement concentrées sur les trous qui correspondaient à un terme clé pour la compréhension de l'argumentation.

Nous appellerons dépendance argumentative la caractéristique rédactionnelle qui sollicite du lecteur la saisie de l'implicite du texte.

Le travail de la lecture ne dépend pas seulement des connaissances du lecteur, il est aussi fonction de l'importance des caractéristiques rédactionnelles du texte. C'est à partir de leurs variations que l'on peut espérer analyser les traitements intermédiaires entre le parcours visuel des mots et la compréhension du texte, celle-ci étant déterminée par un niveau choisi de performance.

## II - DESCRIPTION QUANTITATIVE DES CARACTERISTIQUES REDACTIONNELLES.

L'identification des objets désignés, la réorganisation des informations, l'activité d'inférence conduisent à privilégier dans le texte des types d'indices linguistiques. Pour chacune des caractéristiques rédactionnelles le nombre de ces indices peut être relevé et rapporté à un ensemble d'occurrences. Mais il serait erroné de croire que cet ensemble puisse être principalement le nombre total des mots. Le traitement des indices linguistiques présents se fait en fonction de certains autres indices auxquels ils renvoient directement : un substantif est d'abord en relation avec un verbe ou avec un autre substantif, une préposition est en relation avec un substantif ou verbe... Cet ensemble peut être aussi la phrase ou les énoncés qui la composent.

Ce sont ces indices propres à chaque caractéristique rédactionnelle que nous allons maintenant présenter. Nous illustrerons cette présentation par la description de plusieurs textes (voir annexe). Les trois premiers, ceux de Proust, de Rogers, de Piaget, ont été proposés à des étudiants dans le cadre d'un test de closure. Le texte de Piaget, en particulier, avait été proposé quelques semaines après un cours sur le même sujet. Les trois autres sont destinés à des élèves de quatrième. L'un est une suite d'instructions pour simplifier une fraction (Deledicq), l'autre une description des engrenages en tant qu'illustration des nombres rationnels (IREM). Le troisième n'est pas à proprement parler un texte, mais un corpus de définitions.

### 1° La densité lexico-sémantique.

Deux types d'indices linguistiques sont privilégiés pour l'identification des objets désignés et pour celle des relations mises en oeuvre : ce sont les termes ayant une fonction substantive et ceux ayant une fonction de lien verbal. La liaison entre substantifs et verbes est la liaison fondamentale de la phrase, celle dans laquelle le sens s'articule : par elle est établie une relation entre plusieurs objets, ou bien un objet est posé comme sujet d'une transformation.

Or, en surface, cette liaison fondamentale entre substantif et verbe présente différentes formes qui peuvent être distinguées par les propositions suivantes :

- un substantif pour un verbe, comme dans les énoncés suivants :

"le vent se lève...", "la séance commence", "tracez une droite", "faire l'opération". Si des liaisons de ce type sont prédominantes dans le texte, on a alors une proportion de substantifs, par rapport aux verbes, oscillant autour de 1. Ainsi le texte de Deledicq, composé surtout d'instructions, a une proportion peu éloignée de 1 (voir le tableau p. 14). Cela n'exclut pas bien sûr la présence d'énoncés comme "le chat mange la soupe", mais de tels énoncés peuvent être équilibrés par d'autres comme "et il s'endort".

- Moins d'un substantif par verbe, comme dans "J'ai posé une question à laquelle tu n'as pas répondu". D'une façon plus générale, cela signifie que plusieurs prédicats ou plusieurs transformations sont rapportées au même objet, lequel est alors désigné par une pronominalisation. C'est le cas du texte de Rogers : les verbes y prédominent nettement sur les substantifs et beaucoup d'énoncés sont centrés sur "Je". Notons que les constructions relatives ne peuvent permettre d'obtenir une proportion inférieure à 1. De  $P_1$  "le musicien frappe le clown" et de  $P_2$  "le clown regarde la danseuse", on peut former "Le musicien frappe le clown qui regarde la danseuse" (exemple donné dans [10] p.144).

- Plusieurs substantifs pour un verbe. Plusieurs cas doivent être distingués :

- Le verbe relie deux substantifs comme "la souris échappe au chat".
- Les objets ne sont pas désignés par un substantif, mais par une construction nominale comme dans "l'intersection de deux ensembles est un ensemble", ou comme dans "la transmission des mouvements de rotation est effectuée par des roues dentées".
- Il y a des énumérations.

Dans tous ces cas la proportion prend une valeur moyenne qui soit approche de 2, soit dépasse nettement 2, comme par exemple dans le texte de l'IREM ou dans le corpus de définitions. Cependant pour interpréter les valeurs ainsi obtenues, il faut prendre en compte la complexité syntagmatique. Dans le texte de Piaget, la présence de constructions nominales explique en partie le taux de 1,93, plusieurs énoncés ayant un sujet neutre "il" ou "non". Le Tableau suivant indique les différentes possibilités résultant d'un croisement de deux caractéristiques rédactionnelles.

COMPLEXITE SYNTAGMATIQUE

	faible	forte
proche de 2	prédominance des énoncés dans lesquels deux substantifs sont reliés par un verbe.	en partie des énoncés reliant un verbe et un substantif. en partie des constructions nominales.
nettement plus élevée que 2	énumérations	surtout des constructions nominales

DENSITE LEXICO-  
SEMANTIQUE

D'un texte à l'autre il y a donc des variations qualitatives importantes qui correspondent à la variation du nombre des substantifs présents par rapport à celui des verbes. Mais comme les verbes ne peuvent être considérés comme un ensemble d'occurrences pour les substantifs, il est préférable de définir un rapport  $\frac{\text{substantifs}}{\text{substantifs \& verbes}}$ . L'échelle des repères, esquissée plus haut se trouve conservée. Si  $\frac{S}{S+V}$  est proche de  $\frac{1}{2}$ , cela signifie la prédominance de la liaison un substantif par verbe.

Si  $\frac{S}{S+V}$  est proche de  $\frac{2}{3}$  nous retrouvons la situation correspondant à la valeur 2 évoquée plus haut.

Cependant en définissant cette proportion nous avons négligé un aspect essentiel. Un texte n'est pas un simple corpus de phrases. Celles-ci jouent les unes par rapport aux autres, un rôle de contexte, et de l'une à l'autre, il y a la continuité d'un même thème : "cette continuité correspond précisément au fait que l'on dit plusieurs choses différentes d'un signifié qui se retrouve, lui, dans plusieurs phrases successives" (Le Ny p. 43). D'une phrase à l'autre certains substantifs ou certains verbes peuvent être repris de sorte que la compréhension d'une phrase y est fonction de celle des phrases précédentes ou de questions qui trouveront une réponse dans les phrases suivantes.

Il faut donc prendre en compte le nombre de répétitions des substantifs dans la mesure où une telle répétition rend le coût d'identification moins important. Ce phénomène est sensible au cours de la lecture d'un ouvrage : il se produit une familiarisation avec le vocabulaire de l'auteur qui permet souvent d'accélérer le rythme de la lecture. En revanche la répétition des verbes peut ne pas signifier nécessairement une diminution du coût d'identification. Car si on considère que les verbes déterminent chaque fois une relation entre les objets désignés par les substantifs, il apparaît que la relation est chaque fois à identifier.

Nous aurons donc en définitive la variable rédactionnelle suivante

$$\frac{S_1}{S_1 + V}$$

où  $S_1$  est le nombre de substantifs lexicalement différents et  $V$  le nombre de verbes présents dans le texte.

La comparaison avec la probabilité intermédiaire que nous avons établie  $\frac{S}{S + V}$  fournit une indication sur l'importance des répétitions.

	$\frac{S}{V}$	$\frac{S}{S + V}$	$\frac{S}{S_1}$	$\frac{V}{V_1}$	$\frac{S_1}{S_1 + V}$
PROUST	$\frac{20}{15} = 1,33$	0,57	1	1,25	0,57
ROGERS	$\frac{22}{35} = 0,63$	0,39	1,22	1,41	0,34
PIAGET	$\frac{27}{14} = 1,93$	0,66	1,77	1,11	0,62
DELEDICQ	$\frac{21}{16} = 1,31$	0,57	3,23	2,63	0,30
IREM engrenage	$\frac{65}{23} = 2,83$	0,74	2,17	2,08	0,57
IREM	$\frac{40}{12} = 3,33$	0,77	1,35	2	0,71

## 2° La complexité syntagmatique.

Un texte peut être composé de phrases courtes comme le texte de Deledicq, de phrases longues ou même très longues comme celui de Proust. Ces variations ne sont pas négligeables pour la lecture. Les formules de lisibilité ont pris cet aspect en compte par le rapport  $\frac{\text{nombre de mots}}{\text{nombre de phrases}}$ . Mais cette description quantitative est très pauvre : la longueur moyenne d'une phrase ne peut avoir de signification que relativement à l'empan mnésique : très longue, la phrase en son début risque d'être oubliée avant qu'on ne soit parvenu à son terme.

Cela constitue un obstacle, mais plus à l'audition qu'à la lecture proprement dite. En outre, du point de vue mnésique, il est difficile de fixer un seuil entre ce qui serait une phrase longue et ce qui serait une phrase courte.

Tout dépend du nombre d'unités que le lecteur ou l'auditeur y distinguent, c'est-à-dire du nombre de regroupements effectués dans la suite des mots.

Dans un texte deux types de constructions peuvent contribuer à la longueur moyenne des phrases et surtout exiger un temps de fixation plus ou moins long pour les identifier : ce sont les descriptions et les propositions subordonnées. Que l'on compare de ce point de vue le texte de l'IREM sur les engrenages et celui de Proust.

a) Les descriptions, définies ou indéfinies, sont des expressions référentielles suppléant à la dénomination d'un terme dans la dénomination d'un objet : "le directeur du secrétariat général des Eaux et Forêts", "la transmission des mouvements de rotation". Russel a souligné le piège de telles descriptions qui peuvent poser implicitement l'existence d'objets qui n'existent pas. Mais moins que le problème logique de la décision d'existence, ce qui importe dans la lecture est le coût nécessaire pour identifier l'objet auquel réfère ou semble référer une telle description (Searle [13] P.64, 130-132).

Ces descriptions ont le plus souvent la forme de construction nominales, reliant deux substantifs, ou davantage, par une préposition. Pour les dénombrer il suffit de relever les couples substantifs-prépositions, par exemple " relation de..... " indépendamment du second substantif relié. Dans des constructions nominales plus longues nous pouvons avoir soit un couple de couples, "[transmission des [mouvements de....]]", soit un triplet de couples "[le point de [concours de deux [bissectrices d'un triangle]]]". Une descriptions fine exigerait que l'on tienne compte non seulement du nombre de couples, mais aussi du nombre d'enchâssements. Nous nous en tiendrons ici au seul dénombrement des couples avec les règles suivantes

- les couples pronoms-démonstratif-préposition sont équivalents à un couple substantif-préposition.
- la contraction d'une préposition et d'un article est comptée évidemment comme une préposition.

Dans le texte de Piaget l'expression "celle des inclusions" c'est-à-dire "la construction des ...", sera donc considérée comme un indice de complexité syntagmatique. Mais une expression comme "dans tous les cas..." ne pourra être considérée comme construction nominale.

Nous ne prenons pas en compte les couples substantif-adjectif, équivalents parfois à une construction nominale. Car du point de vue du coût d'identification des unités syntagmatiques, on ne peut considérer comme équivalentes les deux expressions "la construction des ensembles de nombres" et "la construction des ensembles numériques".

Le nombre des couples substantifs-préposition, constituant, à l'intérieur du texte, une expression référentielle peut être rapporté au nombre des mots du texte. De la sorte on peut définir la probabilité de rencontre d'un mot qui appartienne au noyau d'une construction nominale

$$\frac{C(s.p) \times 2}{\text{nombre de mots du texte}}$$

Mais naturellement les constructions nominales comportent plus que deux termes : "le souvenir des songes", "la construction des nombres entiers". Une probabilité de 0,16 sera donc une probabilité élevée : elle signifie que près d'un tiers des mots du texte est élément constituant d'une construction nominale. C'est le cas par exemple du texte de Piaget. Et une probabilité de 0,25 représente un seuil proche du maximum. Ce seuil semble presque atteint dans le corpus de théorèmes donné en annexe.

	couples [subor-prépo] [pron-prépo]	couples [prépo-verb] [verb-prép]	[prép-pron]	[pré-sub]	[adj-prép]	$\frac{C \text{ s.p} \times 2}{\text{nbre de mots}}$
PROUST	4	4		4	1	$\frac{4 \times 2}{140} = 0,06$
ROGERS	5	19		3	3	$\frac{5 \times 2}{170} = 0,06$
PIAGET	10	11		3	2	$\frac{10 \times 2}{125} = 0,16$
DELEDICQ		3		2	2	0 0
IREM engre.	16	10		3		$\frac{16 \times 2}{216} = 0,15$
IREM	20	7		3	2	$\frac{20,2}{159} = 0,25$

b) Outre les constructions nominales, il y a les constructions qui subordonnent une proposition à une autre. De ce type sont les propositions complétives, les propositions circonstancielles, les concessives, les comparatives. Mais nous excluons les propositions relatives [même lorsqu'il y a un double enchâssement], les propositions conditionnelles introduites par "si", et celles introduites par des termes pouvant être considérés comme homologues d'un connecteur logique. On peut donc dénombrer dans le texte le nombre de verbes introduits par "que" ou par un adverbe le plus souvent construit avec "que" ("bien que", "après que") et référer ces occurrences à l'ensemble des verbes du texte présent sous un mode personnel. Nous avons alors une estimation du degré de construction propositionnelle caractéristique d'un texte.

Nombre de verbes introduits par un adverbe ou par "que"  
Nombre de verbes autres que les verbes à l'infinitif ou au participe

Proust	Rogers	Piaget	Deledicq	IREM engre.	IREM corpus
$\frac{7}{10} = 0,7$	$\frac{1}{16} = 0,06$	$\frac{4}{10} = 0,4$	$\frac{1}{11} = 0,09$	$\frac{0}{15} = 0$	$\frac{0}{11} = 0$

Les constructions nominales et les constructions propositionnelles contribuent à la complexité syntagmatique d'un texte. Elles semblent cependant devoir jouer de façon indépendante : un texte peut présenter beaucoup de constructions nominales avec peu ou pas de constructions propositionnelles (le texte IREM) et réciproquement (le texte de Piaget). Il serait donc illusoire à ce stade de chercher une valeur unique pour caractériser la complexité syntagmatique.

### 3° La complexité du réseau référentiel.

Nous avons vu que chaque énoncé est ancré à un objet nommé en position de sujet grammatical. Relativement à la réorganisation des informations présentées séquentiellement, la variation ou la constance des objets d'ancrages au fil du texte n'est pas une caractéristique neutre. Celle-ci peut être aisément décrite par le rapport entre le nombre des objets d'ancrages et le nombre des sujets grammaticaux  $\frac{Oa}{Sg}$ . Si ce rapport est 1 ou s'approche de 1 cela signifie que l'objet change à chaque énoncé.

Ainsi dans l'énoncé du problème suivant, il y a trois sujets grammaticaux et trois objets différents : "Deux pompes concourent à l'épuisement d'un bassin. La première enlève 4,5 l d'eau par coup de piston. La seconde en enlève le double..."

D'un point de vue pratique les expressions "il faut", "on demande" soulèvent un problème. Nous avons décidé de compter l'occurrence d'un sujet grammatical avec une place vide. Ce type de description appelle deux précisions. D'une part elle va à l'encontre de certaines approches qui recourent à la logique des prédicats et qui éliminent la notion de sujet en la décomposant dans le couple prédicat-variable. Pour reprendre l'exemple classique, "le roi de France est chauve" s'analyse en "il existe x tel que x est roi de France et x est chauve". Dans cette perspective la phrase "En 1941, le 7 décembre, l'aviation japonaise attaqua violement la flotte américaine mouillée dans le port de Pearl Harbour à Hawaï et lui infligea des pertes irréparables" se décomposerait en 14 prédicats avec argument (Le Ny [8] p 62-63) une telle approche qui peut avoir un sens pour des phrases prises isolément tend à éliminer ce qui fait la spécificité d'un texte par rapport à une phrase : description d'une situation par une suite de points de vue qui auraient pu être pris autrement.

Au niveau du texte la notion de sujet grammatical garde son importance : le substantif sujet fixe le cadre à l'intérieur duquel joue l'assertion, la question ou la quantification de l'énoncé (Ducrot [3] pp 234-235). Ce cadre peut changer d'une phrase à l'autre.

Cette description, d'autre part, n'est pas une description de la cohérence du texte au sens où on l'entend ordinairement. La reprise d'un petit nombre de signifiés (nous avons préféré parler d'"objets") dans la séquence des énoncés est une des conditions nécessaires pour qu'il y ait vraiment un texte et non pas une collection de phrases indépendantes (Le Ny [8] p.73, Charolles [1] p. 14-15). Mais cette condition n'implique nullement que ces signifiés soient repris dans la fonction de sujet grammatical. La continuité thématique essentielle pourqu'il y ait texte est autre chose que la centration ou la centration des énoncés d'un texte.

	Proust	Rogers	Piaget	Deledicq	I.R.E.M. eng.
$\frac{Oa}{Sg}$	$\frac{7}{10} = 0,7$	$\frac{6}{16} = 0,37$	$\frac{5}{10} = 0,5$	$\frac{3}{8} = 0,37$	$\frac{9}{14} = 0,64$

Une toute autre procédure pourrait être envisagée :

- relever le nombre d'objets différents auxquels le texte réfère. N O
- relever le nombre de verbes lexicalement différents, c'est-à-dire le nombre de relations différentes utilisées dans le texte N V
- faire le produit pour établir l'ensemble des énoncés possibles sur la base de ce qui est explicitement donné dans le texte
- relever le nombre de couple O-V effectivement constitués dans le texte

On obtient alors un rapport  $\frac{N(O-V)}{NO.NV}$  indiquant un degré de choix pour l'occurrence de chaque énoncé.

Une telle procédure pour être acceptable exige cependant les deux restrictions suivantes :

- 1) certains couples (O-V) sont impossibles en vertu de contraintes sémantiques il faut donc les exclure. D'où on a

$$\frac{N(O-V)}{NO.NV - N(O-V \text{ impossible})}$$

- 2) Ce degré de choix n'a de sens que pour le lecteur et non point pour le rédacteur. Pour ce dernier un ensemble des énoncés possibles n'est pas définissable.

Outre l'identification des objets auxquels sont ancrés les énoncés du texte, il y a un autre type de référence qui se révèle important au moins dans les textes mathématiques : ce sont les renvois à "l'activité du sujet qui tient le discours" [M.Darot p42,10]. Dans les textes d'enseignement mathématique, ces renvois sont marqués par des verbes : poser, supposer, définir, conclure... Ces verbes indiquent le statut d'un énoncé dans une démarche de démonstration ou d'exposition : ils signalent l'attitude intellectuelle qui permet de saisir la fonction d'un énoncé dans le discours. C'est une hypothèse, c'est un exemple, c'est un rappel... A cette classe de verbes, on pourrait rattacher les verbes indiquant ce que Russel appelait une "attitude propositionnelle" (croire, désirer, douter) dans la mesure où ils sont suivis d'une proposition subordonnée (Russel p 77, 182).

Ces différentes marques verbales de l'activité du locuteur transforment la nature des énoncés qui les intègrent. Elles relèvent d'un second niveau de discours dans lequel les énoncés sont neutralisés pour devenir eux-mêmes objets d'énoncés. Le nombre de ces différentes marques verbales est à rapporter

à l'ensemble des verbes du texte se présentant sous un mode personnel. Dans les cinq textes donnés en annexe, la fréquence de ces marques est presque nulle.

4° La dépendance argumentative des énoncés.

Les caractéristiques décrites jusqu'à présent privilégient la liaison fondamentale dans laquelle s'articule le sens d'un énoncé. Le problème du passage d'un énoncé à un autre n'a été envisagé qu'en fonction des objets d'ancrage choisis pour établir chacune des liaisons fondamentales posées dans le texte. Or le sens d'un texte n'est pas seulement dans la cohérence référentielle ou pragmatique des différents énoncés, il est aussi dans la connection établie entre plusieurs énoncés, connection qui les rend indissociables.

Le sens d'un texte ne se joue pas seulement dans le sens de chacun des énoncés, il se joue aussi dans l'interdépendance des énoncés, laquelle ne se limite pas à ce qui est donné explicitement dans le texte, comme nous avons pu le voir à propos du texte de Rogers (p.9-10).

Nous sommes ici en présence d'une nouvelle dimension, plus difficile à traduire quantitativement que les précédentes. Il y a en effet des termes qui marquent sans équivoque la connection de deux énoncés. Ce sont les "termes logiques", ceux qui relient des propositions ou des expressions pouvant être considérées comme des propositions implicites : si, non, et, ou (Russel [12] p. 77). Ces termes sont considérés comme les homologues, dans la langue courante, des opérateurs formels (Grice [5] p. 57) : leur intérêt est de permettre l'inférence. On peut remarquer cependant que l'informativité de la conjonction "et" est souvent nulle : la linéarité de l'énoncé conduit le lecteur à réunir les termes de même catégorie qui se suivent, qu'ils soient ou non reliés par "et". L'emploi de cette conjonction signale plus souvent la clôture d'une énumération de termes ou d'expressions qu'il ne signifie leur réunion. Tous les "termes logiques" ne sont pas indices d'une liaison argumentative. En revanche, il y a d'autres termes dont la valeur argumentative est indéniable : "mais", "sans"...

Une approche qui se voudrait exhaustive supposerait une analyse systématique de toutes les relations qui peuvent être établies entre deux énoncés. Nous en signalerons seulement quelques unes, en indiquant les termes qui la signifient :

- négation : "ne...pas", "ne...", "aucun" ...
- restriction : "ne...que", "sans", "seulement" ...
- inversion : "au contraire", "mais", "plus...moins" de sens
- condition : "si".

Tous ces termes marquant une relation entre un énoncé du texte et un autre énoncé donné explicitement ou non dans le texte doivent être considérés comme des indices de dépendance argumentative.

Il reste un autre point délicat : comment définir l'ensemble de référence pour déterminer le degré d'interdépendance des énoncés dans un texte ?

Le seul ensemble possible semble devoir être celui des phrases (définies par la ponctuation ";" ou "."). Si on prenait le nombre des propositions, on serait gêné par les éventualités suivantes : certains termes "mais", "sans" peuvent introduire une expression non verbale et établir, cependant, l'opposition d'un énoncé explicite à un énoncé implicite. Dans un énoncé comme "Je crois qu'il fera beau", il faudrait compter deux propositions, alors que manifestement cet énoncé constitue un tout indécomposable pour entrer dans une unité argumentative plus large "Je crois qu'il fera beau  $\left. \begin{array}{l} \text{\{donc\}} \\ \text{\{mais\}} \end{array} \right\} \dots"$

En d'autres termes il faut éviter de confondre le degré de complexité propositionnelle des phrases et le degré de dépendance argumentative.

PROUST	ROGERS	PIAGET	DELEDICQ	IREM engre.
$\frac{5}{2} = 2,5$	$\frac{5}{7} = 0,7$	$\frac{3}{6} = 0,5$	$\frac{1}{8} = 0,12$	$\frac{1}{10} = 0,1$

La lecture de ce tableau peut surprendre. La dépendance argumentative du texte IREM apparaît nulle et celle du texte de Proust apparaît élevée.

Le texte de l'IREM est un inventaire des procédés de transmissions pouvant être représenté dans le tableau suivant :

3 procédés de transmissions de rotations	sens respectifs des mouvements de rotation	rapport des roues
- roues dentées - chaînes - courroies	- inverse - le même - le même (signe positif) - inverse (signe négatif)	$\frac{\text{Circonf. de la roue menante}}{\text{circonf. de la roue menée}}$

Le texte de l'I.R.E.M. veut décrire une situation qui se prête mal à une représentation linéaire dans un texte. D'où une difficulté artificielle. L'ordre des phrases peut d'ailleurs y être changé sans en affecter en rien le sens ou l'intelligibilité du texte.

Le texte de Proust est au contraire un labyrinthe d'énoncés. Nous pouvons en schématiser les bifurcations de la façon suivante :

Phase 1

Les souvenirs des songes pendent, [incise] ↓ , aux parois obscures d'une chambre { qui s'ouvre...  
 où travaille... duquel → est  
 → et (est)

**MAIS** ils sont si enténébrés qu'on les aperçoit **seulement** ..... quand

Phase 2

Quelques uns (étaient) déjà clairs

**MAIS** ils sont devenus si méconnaissables que, [incise négative] ↓ , ...ainsi que →  
 → ...si...que... **ne** pourrait

Durant le parcours de ce labyrinthe un glissement s'opère avec les adjectifs : obscurs, enténébrés, clairs, méconnaissables, décomposés.

L'atmosphère s'impose de songes irrémédiablement oubliés et que le souvenir ne peut plus ressusciter à la clarté de la conscience. L'interdépendance argumentative se mêle à la complexité proportionnelle des deux phrases : les restrictions et les négations jouent un rôle essentiel dans cette évocation.

Nous avons déjà analysé dans la première partie le texte de Rogers. L'écart des valeurs obtenues pour le texte de Rogers et pour celui de Proust peut apparaître excessif. Cela tient essentiellement à la longueur exceptionnelle des phrases chez Proust. Cependant si pour Proust nous avions pris un passage un peu plus long, le nombre des phrases étant augmentées, nous aurions une valeur de dépendance argumentative moindre. Pour les autres textes nous avons des textes allant de 6 à 10 phrases. Cela paraît être une précaution indispensable pour évaluer la dépendance argumentative d'un texte.

### III - LE PARALLÉLISME LEXICO-SYNTAXIQUE.

Par cette expression nous désignons les répétitions dans le texte, de phrases, de propositions ou de syntagmes nominaux, à une variation ou deux près. Ces variations substituent à un ou deux termes des expressions synonymes ou antithétiques. Ce phénomène de parallélisme, caractéristique de la poésie et des formes orales de discours, [Guiraud] se retrouve dans des textes qui sont loin d'avoir une prétention littéraire. Par exemple les manuels de mathématiques.

" La [suite des nombres de la colonne de droite] est proportionnelle à la [suite des nombres de la colonne de gauche] " IREM 6e.

" Soient [A un axe de support d]  
et [A<sup>0</sup> un axe de support d<sup>0</sup>] " Monge 4e.

D'une façon générale, l'énumération ou l'examen de différents cas est une situation qui favorise le recours au parallélisme lexico-syntaxique [IREM 5e P.25...28 les textes sur fond rouge].

Dans le texte de Rogers ce phénomène de parallélisme n'est pas absent : "plus j'essaie de m'écouter et d'être attentif à mon expérience interne et moins j'essaie d'étendre cette attitude d'écoute à l'autre".

La présence ou l'absence de ce phénomène dans un texte est une caractéristique qui mérite d'être relevée, surtout lorsque le texte doit être lu dans une situation de closure. Dans cette situation, un mot sur cinq étant supprimé, il est relativement aisé de retrouver le mot effacé, si ce trou est dans un syntagme qui a son parallèle avant ou après. Car il y a plus de chances pour que ce soit l'un des termes répétés et non l'un des termes changés qui se trouve effacé selon la règle d'effacement des termes. Le phénomène de parallélisme lexico-syntaxique facilite donc une tâche de closure. En revanche, il peut comporter certains pièges dans une tâche de lecture normale, dans la mesure où l'identification d'une répétition peut masquer la variation locale. Les effets peuvent être différents selon des tâches proposées : facilitateurs dans un cas, distracteurs dans l'autre.

La fréquence des répétitions créant un parallélisme lexico-syntaxique est à rapporter au nombre des phrases du texte. Car le plus souvent c'est entre deux phrases consécutives ou entre les deux membres d'une même phrase que ce phénomène se produit.

CONCLUSION

DESCRIPTION QUANTITATIVE D'UN TEXTE ET ANALYSE DE LA LECTURE

	Densité lexico-sémantique		Construction nominale	Construction proposition	Décentration référentielle	Dépendance argumentative	Parallélisme
	$\frac{S}{S+V}$	$\frac{S_1}{S_1+V}$	$\frac{Cs.px2}{Mots}$	$\frac{Vadv}{Verbes}$	$\frac{Oa}{Sg}$	$\frac{\text{termes logi.}}{\text{phrases}}$	$\frac{N \text{ exp.}}{\text{phrases}}$
PROUST	0,57	0,57	0,06	0,70	0,70	2,5	$\frac{0}{2} = 0$
ROGERS	0,39	0,34	0,06	0,06	0,37	0,70	$\frac{1}{7} = 0,14$
PIAGET	0,66	0,62	0,16	0,40	0,50	0,50	$\frac{0}{6} = 0$
DELEDICO	0,57	0,30	0	0,09	0,37	0,12	$\frac{2}{8} = 0,25$
IREM engre.	0,74	0,57	0,15	0	0,64	0,10	$\frac{6}{10} = 0,60$
IREM Corpus	0,77	0,71	0,25	0	/	/	

Tableau Récapitulatif

- Rappels :
- on aurait en premières colonnes un agrandissement spectaculaire des écarts entre textes, si on prenait le rapport  $\frac{S}{V}$  au lieu de  $\frac{S}{S+V}$
  - la valeur 0,25 en 2e colonne pour  $\frac{Cs.px2}{mots}$  constitue une valeur approchant la saturation d'un texte en construction nominale.
  - la valeur très élevée de dépendance argumentative pour le texte de Proust tient en partie au trop petit nombre de phrases dans le texte.

Ce tableau récapitulatif nous donne la description quantitative des caractéristiques rédactionnelles de cinq textes. Il est très facile d'y faire les observations suivantes.

Dans le texte de Proust il y a peu de constructions nominales et la densité lexico-sémantique dépend surtout de la variété des substantifs employés : aucun n'est répété. S'il y a une difficulté de lecture dans le texte, elle tient essentiellement à la construction propositionnelle et à la dépendance argumentative qui ont des valeurs élevées.

A l'inverse le texte de Piaget présente une proportion importante de constructions nominales, la densité lexico-sémantique y est plus élevée. Mais la dépendance argumentative et la construction propositionnelle y sont moindres.

Le texte IREM engrenage présente deux caractéristiques semblables à celles du texte de Piaget (0,57 et 0,15 contre 0,62 et 0,16). En revanche, la construction propositionnelle et la dépendance argumentative y sont nulles. C'est un texte descriptif, comme celui de Deledicq lequel a aussi pour ces deux caractéristiques des valeurs presque nulles. Le texte de Deledicq présente en outre une faible densité lexico-sémantique en raison de la répétition des substantifs.

Le texte de Rogers présente lui aussi une faible densité lexico-sémantique mais en raison de la prédominance des pronoms. Le seul traitement un peu important qui semble devoir exiger sa lecture porte sur la dépendance argumentative des énoncés, bien que celle-ci y soit moindre que dans le texte de Proust. Le texte de Rogers présente en outre cette particularité que les énoncés sont pour la plupart centrés sur le même objet, à l'opposé de ce qui se passe dans le texte de Proust.

Par comparaison avec ces cinq textes on voit ce qu'est le corpus de définitions : un montage de substantifs.

+

+            +

Au vu de ces descriptions, les textes dont la lecture doit solliciter le plus de traitements sont ceux de Proust et de Piaget. Mais l'intérêt d'une description quantitative des variables rédactionnelles n'est pas dans ce type d'observations triviales. Il est dans la possibilité de comparer soit les variations rédactionnelles possibles à partir d'un texte donné, soit les caractéristiques de deux textes différents, c'est-à-dire de deux textes n'ayant pas le même thème. Et une comparaison contrôlée des différences rédactionnelles doit permettre d'analyser l'activité de lecture comme un système de traitements plus ou moins sollicités selon les caractéristiques du texte.

Dans la première partie de ce travail nous avons avancé des hypothèses sur la nature de certains traitements. Nous ne pouvons prétendre qu'ils suffisent pour décrire l'activité de lecture, ni que les caractéristiques retenues soient les seules qui leur correspondent. Ainsi nous n'avons pas pris en compte la présence des adjectifs : ceux-ci peuvent pourtant jouer un rôle dans l'identification des objets dont le texte parle. Nous n'avons pas davantage retenu les anaphores, pronominalisations ou reprises lexicales synonymes d'un même objet. La description quantitative de cette caractéristique ne soulève pas de difficulté majeure, mais son interprétation en termes de traitement est moins aisée. Le propos de ce travail était de montrer la possibilité d'une description quantitative des textes pour l'analyse de la lecture. Car on ne peut espérer analyser la lecture sans fixer deux points de repère : les connaissances du lecteur et les caractéristiques du texte. Si, en général on a des informations globales concernant le premier repère, ne serait-ce que par l'indication du niveau scolaire, on néglige le second repère.

+

+ +

D'autres problèmes, plus décisifs, restent ouverts. Tout d'abord celui de la réécriture d'un texte de façon à modifier une caractéristique rédactionnelle sans faire varier sensiblement les autres. On pourrait ainsi évaluer le coût pour des lecteurs, des constructions nominales dans le texte de Piaget, des constructions propositionnelles dans le texte de Proust, ou de la dépendance argumentative dans le texte de Rogers. Il semble qu'une réécriture ne modifiant pas le contenu explicite

du texte, c'est-à-dire une opération analogue à celle d'une traduction, exige le passage par des représentations non linéaires, lesquelles peuvent alors être décrites d'une autre manière. Nous avons effleuré ce problème dans la première partie à propos du texte de Piaget. C'est un peu un détour analogue à celui qui s'impose actuellement pour élaborer des programmes de traduction automatique ou de compréhension de textes par ordinateur.

Ce problème de réécriture soulève celui de l'indépendance des différentes caractéristiques prises en compte pour établir la description des textes. Cela est important ne serait-ce que pour établir la possibilité d'une réécriture qui modifie une caractéristique et qui n'affecte pas les autres. Si l'indépendance des constructions nominales par rapport aux constructions propositionnelles ne fait aucun doute, il n'en est pas de même pour la densité lexico-sémantique par rapport à la construction nominale ; car ces deux caractéristiques prennent en compte la présence de substantifs. La simple observation du tableau récapitulatif permet de supposer qu'une valeur élevée de construction nominale implique une forte densité lexico-sémantique mais que la réciproque est fautive.

Le dernier problème est d'un autre type : comment vérifier que la variation d'une ou plusieurs caractéristiques rédactionnelles correspond à une variation des performances de lecture ? Rien n'est plus délicat que l'évaluation de la compréhension d'un texte. Le test de closure est jusqu'à présent considéré comme un instrument fiable pour cette évaluation. Utilisé après analyse des caractéristiques rédactionnelles, ce test pourrait servir non plus à des fins de diagnostic pédagogique mais de vérification expérimentale d'hypothèses concernant la lecture. Mais le test de closure n'est pas la seule procédure envisageable. Car on peut recourir à d'autres variables dépendantes que le taux de reproduction des termes effacés du texte : le nombre de lectures, ou de recours au texte, nécessaire pour répondre à des questions difficiles portant par exemple sur la liaison des énoncés. Ou encore le temps de réaction s'il s'agit de questions faciles concernant l'identification des objets... etc.

Une analyse de l'activité de la lecture en un système de traitements n'a pas seulement un intérêt théorique. Elle ouvre un champ de questions didactiques : pour une catégorie de lecteurs, prise selon le niveau scolaire et selon le niveau socioculturel, à partir de quel seuil l'augmentation de certains coûts de traitement, crée-t-elle une limite à la compréhension d'un texte ?

B I B L I O G R A P H I E

- [ 1 ] CHAROLLES.M. Introduction aux problèmes de la cohérence des textes dans  
Langue Française N° 38 - Mai 1978.
  
- [ 2 ] DAROT. M. Discours mathématique et discours didactique.  
Bureau pour l'enseignement de la langue et de la civilisation  
française à l'étranger Paris 1975.
  
- [ 3 ] DUCROT. O. Dire et ne pas dire. Paris Herrmann - 1972.
  
- [ 4 ] DUCROT & TODOROV. Dictionnaire encyclopédique des sciences du langage.  
Paris, Seuil - 1972.
  
- [ 5 ] GRICE. Logique et conversation. Article traduit dans Communication n°30 1979.
  
- [ 6 ] HENRY. Comment mesurer la lisibilité. Paris, labo. Nathan - 1975.
  
- [ 7 ] LANDSHEERE. Le test de closure. Paris, Nathan - 1973.
  
- [ 8 ] LE NY. La sémantique psychologique. Paris, P.U.F. - 1979.
  
- [ 9 ] MOIRAND S. Le rôle anaphorique de la nominalisation dans la presse écrite.  
Dans Langue Française N° 28.
  
- [ 10 ] NOIZET. De la perception à la compréhension du langage. Paris, P.U.F. -1980.
  
- [ 11 ] RUWET N. Introduction à la grammaire générative. Paris, Plon - 1967.
  
- [ 12 ] RUSSEL Signification et vérité. Paris, Flammarion - 1969.
  
- [ 13 ] SEARLE Les actes de langage. Paris, Herrmann - 1972.

Annexe I

PROUST - Souvenir de songes. (Du côté de guermantes)

Aux parois obscures de cette chambre qui s'ouvre sur les rêves, et où travaille sans cesse cet oubli des chagrins amoureux duquel est parfois interrompue et défilée par un cauchemar plein de réminiscences la tâche vite recommencée, pendant, même après qu'on est éveillé, les souvenirs des songes, mais si enténébrés que souvent nous ne les apercevons pour la première fois qu'en plein après-midi, quand le rayon d'une idée similaire vient subitement les frapper; quelques uns, déjà harmonieusement clairs pendant qu'on dormait, mais devenus si méconnaissables que, ne les ayant pas reconnus, nous ne pouvons que nous hâter de les rendre à la terre, ainsi que des morts trop vite décomposés ou que des objets si gravement atteints et près de la poussière que le restaurateur le plus habile ne pourrait leur rendre forme. --

C.R. ROGERS - Ecoute et respect. (Le développement de la personne p.19-20)

La découverte suivante est difficile à communiquer. Elle consiste en ceci: *Plus je suis prêt à reconnaître ce qu'il y a de réel en moi et chez l'autre, moins j'ai le désir d'essayer à tout prix d'arranger les choses. Plus j'essaie de m'écouter et d'être attentif à mon expérience interne et plus j'essaie d'étendre cette attitude d'écoute à un autre, plus j'éprouve*

*de respect pour les complexités du processus vital. C'est pourquoi je me sens de moins en moins pressé d'arranger les choses, d'imposer des buts, de façonner des individus, de les manipuler et de les pousser dans les voies que je voudrais leur voir suivre. Je suis beaucoup plus satisfait d'être simplement moi-même et de laisser l'autre être lui-même. J'ai pleinement conscience que ceci doit sembler un point de vue étrange et quasi oriental. A quoi sert la vie si nous ne cherchons pas à agir sur les autres?*

PIAGET - (La psychologie de l'enfant. Que sais-je p. 82)

**5. Le nombre.** — La construction des nombres entiers s'effectue chez l'enfant en liaison étroite avec celle des sériations et des inclusions de classes. Il ne faut pas croire, en effet, qu'un jeune enfant possède le nombre du seul fait qu'il a appris à compter verbalement : l'évaluation numérique est en réalité longtemps liée pour lui à la disposition spatiale des éléments, en analogie étroite avec les « collections figurales » (voir plus haut sous 4). L'expérience décrite au chap. III, § IV-5 le montre à l'évidence : il suffit d'espacer les éléments de l'une de deux rangées mises initialement en correspondance optique pour que le sujet cesse d'admettre leur équivalence numérique. Or, on ne saurait naturellement parler de nombres opératoires avant que se soit constituée une conservation des ensembles numériques indépendamment des arrangements spatiaux.

Annexe I (suite)

DELEDICQ - "Faire" des Mathématiques. P. 29

*Fractions irréductibles.*

Parmi toutes les fractions représentant le même rationnel, on préfère souvent écrire celle ayant pour dénominateur le plus petit naturel possible.

En mathématique, on dit :

Une fraction irréductible est une fraction dont le dénominateur et le numérateur n'ont aucun diviseur commun. « Simplifier » une fraction, c'est trouver la fraction irréductible représentant le même rationnel.

Pour simplifier une fraction :

On cherche un diviseur commun au numérateur et au dénominateur, on effectue cette division, on cherche un autre diviseur... et ainsi de suite jusqu'à ce qu'on ne trouve plus de diviseur commun.

I.R.E.M. - "Mathématiques 4e, p. 102-103.

La transmission des mouvements de rotation est effectuée par des roues dentées (exemple : boîte de vitesses d'une automobile), par des chaînes (exemple : bicyclette) ou par des courroies (exemple : machines-outils de menuiserie).

Lors d'une transmission par roues dentées, deux roues qui se touchent tournent toujours en sens inverse. Lors d'une transmission par chaîne, au contraire, les sens de rotation de deux roues reliées sont les mêmes. Les transmissions par courroies offrent le choix : courroie droite pour conserver le sens de rotation, courroie croisée pour inverser les sens de rotation (voir figure). Dans tous les cas, la roue qui transmet le mouvement est appelée roue menante, et la roue qui reçoit le mouvement (qui est entraînée) est appelée roue menée.

Ci-dessus, lorsque la roue menante fait 1 tour, la roue menée fait  $\frac{3}{2}$  tours (donc : 1 tour et demi). Mais la première fois dans le même sens et la deuxième fois en sens contraire. Le premier cas représente le nombre  $\frac{3}{2}$ , et le second le nombre  $(-\frac{3}{2})$ .

Le nombre représenté par une transmission par courroie est le nombre ayant pour partie numérique le rapport de la circonférence de la roue menante à la circonférence de la roue menée, et ayant pour signe + si la courroie est droite et - si la courroie est croisée.

I.R.E.M. - Mathématiques 4e.

p.42

**Théorème 2 :** *Le point d'une droite  $d$  le plus proche d'un point  $A$  est le point  $H$ , pied de la perpendiculaire à  $d$  passant par  $A$ .*

p.42

**Théorème 3 :** *La composée de deux symétries axiales d'axes perpendiculaires est une symétrie centrale ayant pour centre le point d'intersection des deux axes. Ce résultat ne dépend pas de l'ordre dans lequel on prend les deux symétries.*

p.43

**Théorème 4 :** *La composée d'une symétrie d'axe  $d$  suivie d'une symétrie d'axe  $\delta$  parallèle à  $d$  est une translation. Cette translation est celle qui transforme un point quelconque de  $d$  en son symétrique par rapport à  $\delta$ .*

p50

**Théorème 2 :** *L'ensemble des points équidistants de deux droites sécantes est constitué par les bissectrices de ces droites.*

p51

**Théorème 3 :** *Par le point de concours de deux bissectrices d'un triangle, il en passe une troisième.*

p52

**Théorème 4 :** *(troisième cas d'égalité \*). Si deux triangles ont leurs côtés égaux, l'un se transforme en l'autre par la composée d'au plus trois symétries axiales.*