

UNIVERSITÉ GRENOBLE II ■ PIERRE MENDÈS–FRANCE
U.F.R. SCIENCES DE L'HOMME ET DE LA SOCIÉTÉ
DÉPARTEMENT DES SCIENCES DE L'ÉDUCATION
Année 2003

THÈSE

- Pour obtenir le grade de docteur de l'université GRENOBLE II
- Discipline Sciences de l'Éducation
- Présentée et soutenue publiquement par Denis FOUCAMBERT le 16 décembre 2003

SYNTAXE, VISION PARAFOVÉALE ET PROCESSUS DE LECTURE

Contribution du modèle structural à la pédagogie

VOLUME I

Directeur de thèse :

Jacques BAILLÉ, Université Grenoble II

Jury :

Christine BARRÉ de MINIAC, IUFM de Grenoble

Greg BROOKS, University of Sheffield

Jacques CRINON, IUFM de Créteil

Sommaire

Introduction	3
Première partie : Approches théoriques	10
Chapitre 1. L'effet d'oubli de lettres.....	11
Chapitre 2. Syntaxe et linguistique.....	29
Chapitre 3. La vision et la lecture.....	56
Chapitre 4. Traitement syntaxique et compréhension de phrase.....	78
Chapitre 5. Les observations neurophysiologiques.....	97
Deuxième partie : Approches expérimentales	112
Chapitre 6. Hypothèses.....	113
Chapitre 7. Les performances en lecture de la population expérimentale	120
Chapitre 8. L'oubli de lettres : évaluations des fréquences, rôles et natures des mots	130
Chapitre 9. Mise en rapport de l'effet d'oubli de lettres et des performances en lecture	147
Chapitre 10. L'importance de la vision périphérique pour le modèle structural.....	154
Troisième partie : Place de la syntaxe dans la pédagogie de la lecture	167
Chapitre 11. Contexte théorique.....	168
Chapitre 12. Le barrage de lettres chez les enfants de cycle 2	178
Chapitre 13. Peut-on entraîner le lecteur à construire des cadres syntaxiques ?.....	190
Chapitre 14. Différence de résultats en fonction du type d'entraînement dans le seul groupe expérimental	210
Conclusion.....	217
Bibliographie.....	224
Index des figures.....	242
Index des tableaux.....	244
Index des auteurs.....	246
Table des matières.....	250

Introduction

Les évaluations en lecture recueillies dans le cadre de la Journée d'Appel à la Préparation à la Défense fournissent des informations assez précises sur le niveau des jeunes Français, filles et garçons, âgés de 17 ans, à la fin donc de l'école obligatoire. Les résultats font ressortir deux groupes distincts : le premier regroupe les jeunes en grandes difficultés de lecture, soit 12% de la classe d'âge étudiée (14,5% pour les garçons et 9,4% pour les filles) alors que le deuxième, représentant donc les « bons lecteurs », est lui-même scindé en trois sous-groupes que hiérarchisent les niveaux de compréhension en lecture. Si on considère les 88% de « bons lecteurs », 97% d'entre eux ont une « compréhension immédiate » satisfaisante, ce qui correspond à « un niveau simple de compréhension des informations explicitement délivrées par le texte ». Ils ne sont plus que 71,1% à satisfaire aux exigences d'une bonne « compréhension logique » qui correspond à la capacité de « mettre en relation plusieurs informations explicitement fournies dans le but de construire le sens du texte ». Enfin, seulement 58,6% des bons lecteurs (soit 51% de la classe d'âge) sont capables d'avoir une « compréhension fine » qui implique un « traitement approfondi des informations » pour « suppléer aux lacunes du texte », pour « déduire à partir des informations explicites d'autres informations qui en découlent et donc être capables d'accéder à l'implicite »¹. Ces résultats sont similaires à ceux obtenus depuis plusieurs années aux évaluations faites en classe de sixième et, depuis l'année 2003, en classe de cinquième (Dauphin, Rebmeister, & Zelty, 2003a; Dauphin, Rebmeister, & Zelty, 2003b).

Les études internationales, comme l'enquête PIRLS² effectuée en 2001 dans 35 pays, classe la France au-dessus de la moyenne des pays participants. Cependant, une analyse ne prenant en compte que les résultats des pays européens, la situe en deçà de la moyenne générale, comme en témoigne la comparaison avec l'ensemble des pays membres de l'OCDE (Colmant & Mulliez, 2003; Mullis, Martin, Gonzalez, & Kennedy, 2003). Les jeunes

¹ Toutes les citations précédentes ainsi que les pourcentages fournis sont extraits de (Robin & Rocher, 2003)

² PIRLS : Progress in International Reading Literacy Study

Français de CM1 n'apparaissent qu'au 19^{ème} rang pour la compréhension des textes littéraires et au 13^{ème} rang pour celle des textes narratifs. En outre, pour ce qui est des quatre compétences observées par l'enquête (prélever, inférer, interpréter, apprécier), ces jeunes Français montrent une assez grande homogénéité dans leur classement avec toutefois une moins bonne réussite aux deux compétences *interpréter* et *apprécier*. Cette faiblesse se retrouve dans les résultats des items (10% de l'ensemble) reflétant un degré supérieur de compréhension : la France passe sous la barre de la moyenne internationale. Ces items ne devraient toutefois pas être hors de portée des élèves puisqu'il leur est demandé :

- Pour la lecture des textes littéraires, d'être capables d'intégrer différentes idées pour construire une interprétation sur les caractères des personnages, de pouvoir repérer les intentions de l'auteur, et d'être capables d'expliquer l'implicite du texte et les différents thèmes de l'histoire.
- Pour les textes informatifs, d'être capables d'intégrer à leurs propres connaissances antérieures des informations issues de matériaux écrits différents pour les appliquer dans des situations qu'ils peuvent rencontrer dans la vie de tous les jours.

D'après ces résultats, plus de 45% des lecteurs formés par notre système éducatif ne sont pas compétents quand il s'agit d'accéder à un maniement subtil du matériau écrit.

De nombreuses recherches travaillent à décrire les causes de ces résultats et, parfois, à y remédier. On pense ici tout particulièrement aux travaux des sociologues de la culture, qui ont mis à jour les contraintes qui grèvent un accès démocratique à l'écrit. Pour autant, les aspects techniques ne sauraient être négligés et nous partageons, à cet égard, l'avis de Jean-Claude Passeron : « ce qu'on entend par lecture, qui va de la lecture cultivée à la lecture savante, ne peut pas être techniquement abordée à un niveau rudimentaire de maîtrise du code écrit » (Passeron, 1987).

Face à cette hétérogénéité de maîtrise du code écrit, plusieurs disciplines se sont penchées sur le problème pour mieux comprendre les processus à l'œuvre dans une lecture efficace afin d'aider les élèves dans cet apprentissage. Le volume de publication dans ces domaines est considérable, ce qui peut sembler étonnant face à la relative stagnation des résultats des élèves. La psychologie, qu'elle soit expérimentale ou cognitive, et l'ensemble des disciplines regroupées sous la bannière des neurosciences participent de cet effort.

La figure 1 en présente un raccourci saisissant en les organisant d'après la temporalité de l'apparition de leur « objet » de recherche après l'apparition d'un stimulus, ici un mot.

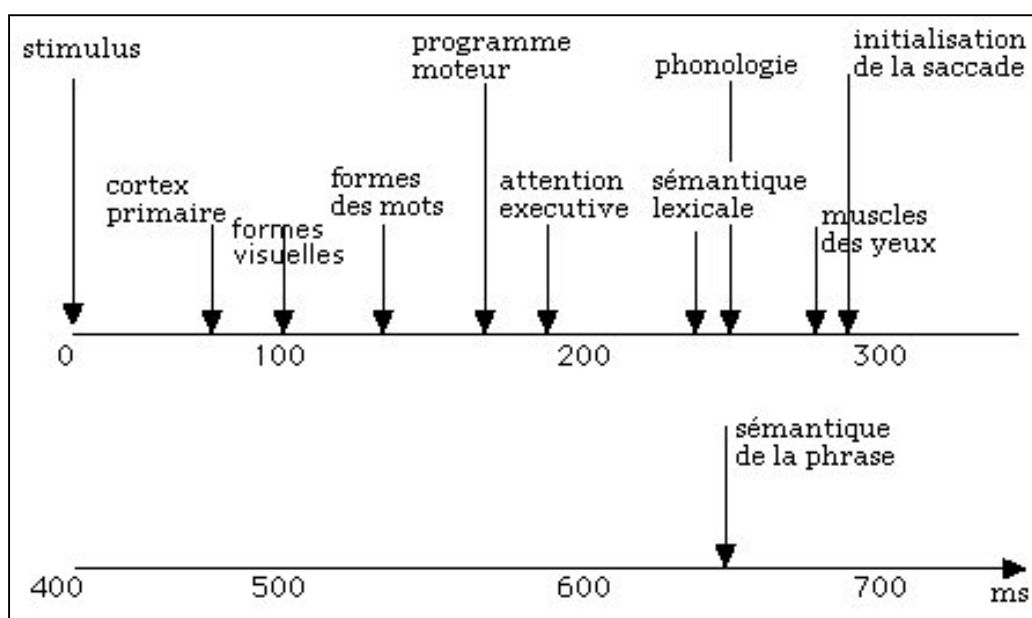


Figure 1 : Ligne du temps montrant différents processus pendant la lecture de mots. D'après (Posner, Abdullaev, McCandliss, & Sereno, 1999)

On y reconnaît les problématiques sur la programmation de la saccade oculaire, sur la place de la phonologie, sur l'accès au lexique que ce soit au niveau du mot ou de son intégration au sens de la phrase, sur les aspects neuronaux et les parcours de l'information et enfin sur le rôle de la forme du mot. Chacun de ces domaines suscite une littérature « controversiale », preuve de la fécondité de ces approches. Pourtant, on notera que si le terme *sémantique* apparaît deux fois dans cette figure, celui de *syntaxe* est absent. Cette absence est d'autant plus surprenante que l'étude de la syntaxe (et les

débats houleux qui l'accompagnent) a été au cœur de l'étude de la langue depuis une quarantaine d'année. Sans doute, certains des débats qui ont accompagné les travaux de Noam Chomsky (en particulier ceux sur l'innéité) ont empêché une approche sereine des processus syntaxiques. Toujours est-il que la syntaxe reste en retrait dans l'étude des processus de lecture, bien que, depuis quelques années, il semble qu'on redécouvre son importance dans les processus langagiers, en grande partie grâce aux très nombreux travaux utilisant l'imagerie cérébrale et qui révèlent une claire dissociation des « circuits » de la syntaxe et de ceux du sémantique.

C'est dans ce cadre que s'intègre le modèle structural de lecture, autour duquel s'organise principalement notre travail. Assez récent puisque les premiers écrits datent de 1991, il est le fruit du travail de Asher Koriat et de Seth Greenberg, auxquels se joignent dorénavant d'autres chercheurs, souvent psychologues expérimentaux. Ce modèle considère que le lecteur extrait très tôt au cours de sa lecture la structure syntaxique de la phrase qu'il parcourt, laquelle va ensuite guider le repérage et l'intégration des unités lexicales sémantiquement riches. Pour ses inventeurs, cette construction de la structure de la phrase repose prioritairement sur des indices morphosyntaxiques. Nous reviendrons en détail sur cette théorie et sur ses différentes implications.

Ce qui nous a intrigué dans cette tentative d'explication du fonctionnement de la lecture, c'est la manière dont ces indices syntaxiques sont prélevés au cours de la lecture. Les thèses de Greenberg et Koriat font ainsi écho à des travaux d'autres disciplines, notamment ceux d'Alain Berthoz dans le domaine de la perception et ceux qui ont été écrits à la suite des travaux de Chomsky sur la syntaxe. La force des Sciences de l'Éducation réside en grande partie dans l'interdisciplinarité (Marmoz, 1988), et la recherche que nous présentons ici tente de s'inscrire dans cette optique. Le premier intérêt du modèle de Koriat et Greenberg est sans doute la multiplicité des approches nécessaires à sa complète explicitation. Le deuxième est qu'il pourrait contribuer à l'amélioration des performances en lecture des élèves, si tant est que deux

hypothèses soient vérifiées : qu'il entretient bien un rapport avec les performances de lecture et qu'une « transposition didactique » en soit possible.

* *
*

Le travail que nous présentons s'organise autour de trois grandes parties :

La première est plus particulièrement consacrée à l'exposition des travaux issus de différents champs de recherches, susceptibles d'aider la compréhension du fonctionnement du modèle structural lui-même. Après sa présentation détaillée, nous synthétiserons les apports de la littérature provenant des disciplines qui l'éclairent et le précisent et parfois suggèrent de nouvelles hypothèses quant à son fonctionnement. C'est ainsi que nous présenterons des travaux de cette partie de la linguistique spécialisée dans l'étude de la syntaxe et d'autres travaux sur le fonctionnement de la vision dans les processus de lecture. Nous décrirons ensuite la place des traitements syntaxiques dans le fonctionnement cérébral, soit d'une manière théorique, sous la forme de l'étude des grands modèles explicatifs, soit par l'exposé des observations du cerveau que ce soit par la méthode classique ou par celle, récente, de l'imagerie cérébrale.

La deuxième partie est la première de nos deux contributions expérimentales. Elle est consacrée à l'étude, chez le lecteur adulte, des composantes du modèle structural de lecture. Nous commencerons par une présentation de l'ensemble des hypothèses expérimentales que nous détaillerons dans les quatre chapitres suivants. Deux grands objectifs organisent cette première contribution : d'abord travailler à partir du modèle structural de lecture pour préciser, d'une part, l'ampleur des structures syntaxiques construites par le lecteur et, d'autre part, les indices utiles à leur construction. Ensuite, nous chercherons à vérifier l'existence d'un lien entre cette habileté syntaxique et le niveau de lecture.

Deux points sont au programme de notre seconde contribution expérimentale : en premier lieu, l'observation, chez plus de 800 élèves de cycle 2, de la mise

en place du modèle structural et ensuite l'exploration et l'évaluation, pour des collégiens de cinquième, d'un entraînement spécifiquement axé sur les composantes du modèle structural et de son transfert sur les performances de lecture.

Première partie :
Approches théoriques

Chapitre 1.

L'effet d'oubli de lettres

Depuis quelques années, l'épreuve dite d'oubli de lettres (*missing letter effect*) (Corcoran, 1966; Healy, 1976) est devenue commune dans les recherches de psychologie expérimentale sur les processus de lecture. On présente à un sujet un texte écrit en lui demandant de le lire à sa vitesse habituelle, dans le but de le comprendre. La contrainte de l'épreuve réside dans la demande supplémentaire qui lui est faite de barrer (ou d'encercler) une lettre à chaque fois qu'il la rencontre au cours de sa lecture, sans effectuer de retour en arrière sur les lettres qu'il aurait eu la sensation d'oublier.

A partir des résultats de cette épreuve, on constate qu'un certain nombre de lettres ont été omises par le lecteur. Qui plus est, la répartition de ces oublis n'est pas aléatoire et semble répondre à certaines régularités. Plusieurs modèles s'efforcent de rendre compte de ces phénomènes.

1.1. Le modèle de Healy et ses évolutions

Une des initiatrices de cette épreuve est Alice Healy de l'Université du Colorado dès la fin des années 70.

1.1.1. Les unités concurrentes

Les travaux de Healy ont montré des taux d'oubli statistiquement plus importants dans les mots fréquents, comme par exemple, *and*, *the* ou *for*. Cette constatation et le modèle explicatif qui en découle sont repris fréquemment dans d'autres travaux similaires. Healy et ses associés (Healy, 1976; Drewnowsky & Healy, 1977; Healy, 1980; Healy & Drewnowsky, 1983; Healy, 1976; Healy & Drewnowsky, 1983) postulent que le processus de lecture d'un texte met en jeu, de manière simultanée, différents domaines d'analyse, portant sur les formes des lettres, les lettres, les syllabes, les mots et des groupes de mots. Ils supposent que, dès que le lecteur identifie une unité

par un de ces domaines, les processus de tous les niveaux parallèles sont stoppés même s'ils n'ont pas été menés jusqu'à leur terme (figure 2).

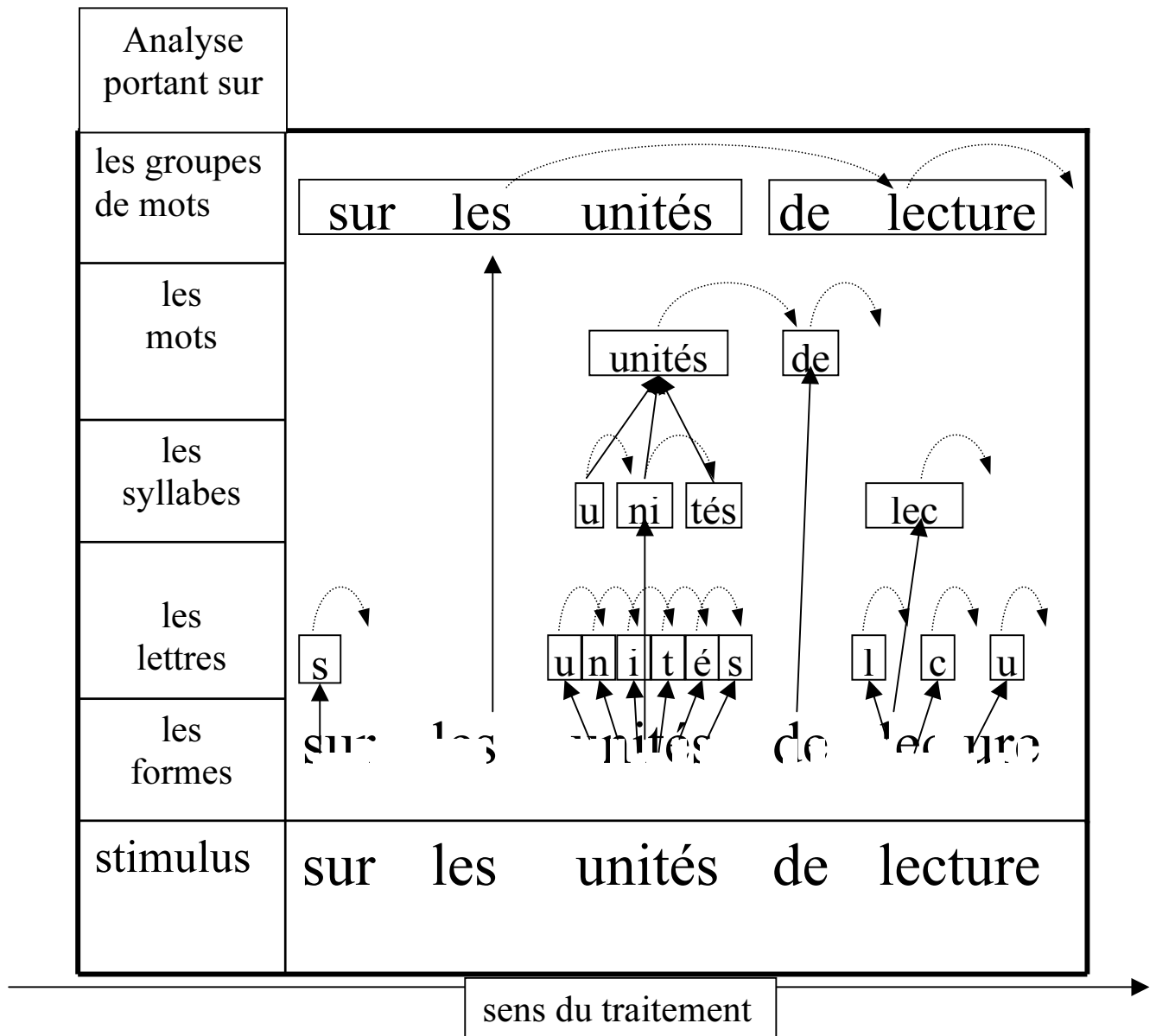


Figure 2 — Les différents domaines d'analyse selon l'hypothèse d'unités concurrentes et parallèles de Healy (d'après (Healy, 1994)).

L'oubli de lettres s'explique, selon Healy, par la reconnaissance rapide (globale) d'un mot fréquent, ce qui stoppe les niveaux d'analyse parallèles comme la reconnaissance des lettres, et provoque donc un taux plus fort d'oubli. Ce modèle a permis d'expliquer bien des observations comme, par exemple, l'oubli plus important de lettres dans les noms fréquents que dans les noms rares (Drewnowsky & Healy, 1982) : la grande fréquence de certains

noms rend probable leur identification avant celle de leurs constituants. En outre, si on perturbe le modèle orthographique de ces mots fréquents (introduction d'astérisques dans la chaîne de caractères), on élimine cet effet d'oubli de lettres. De même, si on modifie l'apparence visuelle de ces mots fréquents par l'introduction de polices rares, de fautes d'orthographe ou de lettres mélangées, l'effet s'amointrit considérablement.

1.1.2. Le traitement parafovéal

Une modification de ce modèle d'unités concurrentes a été proposée récemment par l'équipe d'Healy. L'hypothèse nouvelle d'un traitement parafovéal (Hadley & Healy, 1991), vise à mieux rendre compte des oublis de lettres, par la relation entre le lieu où le mot prend place sur la rétine et les traitements qui en sont faits. Cette hypothèse reprend les études préalables sur le mouvement des yeux³ qui ont montré à la fois que les mots fonctionnels (articles, conjonctions, prépositions) ne sont souvent pas fixés durant la lecture (Carpenter & Just, 1983), et que les mots de trois lettres sont plus souvent sautés quand ils sont des mots fonctionnels (*the*) que quand ils sont des verbes (O'Regan, 1979). En conséquence, les mots fréquents, comme le sont les mots fonctionnels, seraient plus souvent identifiés en vision parafovéale (par les zones périphériques de la rétine) pendant la fixation d'un mot précédent, alors que les mots moins courants devraient être traités par la zone fovéale⁴ de l'œil (Rayner & Pollatsek, 1987).

Dans ces conditions, les différents domaines d'analyse (dont nous avons parlé précédemment : lettres, mots, etc.) ne sont pas tous accessibles selon l'endroit où le mot est perçu. S'il est traité dans la zone fovéale, tous les niveaux d'analyse sont possibles, en particulier celui des lettres. En conséquence, très peu de lettres devraient être omises puisque les omissions sont attribuées au traitement plus rapide du niveau-mot (stoppant par-là l'analyse du niveau-lettre avant qu'il n'ait terminé) ; ceci n'est vrai que dans le cas des mots fréquents.

³ Nous traiterons en détail dans le chapitre 3 ce qui touche à la vision, sa physiologie, les indices traités en fonction des lieux de perception sur la rétine, les saccades oculaires, etc...

⁴ Au centre de l'œil, là où la qualité de la perception visuelle est la meilleure, la plus fine (voir chapitre 3)

En revanche, plus le mot est traité loin de la fovéa, moins les informations concernant les lettres sont accessibles. Dans ce cas, seuls les domaines d'analyse correspondant au moins au mot sont efficaces. L'identification du mot ne serait efficace que pour les mots très fréquents (donc très bien connus), alors que celle des mots plus rares resterait hasardeuse.

En bref, le taux d'omission des lettres dans les mots fréquents est dû à leur plus grande probabilité d'être identifié dans la zone parafovéale, lors de la fixation précédente.

1.1.3. Le temps de travail

A partir de 1994, Healy a reconsidéré toutes ses hypothèses et a proposé une nouvelle approche de son cadre explicatif (Healy, 1994) ; (Moravcsik & Healy, 1995) ; (Moravcsik & Healy, 1998). Tout particulièrement, elle a suggéré une *hypothèse de temps de travail* (processing time hypothesis) : le traitement visuel des lettres continuerait jusqu'à l'accès syntaxique et sémantique, ce qui implique que la détection des lettres serait en rapport avec la vitesse d'identification de la signification (ou du rôle) du mot. Les mots très communs ou les mots proposant peu de significations différentes sont traités plus rapidement que les mots plurisémiqiques. L'analyse des lettres continuant en parallèle jusqu'au succès du traitement sémantique disposerait donc de plus de temps pour aller à son terme dans le cas où celui-ci serait laborieux..

En conséquence, un certain nombre de facteurs pourraient réduire le temps d'accès, comme la familiarité avec le sens d'un mot ou encore sa configuration visuelle. Tous les éléments susceptibles de réduire le temps de traitement des unités augmente la probabilité que des lettres les constituant soient omises. Cette hypothèse ne remet pas en cause celle émise par les mêmes auteurs sur la vision parafovéale, elle la complète pour ce qui est du traitement dans la zone fovéale. En quelque sorte, elle permet de différencier des aspects de bas-niveau (la vision parafovéale) et des aspects de haut-niveau (l'accès sémantique).

1.2. Le modèle de Koriat et ses évolutions

Les expériences menées depuis 1991 par Koriat & Greenberg ont permis d'envisager un autre cadre explicatif à cet effet d'oubli de lettres (Greenberg & Koriat, 1991); (Greenberg, Koriat, & Shapiro, 1992; Koriat & Greenberg, 1991; Koriat, Greenberg, & Goldshmid, 1991; Koriat & Greenberg, 1993; Koriat & Greenberg, 1994; Koriat & Greenberg, 1996a); (Koriat & Greenberg, 1996b).

1.2.1. Le rôle de la syntaxe

En notant que les mots les plus fréquents étaient souvent ceux qui jouaient un rôle bien particulier dans la syntaxe de l'anglais, ils ont supposé que c'était ce rôle syntaxique qui induisait l'effet d'oubli de lettres. Un grand nombre d'expérimentations leur a permis de conforter ce point de vue. Un second pan de ces études s'est construit sur certaines caractéristiques de l'hébreu écrit, permettant de distinguer les effets de fréquence et de rôle (fonctionnel vs. sémantique) des mots étudiés.

- D'une part, il est possible en hébreu d'introduire les morphèmes fonctionnels (par exemple les prépositions, les conjonctions, les articles,...) comme préfixes des mots à rôle sémantique (Koriat & Greenberg, 1991). Ainsi, on peut écrire de deux façons la préposition *à*, soit comme un mot séparé (*EL*), soit comme préfixe du nom le suivant (*L*). On aura donc pour la forme « à HAIFA » deux constructions possibles, de sens identique : « EL HAIFA », où le EL est un mot fonctionnel très fréquent, HAIFA étant le mot à rôle sémantique, et la forme « LHAIFA », de fréquence beaucoup plus rare. Les résultats montrent, comme prévu par Healy, que la lettre L est massivement oubliée dans le mot très fréquent EL, mais aussi dans la forme préfixée LHAIFA, de fréquence pourtant plus rare. Une autre expérimentation montre également que cet effet obtenu sur la première lettre d'un mot ne l'est que si cette lettre représente un morphème fonctionnel (opposition «LHB » = « sang » et « LRB » = « à Rabi »). Ces résultats suggèrent que la détection des lettres dans une unité lexicale dépend de la fonction de cette unité au sein de la phrase.

- D'autre part, les auteurs ont utilisé des homographes hébreux pour montrer l'effet du contexte local sur l'oubli de lettres. Par exemple, la forme « *SMR* » peut signifier soit « *ce monsieur* », soit « *volé* ». Placés dans des contextes qui lui ôtent toute ambiguïté, on observe plus de difficultés à repérer cette lettre (*S*) quand le contexte induit un préfixe fonctionnel. En tout état de cause, en proposant une chaîne identique de caractères, l'effet du patron (pattern) orthographique, tel que le propose Healy, ne peut plus exister. On voit ici que le contexte ne sert pas seulement à identifier la bonne signification du mot ambigu (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Dopkins, Morris, & Rayner, 1992), mais qu'il guide le repérage des unités morphologiques à rôle syntaxique ou sémantique.
- Dans une autre série d'expérimentations, Koriat & Greenberg ont démontré que cet effet d'oubli de lettres pouvait même être obtenu avec des non-mots. Si on place, dans un texte en anglais, des chaînes de caractères vides de sens à des positions normalement occupées par des mots fonctionnels, on observe plus d'oubli de lettres que si ces non-mots occupaient des positions de mots à rôle sémantique. Bien entendu, l'effet de fréquence ne peut pas ici servir d'explication. Qui plus est, on trouve des résultats similaires dans la langue hébraïque : si on compare des non-mots constitués d'un préfixe fonctionnel attaché à un élément pseudo-lexical (*lbgn*, où *l* représente la préposition et *bgn* le non-mot) placés soit dans un contexte influençant le choix de la première lettre (*l*) comme élément fonctionnel, soit dans un contexte influençant le choix de la première lettre comme une partie sémantiquement chargée, les oublis de lettres se feront massivement dans le premier cas et dans le même rapport que s'il s'agissait de mots réels. Ainsi, même quand la chaîne orthographique est inconnue, on trouve davantage d'omissions quand cette chaîne apparaît à une place occupée normalement par un élément à rôle syntaxique plutôt que par un élément à rôle sémantique.

1.2.2. La lecture structurale

Tous ces éléments ont conduit Koriat & Greenberg à proposer un modèle alternatif à celui de Healy. Ce modèle, dit de lecture structurale, postule que les effets d'oubli de lettres sont dus au rôle syntaxique du mot dans une phrase et non à sa fréquence. Le point central de cette hypothèse suggère la prégnance de l'organisation structurale : en lisant, le lecteur tente d'établir un cadre structural de la phrase (ou du groupe de mots) et d'en intégrer le sens. Mais c'est la structure qui mène la danse (*lead the way*), (Koriat & Greenberg, 1996a). D'une certaine manière, le lecteur tente d'élaborer un squelette du groupe de mots ou de la phrase qu'il parcourt, sur la base d'une rapide et superficielle analyse, éventuellement à l'aide de la vision parafovéale. Ce squelette accueille les informations sémantiques de la phrase. Les mots fonctionnels comme les prépositions, articles, conjonctions sont donc les premiers indices traités par le lecteur pour établir ce cadre ; ils sont ensuite rejetés à l'arrière-plan alors que l'intégration du sens se poursuit sur le cadre structural pré-construit. Ce point de vue explique particulièrement bien le taux plus élevé d'oublis dans les morphèmes fonctionnels : traités de prime abord, ils ne sont plus cognitivement actifs au moment du barrage de la lettre. Dans cette optique, la détection des lettres se passerait après l'analyse morphologique des mots, ce qui diffère du point de vue d'Healy, qui considère que les différents traitements ont lieu en parallèle jusqu'au succès de l'un d'entre eux.

La construction « on-line » du cadre structural de la phrase est soumise à des hypothèses générées par le contexte aussi bien qu'à des clés (au sens musical) syntaxiques présentes dans le texte, tout particulièrement les morphèmes fonctionnels. Ces indices dirigent l'attention du lecteur vers les éléments les plus signifiants. En conséquence, ces clés sont repérées très précocement dans la lecture de la phrase mais, une fois utilisées pour établir une structure, elles sont reléguées à l'arrière-plan et cèdent la place aux unités sémantiquement riches.

Dans des expérimentations publiées en 1996, Koriat & Greenberg ont essayé de compléter ce phénomène d'oubli de lettres : si, une fois le cadre établi, les

morphèmes fonctionnels sont rejetés à l'arrière-plan, les morphèmes à rôle sémantique devraient alors, par contraste, mieux ressortir. Ainsi les noms suivant directement un morphème fonctionnel devraient avoir moins d'oublis de lettres que ceux suivant un morphème à rôle sémantique. Cette hypothèse sera vérifiée, aussi bien dans le cas de noms anglais, de noms hébreux préfixés que dans le cas de non-mots, mis à la place des mots à rôle sémantique. Une fois encore, c'est bien la fonction que ces éléments occupent dans la phrase qui explique le phénomène d'oubli de lettres, toujours dans l'optique d'une activation prioritaire de cadres structuraux à remplir par les informations sémantiques. Cette notion d'activation peut se rapprocher de celle développée par Le Ny, même si cet auteur l'utilisait dans le domaine sémantique (Le Ny, 1989). Ici, l'activation est avant tout d'ordre syntaxique.

Avec Koriat qui suppose finalement que l'extraction de ce cadre structural se fait à partir de l'interaction de trois grands facteurs (syntaxique, sémantique et visuel), il faut souligner ce que les données relatives aux non-mots suggèrent quant au rôle de la syntaxe, c'est-à-dire au rôle de la position relative des éléments et de leur relation à l'intérieur de la phrase (Grevisse & Goosse, 1980). On note ainsi que la manipulation de l'ordre naturel des mots diminue l'effet d'oubli de lettres : quand l'ordre est modifié (Drewnowsky & Healy, 1977; Healy, 1976) ou quand les mots sont placés à des endroits inappropriés (Koriat & Greenberg, 1991), la détection des lettres s'en trouve améliorée. En conséquence, les contraintes contextuelles aident à spécifier le statut syntaxique des morphèmes. Cependant, on note moins d'oublis de lettres dans les non-mots que dans les mots (Koriat & Greenberg, 1991) et plus d'oublis dans les mots correctement orthographiés que dans ceux contenant des fautes. Healy suggère que cet effet est dû à une plus grande familiarité individuelle avec les mots (Healy, Conboy, & Drewnowsky, 1987). Koriat & Greenberg l'expliquent par la contribution de facteurs contextuels et lexicaux : les non-mots, eux, ne seraient traités que par des effets de contexte.

La totalité des observations sur lesquelles repose le modèle de Koriat a été pratiquée à l'aide de la tâche de détection de lettres dans du matériel écrit, qu'il soit présenté sous forme de textes, de listes de mots ou encore mot-à-mot sur

un écran d'ordinateur. Il faut noter que cette dernière tâche a été critiquée parce qu'elle corrompt la fluidité du processus normal de lecture, ce qui limite la validité des modèles sur la lecture normale qui en seraient extrapolés (Rayner & Pollatsek, 1989). C'est sans doute la raison principale qui a conduit Koriat, Greenberg et Kreiner à proposer une nouvelle façon d'observer cette habileté du lecteur dans la construction préalable de cadres structuraux (Koriat, Greenberg, & Kreiner, 2002). L'étude du rapport entre la prosodie de la lecture à haute voix et la syntaxe des phrases lues prend place dans les travaux de Koriat. Les auteurs rappellent (pour la deuxième fois dans l'ensemble de leurs articles sur le modèle structural) que quand on demande de lire à voix haute le poème de Lewis Carroll « *A travers le miroir* » (Carroll, 1971) qui se termine ainsi :

« Tu as tué le Jabberwock !
Dans mes bras, mon fils rayonnois !
Ô jour frableux ! callouh, calloc ! »
Le vieux glouffait sa joie.

Il était reveure : les sclitueux toves
Sur l'alloinde gyraient et vriblaient ;
Tout flivoreux vaguaient les borogoves ;
Les verchons fourgus bourniflaient.

les lecteurs sont tout à fait capables de produire une lecture fluide, avec une prosodie cohérente, à peu près semblable d'une personne à l'autre. On explique généralement ce fait en remarquant que ce poème respecte les règles morphosyntaxiques ce qui offre au lecteur une prosodie sans surprise⁵.

Cette observation vient confirmer le modèle structural de lecture. D'une part, en montrant que la prosodie intervient dès la première lecture à haute voix

⁵ On pourrait également citer les écrits de Jean Tardieu, qui a subtilement joué d'une syntaxe cohérente et d'un lexique surprenant. Extrait de **Un mot pour un autre** :

IRMA

Y-a que, Madame, yaque j'ai pas de gravats pour mes haridelles, plus de stuc pour le bafouillis de ce soir, plus d'entregent pour friser les mouches... plus rien dans le parloir, plus rien pour émonder, plus rien... plus rien... (*Elle fond en larmes*).

MADAME, *après avoir vainement exploré son sac de nouveau et l'avoir montré à Irma*.

Et moi non plus, Irma ! Ratissez : rien dans ma limande !

IRMA, *levant les bras au ciel*.

Alors ! Qu'allons-nous mariner, Mon Dieu ?

MADAME, *éclatant soudain de rire*.

Bonne quille, bon beurre ! Ne plumez pas ! J'arrime le Comte d'un croissant à l'autre.

(*Confidentielle*.) Il me doit plus de cinq cents crocus ! (Tardieu, 1951)

d'une phrase, suggérant par là l'immédiateté de l'intégration de la prosodie. D'autre part en soulignant que la prosodie est particulièrement tributaire de l'existence d'une syntaxe cohérente mais n'est que peu sensible à une modification ou à une destruction complète du sens de la phrase lue. A une même structure correspondent des prosodies très similaires, quelle que soit la nature des informations sémantiques de la phrase ; des phrases avec des structures différentes obligent à des prosodies différentes, respectant notamment une répartition des pauses recouvrant largement la succession des syntagmes principaux de la phrase. Cependant, il serait inexact de dire que les informations sémantiques n'apportent aucun renseignement utile à la prosodie. On note ainsi que la cohérence sémantique influe sur une prosodie naturelle, mais uniquement si la cohérence syntaxique est intacte, jamais si elle est dégradée. Selon nos auteurs, cette information signifierait que le contenu sémantique ne peut être traité correctement que si le cadre syntaxique est cohérent, alors que, si le cadre syntaxique est déficient, l'analyse sémantique le sera également⁶. Ces résultats sont donc en totale cohérence avec le modèle de lecture structurale de Koriat, qui, rappelons-le, suppose que l'extraction immédiate de la structure de la phrase ouvre la route à l'intégration sémantique.

1.2.3. La question de la taille des structures

Dans leurs premiers travaux, Koriat et Greenberg avaient supposé, sans le vérifier, que la taille des structures ouvertes par le lecteur était courte, se situant autour de celle d'un syntagme réduit (Greenberg & Koriat, 1991; Koriat & Greenberg, 1991; Koriat et al., 1991; Koriat & Greenberg, 1994). Cependant, certains résultats ont introduit le doute quant à la taille des unités créées par le lecteur. Les caractéristiques propres à la langue allemande ont permis d'y voir un peu plus clair (Musseler, Koriat, & Nisslein, 2000). Avant d'exposer les observations qu'ils ont réalisées, il faut rappeler que les phrases organisées selon l'ordre Sujet-Verbe-Objet (S-V-O) sont plus faciles à

⁶ Nous reviendrons sur un processus similaire observé grâce aux nouvelles techniques d'imagerie cérébrale au chapitre 5

appréhender que les phrases écrites dans l'ordre inverse Objet-Verbe-Sujet (O-V-S). Les travaux de van Dijk et Kintsch ont proposé qu'un syntagme nominal en position de sujet avait un rôle significativement plus important dans le traitement de la phrase qu'en position d'objet ((van Dijk & Kintsch, 1983) cité par (Musseler et al., 2000)). L'allemand, admet une double-construction pour certaines phrases, les deux étant syntaxiquement correctes, même si la structure S-V-O reste préférée par les lecteurs. Le format S-V-O est attendu par le lecteur et, dans le cas où la phrase est présentée selon l'ordre O-V-S, un effort est nécessaire pour construire le cadre syntaxique correct. Dans ces conditions, l'effet d'oubli de lettres devrait être moins prononcé dans les phrases présentées selon l'ordre O-V-S.

Les observations effectuées sur la langue allemande offrent comme premier intérêt la réplication du modèle structural de lecture dans cette langue, après l'hébreu, l'anglais et le français (pour lequel les premières observations remontent à 1997 (Saint-Aubin & Poirier, 1997)). Le second est de montrer que les structures créées par les lecteurs se rapportent soit à des syntagmes simples soit à des unités fonctionnant au niveau de la phrase entière. Les lettres sont plus souvent oubliées quand le mot qui les contient est dans une phrase au format S-V-O que dans une phrase O-V-S. Par ailleurs, dans une phrase S-V-O, ou dans une phrase O-V-S, le mot cible peut apparaître soit dans le premier syntagme nominal de la phrase dans le nominatif, soit dans le dernier syntagme nominal si la phrase est à l'accusatif. Dans ces conditions, on constate que la position du mot n'interfère pas avec l'opposition S-V-O / O-V-S. La plus forte différence d'oubli de lettres dans le cas S-V-O est cohérente avec l'idée que le format S-V-O est le format préféré du lecteur, favorisant en cela plus facilement son repérage. Cependant, la présence d'un effet également dans le cas non standard O-V-S montre la création d'une structure plus réduite, correspondant alors au syntagme nominal présent à l'intérieur de la phrase.

Ces résultats suggèrent que la magnitude de l'effet est sensible aux contraintes structurales qui existent à une échelle plus vaste que le simple groupe syntagmatique introduit par le mot fonctionnel observé. Les auteurs en concluent que l'établissement des cadres structuraux est sans doute plus

sensible aux différents niveaux d'organisation de la phrase qu'ils ne le pensaient initialement.

1.3. Modèle structural et vision parafovéale

L'hypothèse de traitement parafovéal prédit que les lettres sont plus souvent oubliées quand les mots sont perçus dans la zone parafovéale, ce qui serait majoritairement le cas des mots fonctionnels (Hadley & Healy, 1991). Dans ce cas, il faut se demander si les lettres présentes dans les mots fonctionnels sont également oubliées, que le mot soit perçu en vision fovéale ou parafovéale. Si, en situation de lecture, on neutralise la zone parafovéale, on devrait observer la même fréquence d'oubli dans les mots à rôle syntaxique et dans les mots à rôle sémantique. Qui plus est, même s'il existe une forte corrélation positive entre l'identification parafovéale des mots et l'omission de lettres, il reste que ces deux événements peuvent être dus à un troisième facteur qui serait responsable à la fois de l'effet d'oubli de lettres et de l'identification parafovéale des mots fonctionnels.

Les travaux de Saint-Aubin et Klein ont précisément pour objectifs de décider si, quand on élimine la perception parafovéale, l'effet d'oubli de lettres persiste (Saint-Aubin & Klein, 2001). Dans l'ensemble, ils ont repris les protocoles expérimentaux mis en place par Healy (Drewnowsky & Healy, 1977; Healy, Oliver, & McNamara, 1987) afin d'élucider l'hypothèse parafovéale, tout en affinant et en améliorant le matériel expérimental : en effet, Healy et ses associés n'avaient pris en compte ni la longueur des mots-cibles, ni leur rôle syntaxique ni la position de la lettre cible dans les mots.

- Leur première expérience compare la moyenne d'oubli de la lettre *t* dans le mot à rôle syntaxique *the* et dans des mots à rôle sémantique de trois lettres commençant également par *t* (*tea*, *tax*). Les mots-cibles sont intégrés dans des textes en prose, qui sont présentés soit sous une forme classique, soit sous forme de liste (un mot par ligne). Les résultats montrent un plus fort taux d'oubli pour le mot fonctionnel *the*, même dans la présentation sous forme de liste où la vision périphérique à droite est impossible.

- La deuxième expérience proposée montre encore un taux supérieur d'oubli pour *the*, cette fois présenté dans un texte, mais dans lequel le lecteur déplace une fenêtre, soit de 5 signes (vision parafovéale impossible), soit de 15 signes (vision parafovéale possible). On note entre ces deux passations que le taux d'oubli du mot fonctionnel *the* diminue pour la fenêtre de cinq signes par rapport à la fenêtre de quinze signes, donc entre les présentations où la vision parafovéale est possible par rapport aux présentations où elle est impossible. Cette diminution n'est pas observée pour les mots à rôle sémantique, confirmant par-là le rôle accru de la vision parafovéale dans le traitement des mots à rôle syntaxique.
- La troisième expérience fait apparaître les textes mot à mot sur un écran d'ordinateur. Cette présentation empêche toute prise d'information dans la zone parafovéale de l'œil. Encore une fois, et contrairement à ce que stipulaient les hypothèses parafovéales, on observe un effet important de la fonction syntaxique du mot, même en vision fovéale.
- La quatrième expérience est une réplique de la troisième et utilise comme mot-cible à rôle syntaxique la préposition *for* puisque les taux d'oublis sont environ 2 fois plus importants pour les mots à rôle syntaxique que pour les mots à rôle sémantique (61,9 vs. 36,1 pour *the* ; 44,6 vs. 19,8 pour *for*)⁷. Ces résultats sont en accord avec le modèle structural, puisqu'il prédit que l'oubli s'explique par le rôle structural du mot dans la phrase (Koriat & Greenberg, 1996a; Koriat & Greenberg, 1996b), que ce mot soit vu par la fovéa ou perçu par les zones périphériques de l'œil.
- La cinquième expérience, quant à elle, enregistre les mouvements des yeux pendant que le lecteur lit, sur un écran d'ordinateur, un texte dans un format habituel. Il doit appuyer sur un bouton de la souris au moment où il repère la lettre-cible. Comme dans les procédures expérimentales précédentes, les auteurs observent une proportion d'oubli plus importante dans les mots à rôle syntaxique parmi le sous-ensemble des mots perçus en vision centrale. De même, pour ce qui est des mots perçus en vision périphérique, on observe

⁷ On notera un oubli plus important pour *the* que pour *for*. Nous reviendrons plus tard sur ce phénomène.

également plus d'oublis dans les mots à rôle syntaxique. D'une manière générale, les mots fixés par la fovéa présentent moins d'oublis que les mots perçus parafovéalement et un plus grand nombre de mots fonctionnels sont perçus parafovéalement.

		Mots fixés	Mots sautés
Type de cible	the	68	79
	Mots – contrôles	26	51
	for	51	72
	Mots – contrôles	18	42

Tableau 1 — Pourcentage d'omission de lettres en fonction du type de mot et du type de fixation. d'après exp. 5 de (Saint-Aubin & Klein, 2001)

En résumé, l'hypothèse de traitement parafovéal permet tout à fait d'interpréter les expériences 1, 2 et 5, lorsque la vision périphérique n'est pas supprimée. Cependant, il semble difficile d'expliquer par cette hypothèse, même complétée par celles d'unités concurrentes et du temps de travail, les observations réalisées quand la vision parafovéale est impossible : dans ces conditions, en effet, il ne devrait plus y avoir de différences d'oubli de lettres selon le rôle des mots.

Le modèle structural, quant à lui, sort renforcé de cette série d'expériences. Rappelons que ce modèle postule l'extraction rapide et première de la structure de la phrase lue, cette structure étant ensuite complétée par les informations sémantiques :

- Les morphèmes à rôle fonctionnel aident à construire le squelette de la phrase.
- L'établissement de ce squelette est basé sur une analyse rapide et superficielle des marques structurelles (mots fonctionnels, ponctuations,...) comme points d'ancrage (Greenberg et al., 1992).

- Cette analyse est construite en partie à l'aide des informations extraites de la vision périphérique du texte (Koriat & Greenberg, 1996a).

Dans ce modèle, les mouvements des yeux reflètent les processus cognitifs actifs, et plus précisément, l'établissement du cadre structurel (Koriat & Greenberg, 1996b; Saint-Aubin & Klein, 2001). A l'inverse, l'hypothèse du traitement parafovéal postule que les processus cognitifs dépendent du lieu de la rétine qui envoie des informations. L'ensemble des observations que nous venons de rapporter à partir du travail de Saint-Aubin et Klein est compatible avec le modèle de Koriat. Que le mot-cible soit perçu en vision fovéale ou parafovéale, on trouve plus d'oublis de lettres dans les mots à rôle syntaxique que dans les mots à rôle sémantique : ils contribuent, de toute façon, à l'installation du cadre structurel. Cependant, il reste plus difficile d'expliquer à l'aide du modèle structural qu'on trouve moins d'omissions de lettres dans les mots fonctionnels quand on supprime la vision parafovéale. Certes, dans ce cas, on a toujours une différence importante entre les mots à rôle sémantique et à rôle syntaxique (tableau 1) mais une différence significative existe aussi selon le type de présentation.

1.4. Commentaires et hypothèses

1. Les premiers commentaires concernent le fait que différentes informations sur la lettre et le mot semblent ne pas avoir été abordées dans l'ensemble des études que nous venons d'évoquer. Saint-Aubin critique à juste titre le fait que, dans ses expérimentations, Healy (1987) fait des comparaisons d'oublis de lettres en confrontant l'oubli de la lettre t dans des mots à rôle syntaxique (*the*) et dans des mots à rôle sémantique (*bathed*) : la lettre ne se situe pas à la même place (Saint-Aubin & Klein, 2001). Doit-on considérer qu'une lettre située à la fin d'un long mot a les mêmes chances d'être barrée, toutes choses égales par ailleurs, qu'une lettre située au début d'un mot court ? Pourtant, depuis longtemps déjà, différents travaux ont étudié les mécanismes de positionnement de l'œil dans le mot, soit lorsqu'il est présenté de manière isolée, soit lorsqu'il fait partie d'un texte, et les recherches continuent

aujourd'hui. Dès 1979, Rayner montrait que les lecteurs tendent à fixer le mot vers la gauche de son milieu (Rayner, 1979). On parlera désormais, à la suite des travaux de O'Regan, de position optimale dans le mot (Vitu, O'Regan, & Mittau, 1990; O'Regan & Jacobs, 1992; O'Regan, 1992). D'autres types de travaux ont également montré qu'une combinaison peu fréquente de lettres perçue de manière parafovéale attirait la fixation suivante (Hyönä, 1995). Enfin, il existe également des travaux sur la visibilité des lettres en fonction de l'écart entre le point de fixation et la lettre. Par exemple, la probabilité d'identifier une lettre dans une chaîne aléatoire de caractères qui est de 1 quand on fixe la lettre passe à 0,4 quand l'œil se situe seulement à cinq caractères de la lettre à identifier (Townsend, Taylor, & Brown, 1971). Il semble surprenant que ce type d'investigation n'ait pas été pris en compte, même si les mots proposés par les auteurs présentés ici étaient le plus souvent des mots courts.

On peut donc légitimement se demander si la taille du mot dans lequel figure la lettre à barrer a une influence sur son oubli. La plupart des expériences que nous avons relatées contournent cette question en fixant une taille de mots identique pour les mots fréquents à rôle syntaxique (*the*) et pour les mots plus rares à rôle sémantique (*tax*).

Nous tenterons, pour notre part, de faire fonctionner ces différentes variables dans l'explication de l'effet d'oubli de lettres.

2. Si les observations sur l'allemand (Musseler et al., 2000) ont permis de différencier les taux d'oubli selon l'ordre de la phrase (S-V-O vs. O-V-S) et par là de mettre en évidence que l'effet d'oubli de lettres est susceptible de révéler une capacité d'organisation structurelle aussi bien au niveau du syntagme qu'au niveau de la phrase, d'autres observations seraient nécessaires pour corroborer ces faits. Koriat avait déjà envisagé de réaliser une observation des oublis de lettres en fonction des natures de mots, donc en différenciant, par exemple, déterminants, prépositions et conjonctions. Notons que, si les déterminants organisent le syntagme nominal, les conjonctions, en permettant l'articulation des propositions entre elles, sont parties prenantes de la structure de la phrase

entière⁸. En conséquence, l'observation de l'oubli de lettres en fonction des natures devrait permettre de préciser ce que Koriat appelle la magnitude de l'effet d'oubli, en fonction du type d'organisation sur lequel ouvre le mot à rôle structurant. Cette observation sera un de nos objectifs du chapitre 8.

3. Il ressort de l'analyse de la littérature que la vision parafovéale a une importance non négligeable sur l'établissement de cadres syntaxiques au cours de la lecture. Les observations de Saint-Aubin, considérant que le modèle structural de lecture pourrait être comparé à *un troisième facteur qui serait responsable à la fois de l'effet d'oubli de lettres et de l'identification parafovéale des mots fonctionnels* (Saint-Aubin & Klein, 2001) vont dans ce sens. Si on accepte ce raisonnement, on peut considérer que l'identification parafovéale (c'est-à-dire ce qui est à droite du point de fixation) n'est pas seulement une conséquence inerte de la vision, mais que c'est un mécanisme volontaire répondant à une demande cérébrale dans le but de construire le cadre syntaxique de la phrase. C'est ce que nous essaierons de vérifier dans l'expérimentation rapportée au chapitre 10.

4. Peu de travaux ont porté sur la mise en rapport de l'oubli de lettres avec les caractéristiques du lecteur. Il a été montré que l'âge pouvait introduire des différences (Greenberg, Koriat, & Vellutino, 1998) : l'effet d'oubli de lettres est plus marqué à 13 ans qu'à 7 ans, ce qui conduit les auteurs à conclure à une prise de conscience progressive du rôle de certains mots dans l'organisation de la structure des phrases. Cependant, il ne semble pas qu'il y ait eu d'autres travaux prenant en compte le niveau de performance du lecteur, et l'éventuelle relation entre la vitesse de lecture ou la compréhension et l'oubli différentiel de lettres. Les populations étudiées jusqu'ici sont considérées comme étant homogènes. Nous pourrions faire l'hypothèse que les lecteurs les plus habiles, possédant une vitesse de lecture plus rapide et une compréhension meilleure, disposent soit d'une meilleure capacité à construire des cadres syntaxiques, avec un plus faible taux d'erreurs, soit d'une capacité à faire ces anticipations

⁸ Cet aspect sera détaillé au chapitre 2

syntaxiques sur des unités plus longues que les mauvais lecteurs. C'est aussi ce que nous essayerons de vérifier au chapitre 9.

Auparavant, et pour éclairer ce qui est en débat à travers les deux modèles qui se proposent d'expliquer l'effet d'oubli de lettres, il est nécessaire :

1. de préciser le fonctionnement de la syntaxe de la langue. Dans le deuxième chapitre, nous évoquerons l'apport de la linguistique à la compréhension du fonctionnement de la structure des langues.
2. de décrire le fonctionnement de notre système de vision. Vision fovéale, vision parafovéale, saccades oculaires, type d'informations traitées, rapport anticipation/vérification, zones visuelles spécialisées dans les formes écrites constitueront le troisième chapitre.
3. de faire le point sur les principaux modèles théoriques tentant d'expliquer le traitement du langage par le cerveau. Nous y consacrerons le quatrième chapitre.
4. de détailler certaines observations issues principalement de la neuro-imagerie qui nous permettent d'explorer, pendant des temps de lectures, les zones cérébrales en action et la temporalité de leur activation.

Chapitre 2. **Syntaxe et linguistique**

2.1. Introduction

A l'intérieur de la linguistique théorique, la syntaxe est l'étude de l'architecture des groupes de mots, des propositions et des phrases (Shapiro, 1997). La parution du livre de Noam Chomsky en 1957, *Structures Syntaxiques*, a profondément modifié la manière d'aborder la syntaxe (pour la traduction française : (Chomsky, 1969)). Même si le modèle décrit par Chomsky est aujourd'hui critiqué, le plus souvent davantage pour les positions qu'il défend sur l'innéité que pour son analyse purement linguistique, il conserve une grande force d'interprétation et regroupe de nombreux partisans tant dans le domaine de la linguistique que dans celui des neurosciences (Pinker, 1999). Nous ne nous intéresserons pas ici aux débats (par ailleurs très instructifs) relatifs à ce qui serait inné ou acquis dans le langage et dans son apprentissage (Piatelli-Palmarini, 1982), mais nous nous servons de l'apport indiscutable de la théorie chomskienne à l'analyse linguistique de la syntaxe, et par là, peut-être, à la façon dont l'esprit humain est organisé : « Les traces, les cas, les X-barres et autres accessoires de la syntaxe [...] doivent faire partie de notre vie mentale inconsciente. » (Pinker, 1999).

Il est d'ailleurs surprenant que les auteurs du modèle de lecture structurale que nous venons de présenter ne fassent aucune référence aux travaux de Chomsky bien qu'ils manipulent abondamment à la fois les notions de *structure* et de *syntaxe*. Ce faisant, ils se privent sans doute de schémas explicatifs pouvant renforcer leur modèle, d'autant qu'ils postulent l'indépendance des processus syntaxiques et sémantiques, à l'instar de ce que Chomsky revendique dès ses premiers travaux où il considère que les outils qu'il utilise devront être purement syntaxiques, donc élaborés sans aucun rapport avec des considérations de sens ou de signification.

Nous essaierons à la fin de ce chapitre de relire le modèle de lecture structurale à l'aide des outils conceptuels fournis par la linguistique moderne, seulement

après avoir présenté le fonctionnement de celle-ci dans ses acceptions les plus actuelles.

2.2. Un peu d'histoire

Il n'est bien évidemment pas question de refaire ici l'histoire de la linguistique. Même si, comme le rappelle Chomsky dans son introduction du livre de Pollock, *Langage et Cognition*, « l'étude du langage compte parmi les domaines d'analyse systématique les plus anciens » (Pollock, 1997), on a coutume de dater les grands débuts de la linguistique, en temps que science autonome, à l'apparition des travaux de Saussure et de Sapir. Cette linguistique synchronique se caractérise par une démarche structuraliste de surface (Bronckart, 1977), où l'objectif est avant tout de définir les unités linguistiques pertinentes et d'analyser leur relation avec le monde extralinguistique. Ces unités, que ce soit le signe chez Saussure ou le symbole chez Sapir, sont de l'ordre du mot. Comme le note Bronckart, la publication en 1916 du *Cours de Linguistique Générale* de Saussure a relativement peu d'influence sur son époque, encore largement dominée par le courant néo-grammairien dont « les membres se consacraient à la grammaire historique et comparée ; ils s'efforçaient d'expliquer les changements historiques ou les parentés entre les langues en invoquant essentiellement les filiations linguistiques [...] de telle sorte que leur système d'explication a pu être qualifié d'atomiste » (Bronckart, 1977).

S'inspirant des travaux de Saussure, quelques linguistes se sont regroupés ensuite au sein de ce qui s'appellera le Cercle Linguistique de Prague (C.L.P.). Jakobson, Tesnière ou encore Martinet en feront partie. Ce structuralisme européen utilisera principalement deux démarches : d'une part, la description structurale qui consiste à isoler et à disséquer les composants et les processus élémentaires d'une donnée linguistique et, d'autre part, la description fonctionnelle qui cherche à observer en quoi l'organisation des structures est subordonnée à une finalité, à une intention des locuteurs. Dans le cadre du C.L.P., on retiendra tout particulièrement :

1. les travaux de Jakobson sur l'analyse des fonctions du langage : expressive (ou émotive), conative, référentielle, phatique, métalinguistique et poétique.
2. les travaux sur l'aspect phonique du langage avec la différenciation de la phonétique qui se consacre à l'analyse des sons émis dans toutes les langues (l'étude de leurs caractéristiques articulatoires et acoustiques) et de la phonologie qui s'intéresse aux sons en tant qu'éléments pertinents dans le système phonologique de la langue parlée par le sujet. Dans cette catégorie, on notera également les travaux de Martinet autour de la célèbre double-articulation⁹ qui constitue, selon lui, le seul aspect véritablement universel du langage.
3. les travaux de Martinet qui ont permis une définition d'unités linguistiques reprises par la suite dans bien d'autres travaux. Le monème y est défini comme la plus petite unité porteuse de sens. Dans la phrase
Françoise s'est cassé un ongle.
chaque mot est un monème. Mais le monème ne se confond pas nécessairement pas avec le mot. Dans *supermarché*, ou *redire*, on trouve deux monèmes. Les monèmes qui constituent les racines sont des lexèmes (*dire*, *marché*, dans l'exemple) (les adjectifs, verbes, noms...) et ceux qui servent de foncteur grammatical sont des morphèmes (les préfixes, suffixes, articles, pronoms, prépositions, conjonctions...). Les morphèmes sont une classe fermée dans la mesure où on ne peut en introduire de nouveaux sans modifier l'équilibre du système. Les lexèmes constituent une classe ouverte : les néologismes sont un exemple de la capacité d'accroissement de cette classe sans remise en cause du système.

Pour terminer ce rapide inventaire des courants linguistiques, parmi lesquels les travaux de Chomsky vont apparaître, il convient de citer la théorie

⁹ La première articulation est constituée par les unités significatives que sont les unités lexicales et grammaticales (morphèmes). Les unités distinctives dépourvues de sens en elles-mêmes forment la deuxième articulation (phonèmes).

glossématique fondée par le linguiste danois Hjelmslev. Son principe majeur est qu'« à toute séquence de comportements correspond une structure (« sous-jacente »), qui permet de l'analyser et de la décrire, en utilisant un nombre restreint de prémisses et de théorèmes » (Bronckart, 1977).

2.3. L'analyse syntaxique

Rappelons tout d'abord que la syntaxe peut être définie comme étant « la partie de la grammaire qui ordonne les mots en syntagmes¹⁰ et en phrases » (Pinker, 1999). Pinker, en se situant dans la mouvance de Chomsky, considère donc que la syntaxe fait partie de la grammaire, et que la grammaire doit être entendue comme un « ensemble de règles qui détermine la forme et le sens des mots et des phrases dans une langue donnée telle qu'elle est parlée par une communauté donnée ». Cette inclusion de la syntaxe dans la grammaire sera toujours sous-entendue dans la suite de ce travail.

Un des objectifs principaux des travaux initiés par Chomsky est de construire un modèle qui « rendrait compte des principes et des processus selon lesquels les phrases sont construites dans les langues particulières ». Pour Chomsky, le processus majeur est ce qu'il appelle la créativité, reprenant ce que Humboldt déclarait il y a bien longtemps, à savoir que le langage « fait un usage infini de moyens finis » (cité par (Changeux, 2002)). Ce processus est à même de produire ou de comprendre un nombre infini de phrases à l'aide d'un nombre restreint et fini d'éléments et de règles pour les ordonnancer. Cette faculté de langage (la capacité à créer la totalité de la langue) attesterait d'un mécanisme sous-jacent qui permettrait son élaboration.

La grammaire sera donc ce modèle complet qui permet de rendre compte de la totalité de la langue, l'étude des processus qui témoignent de la créativité linguistique. Un processus inconscient nous permet de juger instantanément si une phrase est correcte ou non grammaticalement à partir d'une intuition d'acceptabilité. Si l'on considère les deux phrases suivantes :

¹⁰ groupe de mots qui se comporte comme une unité dans une phrase et qui a un sens cohérent.

- (a) Benoît et Jeanne sont à la piscine.
- (b) Benoît et Jeanne nagent la piscine.

l'intuition linguistique nous permet de décider sans effort que la phrase (a) est correcte, c'est à dire grammaticale, bien formée, tandis que la phrase (b) est agrammaticale, mal formée. Cette notion de grammaticalité permet de définir le corpus que le linguiste doit étudier. En outre, l'intuition d'acceptabilité, partagée par tous les utilisateurs d'une langue et qui se réalise dans le jugement de grammaticalité, est, selon Chomsky, une intuition de la grammaire de la langue (au sens défini plus haut). Cette grammaire¹¹ ne fait pas appel au sens des phrases mais seulement à leur acceptabilité structurelle. La célèbre phrase

- (c) Des idées vertes incolores dorment furieusement.

est parfaitement acceptable sur le plan grammatical, même si elle ne veut pas dire grand chose. Le fait que tout locuteur d'une langue puisse émettre ce type de jugement vis-à-vis des séquences qu'on lui soumet implique l'existence d'une compétence linguistique sous-jacente partagée par l'ensemble des locuteurs. Cette compétence comporte, entre autres, des connaissances et des règles qui déterminent la bonne formation syntaxique, et qui associent une (des) interprétation(s) sémantique(s) appropriée(s) aux séquences bien formées. Cet ensemble de connaissances et de règles permet aux locuteurs de produire et d'interpréter des énoncés. Or, il est certain que ces règles ne leur ont jamais été enseignées de façon explicite : aucune grammaire n'en fournit la liste exhaustive. L'objectif principal de la linguistique consiste à rendre compte

¹¹ Dans son livre « L'instinct du langage » (1997), Pinker accorde plusieurs acceptions au terme de grammaire :

1. *grammaire générative* : ensemble de règles qui déterminent la forme et le sens des mots et des phrases dans une langue donnée
2. *grammaire mentale* : grammaire générative hypothétique qui est stockée inconsciemment dans le cerveau d'un individu.
3. *grammaire prescriptive ou stylistique* : grammaire enseignée à l'école et dans les livres, ensemble de directives sur la manière dont on « doit » utiliser un dialecte.
4. *grammaire de structure syntagmatique* : grammaire générative constituée uniquement des règles qui définissent les structures syntagmatiques.
5. *grammaire transformationnelle* : grammaire composée d'un ensemble de règles de structure syntagmatique qui construisent un arbre de structure profonde et d'une ou plusieurs règles de transformation qui déplacent les syntagmes dans la structure profonde pour produire un arbre de structure de surface.
6. *grammaire universelle* : schéma de base sous-jacent aux grammaires de toutes les langues humaines ; désigne également la circuiterie présente dans le cerveau de l'enfant, qui lui permet d'apprendre la grammaire de la langue de ses parents.

de cette compétence¹², à en expliciter le contenu et le fonctionnement. Si cette entreprise est menée à bien, elle nous renseigne sur les propriétés universelles du langage. Le travail du linguiste est donc de créer un modèle (ou une grammaire) de la langue qui rend compte des énoncés grammaticaux. Trois grands critères doivent répondre de cette grammaire :

1. un critère d'adéquation aux données : les phrases générées par le modèle devront être acceptables par un locuteur humain et le modèle doit pouvoir analyser toutes les phrases générées par tout locuteur.
2. un critère de généralité : Chomsky rappelle que la construction de la structure linguistique doit être établie à l'aide de concepts (tel que phonème, syntagme) indépendants de toutes langues particulières
3. un principe de simplicité : l'objectif est de rendre compte des faits de langue de la manière la plus économique.

La théorie chomskienne a subi de profonds bouleversements depuis une quarantaine d'années et il n'y a pas lieu ici de les préciser. Cependant, la représentation en graphes arborescents (même si elle est antérieure aux travaux de Chomsky), constitue une base sur laquelle ont pu se développer non seulement les premiers modèles chomskiens mais aussi des théories plus récentes.

2.4. De la succession des mots à l'analyse en constituants

Il est devenu banal de préciser que les mots ne sont pas seulement des éléments disposés de gauche à droite dans un ordre sériel. Ils appartiennent à des "catégories" différentes. Les catégories traditionnelles de la linguistique incluaient des acceptions vagues pour les noms (on pourrait dire par exemple qu'ils représentaient des personnes, des lieux ou des choses, mais alors qu'en

¹² Chomsky précise ainsi la différence entre compétence et performance : "La compétence est la connaissance que le locuteur-auditeur idéal a de sa langue. La performance, elle, concerne l'emploi effectif de la langue, les énoncés du vrai locuteur-auditeur, avec ses faiblesses et ses limites, dans les situations concrètes". Nous maintiendrons cette différence dans l'ensemble du travail que nous présentons, où nous aurons maintes fois l'occasion de recourir à ces notions de performance et de compétence.

est-il du mot *boucher*), pour les verbes (s'ils réfèrent à des actions, que penser du mot fabrication ?) ou encore pour les prépositions. Les linguistes ont examiné les propriétés phonologiques, morphologiques et distributionnelles des mots pour en extraire des catégorisations lexicales plus rationnelles. Sur le plan distributionnel, les noms apparaissent à certains endroits de la phrase où les verbes ne peuvent prétendre. Les noms peuvent être accompagnés par des adjectifs (*gros mangeur*) à l'inverse des verbes (*gros manger*). Les noms peuvent être quantifiés et spécifiés (*trois chats/le chat*), à l'inverse des verbes (*trois chasser / un chasser*). En conséquence, on ne dira plus d'un nom qu'il représente une personne, mais qu'il appartient à la classe distributionnelle des noms puisqu'il peut être précédé et/ou suivi d'un adjectif, précédé d'un article... Qui plus est, deux termes appartenant à la même classe sont susceptibles de se substituer l'un à l'autre, la substitution étant, en fait, le test d'appartenance à une classe. A ces catégories, s'en rajoutent d'autres, de nature fonctionnelle, comme par exemple les déterminants. En conséquence, « une phrase, ce n'est pas tant l'association de mots en fonction de leur sens que le choix d'éléments de diverses classes complémentaires et leur agencement sur l'axe syntagmatique selon l'ordre imposé par les contraintes distributionnelles propres au français » (Chiss, Filliolet, & Mainguenaud, 1998).

Soit les deux phrases suivantes :

(d) Mon frère déteste cet homme mesquin.

(e) Mon frère trouve cet homme mesquin.

Même si ces deux phrases sont formées de mots appartenant à des catégories syntaxiques identiques (Dét N V Dét N A), elles ne s'interprètent pas de la même manière. Leurs structures respectives diffèrent, c'est à dire que les mots ne sont pas regroupés semblablement dans (d) et (e) et n'entretiennent pas des relations identiques entre eux. Dans (d), les mots de la séquence *cet homme mesquin* sont regroupés et forment un constituant ; on pourrait écrire : *C'est cet homme mesquin que mon frère déteste*. Dans (e), les mots de la séquence *cet homme mesquin* peuvent être séparés en deux groupes et former ainsi deux constituants distincts. On pourrait écrire : *C'est cet homme que mon frère trouve mesquin*. Dit d'une façon intuitive et informelle, dans (d) le verbe est

suivi d'un seul syntagme, qui est son complément d'objet direct : V SN. Dans (e), en revanche, le verbe est suivi de deux syntagmes, un syntagme nominal et un syntagme adjectival : V SN SA. On ne forme donc pas une phrase avec des mots, mais avec des syntagmes. Pour passer des mots à la phrase, on peut rencontrer plusieurs niveaux intermédiaires formés de différents types de syntagmes. Les groupements de mots qui respectent les règles de la bonne formation syntaxique et qui se comportent comme une unité face aux opérations syntaxiques sont des syntagmes. L'élément central de chaque syntagme en détermine la nature : on parlera ainsi de syntagme nominal, syntagme adjectival, syntagme verbal, syntagme prépositionnel, syntagme adverbial, ... L'élément central d'un syntagme est aussi appelé la tête¹³ du syntagme. Ainsi, tout syntagme peut être considéré comme la projection d'une tête.

On représente sous la forme d'un arbre les phrases (d) et (e) :

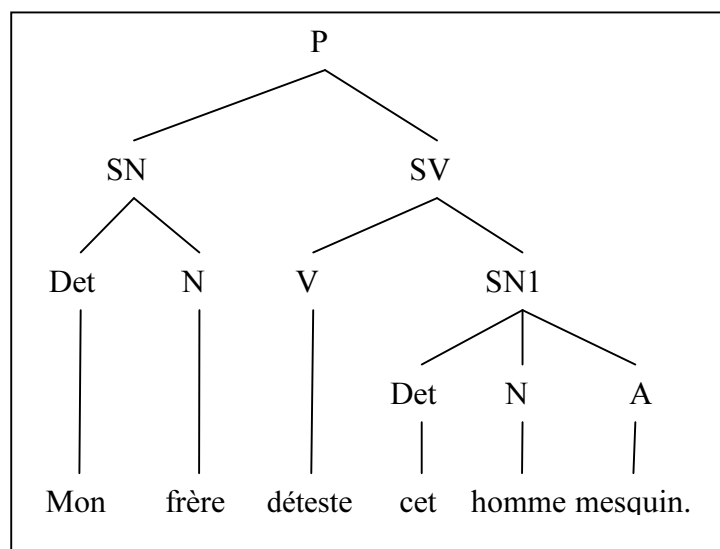


Figure 3 — Arbre représentant la phrase (d)

¹³ Pour des concepts similaires, on trouve la dénomination *tête* (Pollock, 1997), et *noyau* (Pinker, 1999; Chiss et al., 1998). Dans la suite du texte, nous utiliserons indifféremment les deux termes.

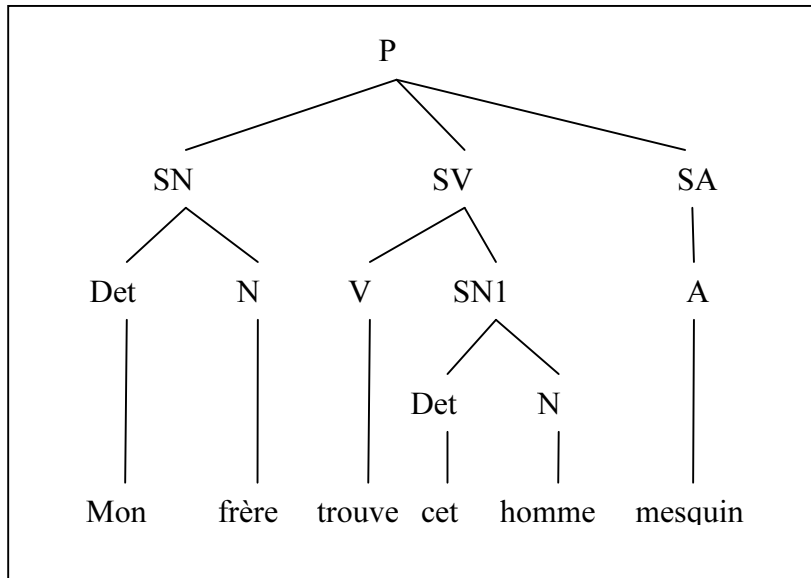


Figure 4 — Arbre représentant la phrase (e)

Cette analyse distributionnelle des constituants de la phrase permet de la considérer comme une hiérarchie de niveaux et un système de dépendances syntaxiques entre diverses catégories. Ainsi, un arbre nous fournit instantanément trois informations :

1. la hiérarchie des constituants présents dans la phrase. Elle est donnée par une lecture descendante de l'arbre.
2. les catégories auxquelles appartiennent les différents composants de la phrase (trouve est un verbe, mesquin est un adjectif, etc.).
3. une représentation des fonctions des éléments et de leurs relations avec les autres éléments.

Les relations syntagmatiques entre les éléments sont données par une lecture ascendante de l'arbre. Grâce à cette représentation, la phrase n'apparaît plus comme une suite linéaire de mots, mais comme une hiérarchie de groupes syntaxiques s'emboîtant les uns dans les autres pour former des unités de plus en plus vastes convergeant vers la phrase (Chiss et al., 1998).

Considérons maintenant la phrase :

(f) Mon frère ferme la porte avec la clé.

Cette phrase comporte une ambiguïté d'ordre syntaxique dans la mesure où elle supporte deux interprétations :

- (g) C'est avec une clé que mon frère ferme la porte
(h) Mon frère ferme la porte qui a une clé.

Deux arbres syntaxiques peuvent être construits, qui symbolisent chacun une interprétation possible (figures 5 et 6).

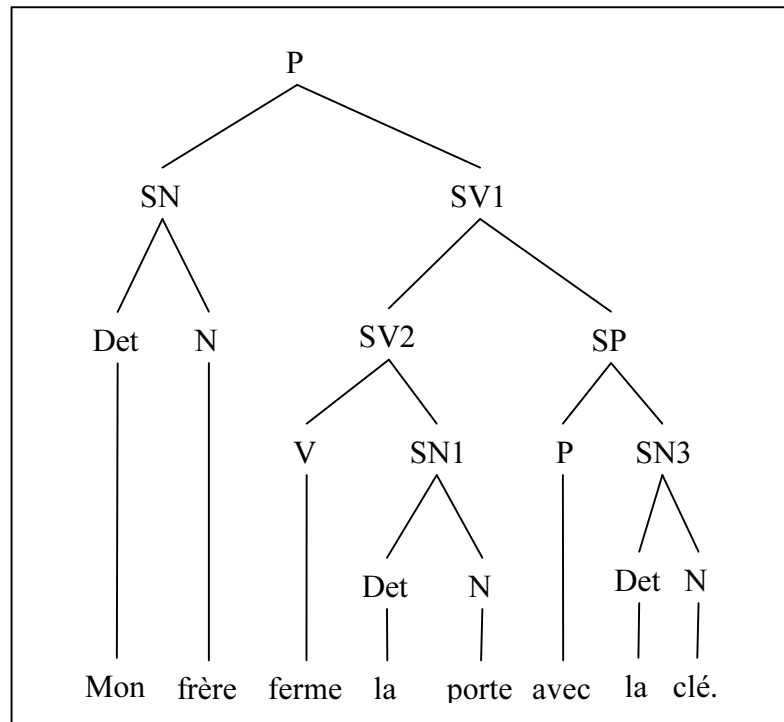


Figure 5 — Représentation syntaxique de la phrase (f) avec l'interprétation (g)

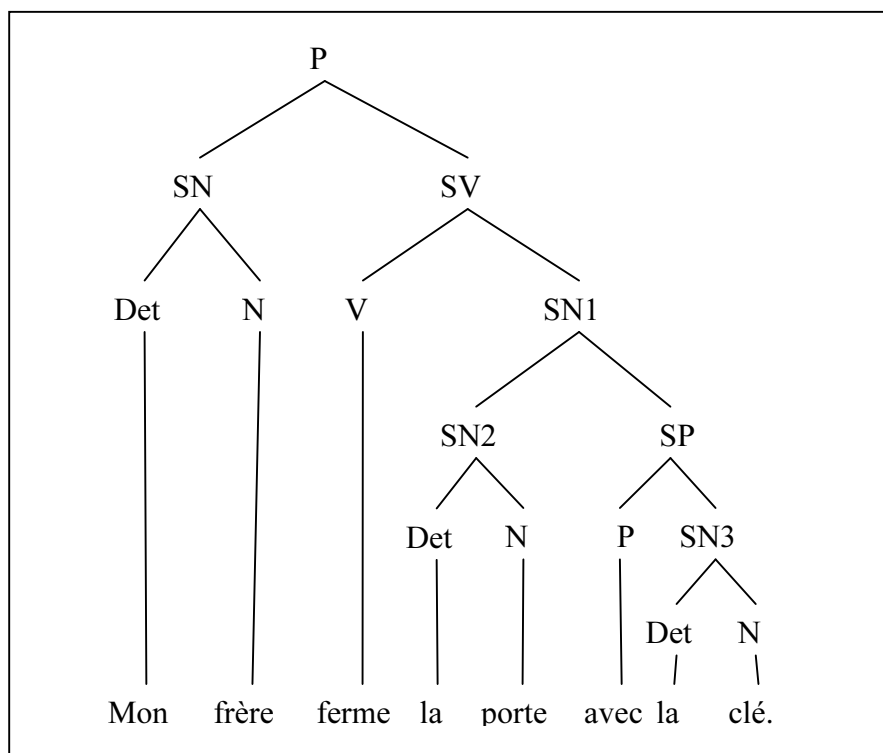


Figure 6 — Représentation syntaxique de la phrase (f) avec l'interprétation (h)

Si on s'intéresse au syntagme verbal (SV), on remarque que :

1. Dans l'interprétation (g), le syntagme prépositionnel *avec la clé* modifie directement le syntagme verbal SV2 et les deux syntagmes sont rattachés au syntagme verbal SV1 hiérarchiquement plus élevé, formant l'ensemble *ferme avec la clé*.
2. Dans l'interprétation (h), le syntagme prépositionnel SP *avec la clé* modifie le syntagme nominal *la porte* (SN2), ces deux syntagmes étant attachés au syntagme nominal SN1 hiérarchiquement plus élevé formant l'ensemble *la porte avec la clé*.

Dans les deux arbres que nous venons de construire, on peut faire une distinction entre les nœuds lexicaux qui sont constitués uniquement des catégories lexicales (Nom, Verbes, etc.) et les nœuds syntagmatiques comme les syntagmes nominaux, les syntagmes prépositionnels qui, eux, incluent à leur tour différentes catégories lexicales. Par exemple, le syntagme prépositionnel *avec une clé* contient une préposition suivie d'un déterminant et d'un nom. Une propriété importante de cette théorie de la structure de la phrase est que chaque catégorie syntagmatique doit contenir un *noyau*, et que ce noyau et l'ensemble du syntagme partagent les mêmes propriétés. Il ne peut donc y avoir de syntagme verbal sans verbe, ni de syntagme prépositionnel sans préposition.

La plupart du temps, une phrase consiste en un syntagme nominal suivi d'un syntagme verbal. Par exemple :

(i) P \Rightarrow SN + SV : [P [SN Mon frère] [SV ferme la porte avec la clé]]

Un SN consiste obligatoirement en un Nom qui peut être précédé éventuellement d'un déterminant et précédé ou suivi d'un syntagme adjectival. Un syntagme verbal contient au moins un verbe et potentiellement plusieurs autres éléments, parmi lesquels d'autres syntagmes nominaux ou prépositionnels et même d'autres phrases (P). On a donc :

- (j) SV \Rightarrow V : [SV [V nage]]
SV \Rightarrow V SN : [SV [V ferme] [SN la porte]]
SV \Rightarrow V SN SP : [SV [V ferme] [SN la porte] [SP avec la clé]]
SV \Rightarrow V P : [SV [V ferme] [SN la porte] [P que le vent avait fermée]]

Un point pose problème par rapport aux affirmations de Chomsky, en particulier sur le fait qu'une théorie ne doit pas pouvoir générer de phrases agrammaticales. Considérons que :

- (k) P \Rightarrow SN SV
SV \Rightarrow V NP

En insérant des items lexicaux à la place des catégories lexicales, on peut d'après (k) générer la phrase grammaticale

(l) Mon frère partage le pain.

et la phrase agrammaticale

(m) Mon frère pense le pain.

Les deux phrases sont toutes deux représentées par la schématisation (k) ce qui pose un problème théorique et montre les insuffisances du modèle dans cet état.

2.5. De l'analyse en constituants à la théorie X-barre

Nous avons vu qu'un syntagme nominal nécessitait la présence d'un nom, un syntagme verbal d'un verbe, etc. et que les informations (comme le nombre, le temps) propres à l'item lexical nécessaire sont celles de l'ensemble du syntagme et, de ce fait, sont transmises aux nœuds supérieurs. Cet élément lexical indispensable est le noyau du syntagme. Dans le syntagme verbal « *en naviguant vers l'Algérie pendant que la tempête faisait rage* », la tête est *naviguer* et le sens premier du syntagme se réfère à la navigation et non à la tempête. En quelque sorte, le propos du syntagme, c'est le propos du noyau.

Un deuxième aspect important est le principe qui permet au syntagme de « référer non pas à des choses ou à des actions uniques au monde, mais à des ensembles d'acteurs qui agissent en interaction d'une manière particulière, chacun avec un rôle spécifique » (Pinker, 1999). Ainsi, dans la phrase

(n) Georges joue au tennis avec Benoît.

jouer implique outre un joueur (Georges), un jeu (le tennis) et un partenaire (Benoît). Ces trois entités sont nommées des actants. Un syntagme nominal peut aussi rattacher des rôles aux actants : le bateau *à voile*, des accords *avec lui*. Le noyau et ses actants seront assemblés dans un sous-syntagme du syntagme nominal ou verbal ; le symbole utilisé sera le N-barre et le V-barre :

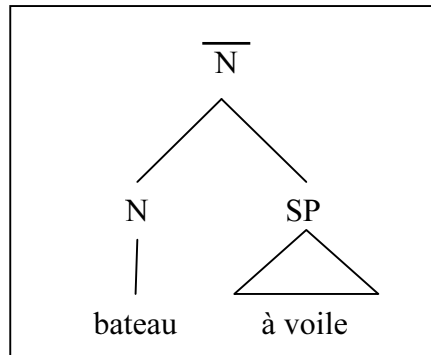


Figure 7 — Schéma d'un N-Barre

Un élément supplémentaire peut s'adjoindre à un syntagme : le (ou les) modifieur(s). Par exemple, on peut rajouter au syntagme précédent un modifieur *de Benoît* pour construire *bateau à voile de Benoît*. La caractéristique principale du bateau, son essence même, c'est d'être à voile. Le bateau peut être vendu, il n'en restera pas moins à voile. Le fait d'appartenir à Benoît n'est qu'une information secondaire dans la description du bateau. Cette différenciation entre actants et modifieurs se retrouve dans la géométrie de l'arbre : l'actant reste près du noyau, à l'intérieur du N-barre, pendant que le modifieur se place à l'étage supérieur, tout en restant dans le syntagme nominal.

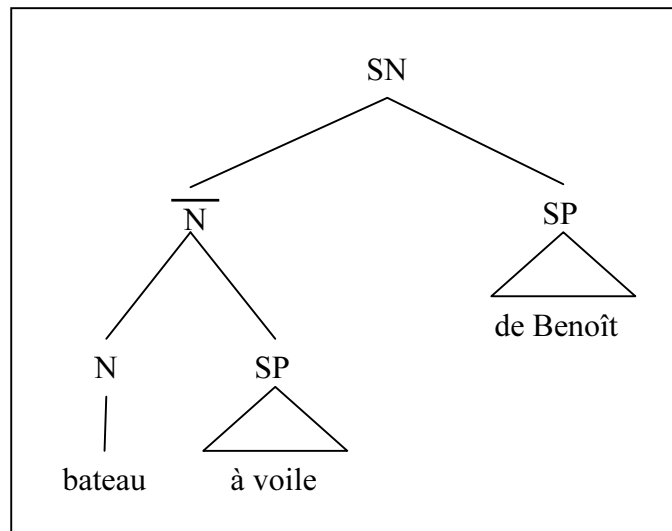


Figure 8 — Schéma d'un N-Barre avec actant et modifieur

Le même raisonnement s'applique aussi pour les syntagmes verbaux. Le syntagme verbal de la phrase (n) peut être symbolisé de la manière suivante :

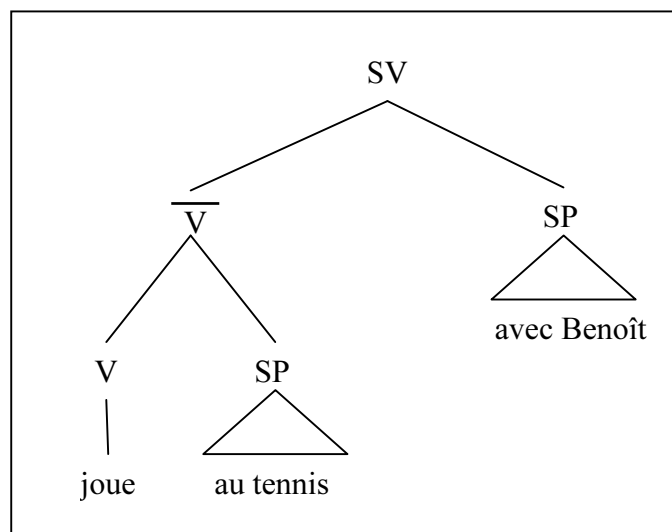


Figure 9 — V-Barre avec actant et modifieur

Bien entendu, on peut étendre ces représentations à tous les types de syntagme : par exemple pour un syntagme adjectival tel que *très fier de ses exploits*, on aura SA (Adv A' (A SP))¹⁴.

¹⁴ Pour des raisons de commodités, on note souvent le A-barre par A', le N-barre par N', etc.

Venons-en maintenant à la notion un peu plus complexe de sujet du syntagme. On considère que le sujet est un actant spécial dans la mesure où c'est l'agent de la cause, s'il en existe. Par exemple, dans la phrase

(o) Les malfaiteurs ont pillé la banque.

ce sont les malfaiteurs qui ont causé l'action. De ce fait, on peut adjoindre une fonction sujet au syntagme verbal *ont pillé la banque*. Ces sujets des syntagmes sont appelés spécifieurs. De la même manière que pour le syntagme nominal *le pillage de la banque par les malfaiteurs*.

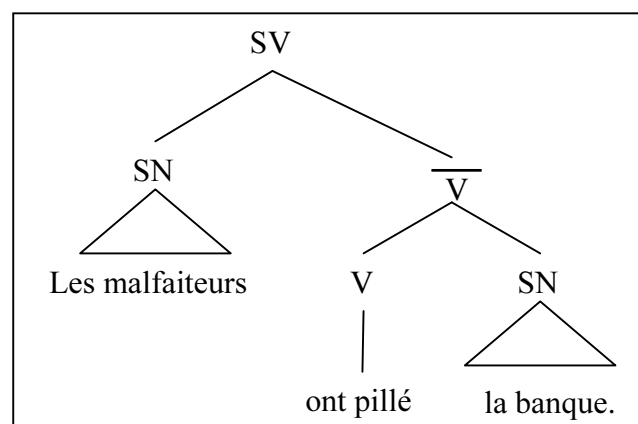


Figure 10 — Spécifieur dans un syntagme verbal

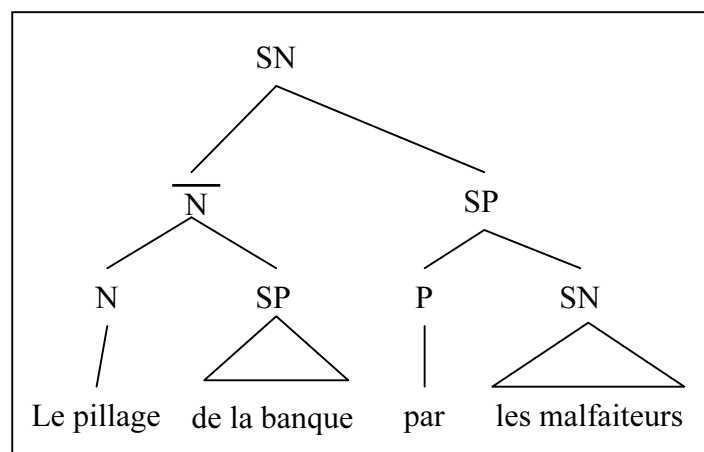


Figure 11 — Spécifieur dans un syntagme nominal

Pour nous résumer, l'ensemble des syntagmes partage un certain nombre de caractéristiques. Ils possèdent :

1. un noyau qui donne son nom au syntagme et qui précise de quoi il est question dans le syntagme.
2. des actants qui sont groupés avec le noyau dans un sous-syntagme
3. des modifieurs qui apparaissent à l'extérieur du N-barre, du V-barre, du P-barre ou du A-barre.
4. un spécifieur.

En conséquence, si on considère que le type de syntagme devient une variable que l'on nommera X, les règles de fonctionnement étant les mêmes pour tous les syntagmes, on se retrouve avec deux super-règles que l'on pourra décliner sur tous les types de noyau.

1. Le syntagme X est donc constitué d'un sujet facultatif, suivi d'un X-barre suivi de n'importe quel type de modifieurs.
2. Un X-barre est constitué d'un mot noyau, suivi d'un nombre quelconque d'actants. (Pinker, 1999)

Il suffira donc d'associer, dans tout syntagme X, un nom, un verbe, un adjectif ou une préposition aux noyaux, actants et modifieurs pour illustrer complètement les règles de formation syntaxique d'une langue. Ceci représente à l'intérieur de la théorie X-barre, les super-règles. Qui plus est, en français, le noyau d'un syntagme est placé avant ses actants : par exemple, on note la présence du verbe avant son complément d'objet, de la préposition avant le complément qu'elle introduit. Il y a donc un ordre de la position du sujet dans la langue, soit à gauche en français ou en anglais, soit à droite par exemple en japonais. La représentation arborescente d'un syntagme présentera donc en français son noyau à gauche et ses compléments à droite. Cette notion d'ordonnement des syntagmes s'appelle un paramètre et est spécifique à une langue.

Si cette caractérisation X-barre permet de décrire l'ensemble des syntagmes potentiels d'une langue, qu'en est-il d'une représentation concernant les phrases ? Si on comprend bien que le verbe est la tête du syntagme verbal, la question de savoir quelle est la tête d'une phrase reste entière. Pour répondre à

cette question, il faut ajouter que les phrases comportent des flexions, des genres, des modes, lesquels sont associés au verbe de son syntagme verbal. Dans la phrase :

(p) Le maçon va réparer bientôt le mur.

L'auxiliaire *va* indique le temps du syntagme verbal principal de la phrase (l'action est dans le futur). Ce que les linguistes considèrent, c'est que cet aspect flexionnel sera la tête de la phrase, car il affecte tout son fonctionnement. Cette marque flexionnelle aurait aussi pu être portée par la seule marque du futur dans la phrase (Le maçon réparera...).

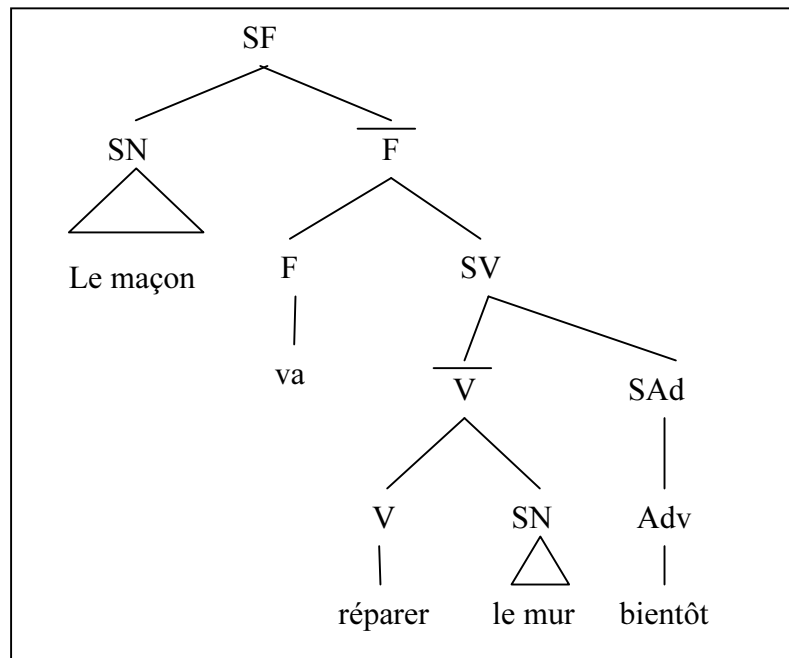


Figure 12 — Exemple de syntagme flexionnel

On voit que le syntagme nominal *Le maçon* est le sujet ou le spécifieur du syntagme flexionnel dont le noyau est *va* et le complément est le syntagme verbal *réparer le mur bientôt*. De nouveau, à l'intérieur de la structure intermédiaire V-barre, on retrouve le noyau *réparer* avec un complément qui est le syntagme nominal *le mur*. Notons que nous pouvons également représenter les deux syntagmes nominaux à l'aide d'une notation N-barre :

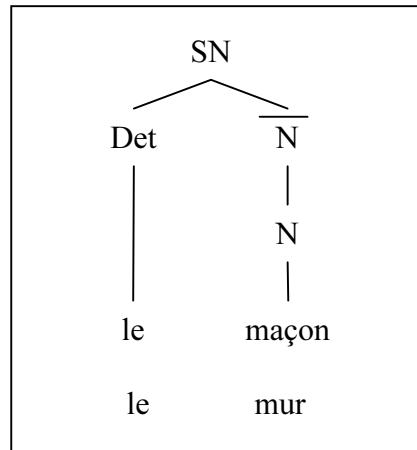


Figure 13 — Syntagme nominal représenté à l'aide de la notation N-Barre

Cette flexion qui devient noyau de la phrase fait partie de ce que Pinker appelle des mots de fonction. Ces mots de fonction appartiennent à une catégorie particulière du lexique : ce sont les articles, les pronoms, les prépositions et les conjonctions. Pour Pinker, ce sont des « fragments de grammaire cristallisée » dans la mesure où ils commandent et déterminent des syntagmes plus grands dans lesquels s'insèrent les syntagmes nominaux, verbaux et adjectivaux. Pinker considère qu'ils fournissent un échafaudage pour la phrase. Ce sont les mots les plus sûrs pour échafauder la structure syntagmatique de la phrase. C'est ce qui explique qu'ils appartiennent à des catégories dites fermées, où aucun élément nouveau ne peut s'immiscer, à la différence des mots non fonctionnels qui intègrent chaque jour de nouveaux membres.

Pour terminer avec la représentation des phrases entières, il convient encore de préciser le rôle de ce qu'on appelle les complémenteurs qui ont la capacité de lier deux syntagmes flexionnels : ce seront donc également des mots fonctionnels comme les conjonctions ou les pronoms relatifs. Ainsi pour la phrase

(q) Le maçon va réparer le mur quand la famille trouvera l'argent.

on aura le graphe affiché par la figure 14. On remarque que le syntagme complémenteur intermédiaire C-barre, comme pour tous les autres niveaux intermédiaires X-barre domine son noyau, le complémenteur C (quand) et un nouveau syntagme flexionnel (« *la famille trouvera l'argent* »). Le C-barre

forme avec son syntagme flexionnel un constituant. On peut dire que C sélectionne une inflexion ou une proposition comme le verbe sélectionne le type de ses compléments. Le complémenteur C est donc bien le noyau d'une projection fonctionnelle qui prend comme complément la projection maximale d'un syntagme flexionnel.

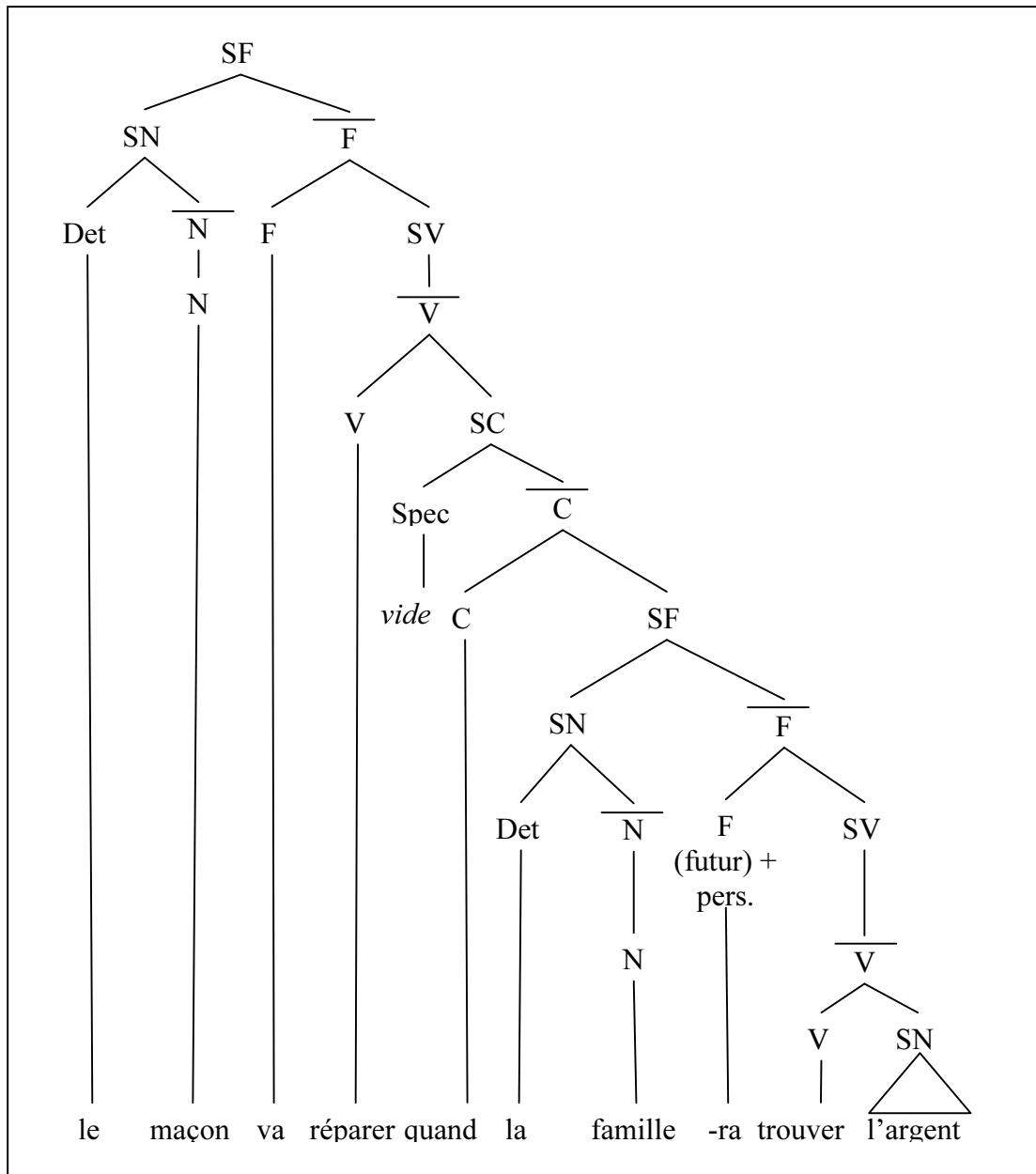


Figure 14 — Représentation syntaxique de la phrase (q)

En conséquence, on peut conclure que tout syntagme, qu'il soit nominal, verbal, adjectival, prépositionnel, mais aussi flexionnel (la phrase) ou encore complémenteur répond à une structure identique : une projection maximale SX, une ou plusieurs projection(s) intermédiaire(s) X-barre, un noyau X, un actant du noyau (SP) qui est au même niveau que le noyau, éventuellement des

modifieurs (SY) du noyau, et bien sûr des spécifieurs. La figure 15 représente le schéma X-barre complet, s'appliquant à tout type de syntagme.

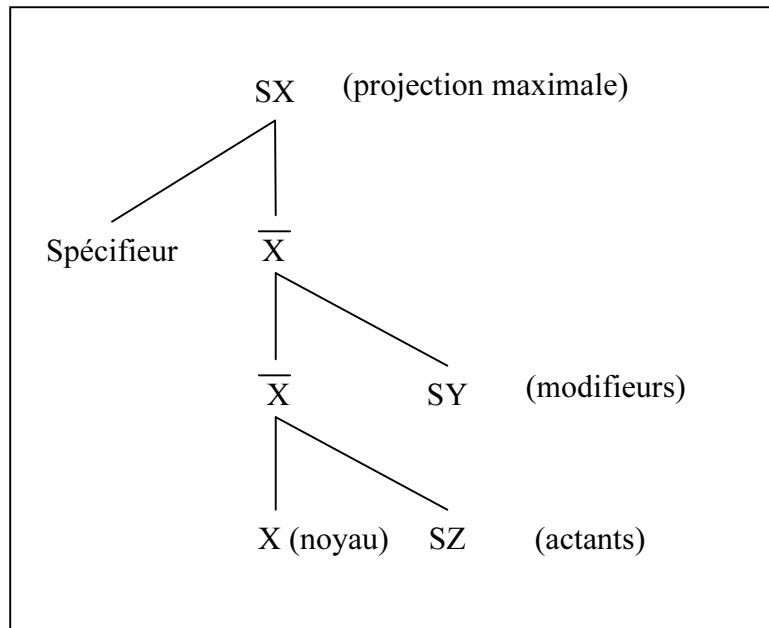
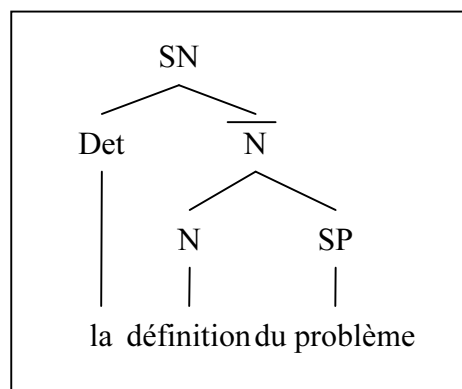


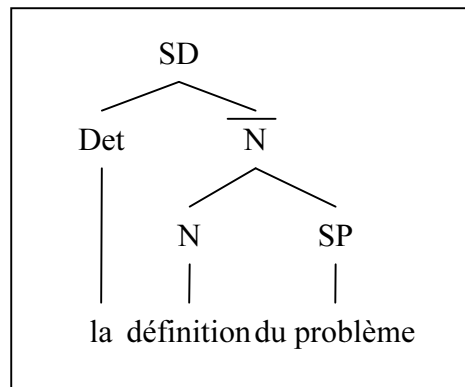
Figure 15 — Schéma X-Barre pour tout type de syntagme

Pour terminer, nous voudrions revenir sur un point. On a vu, et la dernière figure l'illustre encore une fois, que le noyau d'un syntagme est toujours situé à gauche de ses actants. Cependant, dans les syntagmes nominaux, on a le plus souvent le déterminant à gauche du noyau, comme on peut le voir dans la figure 13 du présent chapitre. On considère alors que le déterminant est le spécifieur du syntagme. Dans cette première acception, on a pour le syntagme « *la définition du problème* », la représentation suivante :



Cependant, Shapiro note que de nombreux linguistes ont pris le parti de reconsidérer ce type de projection (Shapiro, 1997). En effet, il est possible de

réanalyser le syntagme précédent en un syntagme fonctionnel ayant comme noyau un déterminant. Sa représentation donne ceci :



Dans ce cas, le constituant *la définition du problème* n'est plus la projection de la tête lexicale nominale *définition*, mais celle de la tête fonctionnelle constituée par le déterminant *la* qui prend pour complément la projection maximale N-barre *définition du problème* (Pollock, 1997). De ce fait, tous les mots fonctionnels dont nous avons déjà rappelé plus haut les caractéristiques de classe fermée se retrouvent têtes de syntagme.

2.6. Relations structurelles

Deux relations formelles sont appliquées aux nœuds d'un arbre.

1. On dit d'un nœud X qu'il *commande* un autre nœud Y si et seulement si le premier nœud qui domine X, domine également Y. Dans la phrase (q), le syntagme nominal (le maçon) domine le nœud F-barre puisque le nœud SF qui domine le SN domine également F-barre. De même, le complémenteur *quand* commande le deuxième syntagme flexionnel (*la famille...*) puisqu'ils sont tous deux dominés par un C-barre.
2. On dit d'un nœud X qu'il *gouverne* un nœud Y si et seulement si :
 - X commande Y
 - X et Y sont dans la même projection maximale
 - Il n'y a pas de projection maximale entre X et Y

2.7. Syntaxe et lexique

D'après la théorie, on devrait donc ne pouvoir construire que des phrases correctes à partir de l'application des super-règles que nous venons de préciser. En particulier, il suffirait de remplir la catégorie lexicale *verbe* du syntagme verbal par un de ses représentants pour produire une phrase. Malheureusement, les phrases générées sans autre contrainte peuvent être surprenantes, puisque les phrases :

(r) Mon frère embrassa.

(s) Mon frère mit la voiture.

respectent les règles édictées mais ne sont pas vraiment acceptables. On admettra alors que notre lexique interne (ou dictionnaire mental) contient des informations sur chacune de ses entrées qui interagissent avec les super-règles syntaxiques. Ainsi, pour chacune des entrées lexicales, on aurait des renseignements sur sa forme phonologique, sa forme graphique, sa catégorie lexicale, sur les informations sémantiques qu'il transporte et surtout sur l'environnement syntaxique dans lequel il doit s'intégrer. En conséquence, si les phrases (r) et (s) sont incorrectes, c'est parce qu'il n'y a pas correspondance entre les informations contextuelles affectées à chacun des verbes dans notre lexique interne et la réalité des contextes dans lesquels ils sont intégrés. Les linguistes considèrent que les verbes sont cruciaux dans la représentation syntaxique de la phrase. Chaque verbe possède deux grandes informations de bonne construction syntaxique : la sous-catégorisation et la grille thématique. Cette deuxième notion est partagée par les autres natures lexicales.

2.7.1. La sous-catégorisation

Il s'agit du type d'environnement dans lequel le verbe peut apparaître. La phrase (r) est incorrecte parce qu'il manque un syntagme nominal (ici complément d'objet) qui indique qui ou quoi embrasse *mon frère*. La sous-catégorisation du verbe embrasser n'est pas satisfaite dans la phrase (r). Le même raisonnement s'applique pour la phrase (s), où le verbe *mettre* demande à la fois un syntagme nominal (*la voiture*) et un syntagme prépositionnel (ici absent). On peut schématiser les entrées lexicales de la manière suivante :

embrasser :

V (Verbe)
Signifie... (sens)
[__ SN] (sous-catégorisation)

mettre

V
[__ SN SP]

Un principe de projection lexicale est donc au cœur de la structuration syntaxique qui postule que les possibles sous-catégorisations sont projetées sur la syntaxe. Les phrases (r) et (s) seront rejetées en son nom parce qu'il n'y a pas accord entre la sous-catégorisation des verbes et la syntaxe proposée. Un verbe peut aussi sous-catégoriser une proposition entière :

(t) Mon frère découvre la montagne.

(u) Mon frère découvre que la montagne est dangereuse.

(t) et (u) montrent que le verbe *découvrir* exige soit un syntagme nominal (objet), soit un syntagme complémenteur SP.

2.7.2. La grille thématique d'un item lexical

Une autre information qui accompagne les entrées du lexique mental concernerait sa grille thématique, c'est à dire le nombre et la fonction (ou rôle) thématique de ses arguments (Pollock, 1997). Le nombre des arguments est proche de la notion de sous-catégorisation que nous venons de voir. Les rôles thématiques sont généralement définis comme se référant aux notions d'agents, de bénéficiaires, de lieux (psychologiques), de sources, de buts, de thèmes ou de patients. Ainsi, non seulement on sait que le verbe embrasser demande deux arguments, mais l'analyse de sa grille permet de savoir qu'il s'agit de l'agent et du patient. Le verbe acheter peut avoir un agent, un thème et une source ([Agent]Mon frère [V]achète [thème])une voiture [Source]au garagiste), mais la source pourrait être remplacée par un but ou un destinataire ([Agent]Mon frère [V]achète [Thème]une voiture [But]pour sa fille). De plus, une propriété thématique peut être marquée comme indispensable ou facultative. L'ensemble de ces critères oblige à ce que les propriétés thématiques obligatoires spécifiées dans les entrées lexicales soient retrouvées une à une dans les arguments présents dans la phrase.

2.8. Conclusion

Pour commencer cette conclusion, nous voudrions rappeler que cette présentation ne prétend pas être un résumé de l'analyse syntaxique moderne. Majoritairement issue des travaux de l'école de Chomsky, elle a beaucoup évolué entre les premiers écrits et les développements actuels. Par exemple, nous n'avons pas mentionné les notions de structures profondes, de transformations et de structures de surface. Elles ont toujours été très controversées et Chomsky lui-même semble y renoncer dans ces derniers travaux (Bronckart, 1977; Pinker, 1999). Nous ne parlerons pas non plus de l'innéité des mécanismes humains nous permettant le langage¹⁵.

La seule question qui nous intéresse ici, c'est de comprendre quels liens on peut observer entre cette analyse linguistique et le modèle structural de lecture tel que Koriat l'a énoncé. De fait, deux grands éléments sont pris en considération lors d'une analyse linguistique : d'une part l'analyse en constituants et d'autre part l'analyse de la phrase.

1. Nous avons essayé de décrire l'analyse en constituants. Ces constituants sont soit des projections maximales de tête lexicale (Adjectif, Verbe et Nom), soit des projections maximales de tête fonctionnelle.
2. L'analyse de la structure syntaxique du niveau supérieur, la phrase, montre de grands parallèles avec l'analyse en constituants. La tête de la phrase est cette fois une catégorie fonctionnelle 'inflexion' qui est le lieu de réalisation des marques de mode, de temps et d'accord. Ce syntagme inflexionnel peut lui-même être le complément d'une tête fonctionnelle complémentateur.

L'intérêt de ces deux grands éléments pour le modèle structural, c'est qu'ils rendent compte des deux types d'observations sur lesquels s'est élaboré le modèle structural : d'une part, celle que Koriat nomme un peu trivialement *de taille réduite* (Koriat & Greenberg, 1994), et qui recouvre en fait le découpage en constituants : les mots observés dans les différentes recherches sur le

¹⁵ Très schématiquement, Chomsky propose que les super-règles sont innées et universelles, tandis que les paramètres sont construits par l'apprentissage. Pour de plus amples développements, on pourra voir (Pinker, 1999) et surtout (Piatelli-Palmarini, 1982).

barrage de lettres étant principalement des déterminants ou des prépositions¹⁶, les syntagmes observés sont soit nominaux, soit prépositionnels. D'autre part, les autres observations se situant au niveau du syntagme inflexionnel sont, nous l'avons vu, très peu nombreuses, et elles se cantonnent aux travaux de Musseler et Koriat sur la langue allemande (Musseler et al., 2000).

Nous voudrions également insister encore une fois sur l'importance des mots fonctionnels dans la linguistique chomskienne. Tous ces déterminants (articles, possessifs) et complémenteurs (conjonctions, pronoms relatifs) appartiennent à une classe fermée, dans la mesure où cette liste ne change pas, où aucun nouvel élément ne peut s'immiscer sans faire vaciller l'ensemble de l'édifice. Cependant, si la classe des déterminants qui introduit, sans conteste, les syntagmes nominaux a été à l'origine des observations sur le modèle structural, la catégorie des complémenteurs, elle, n'a jamais fait l'objet de recherches systématiques, ce qui explique en partie les propos initiaux de Koriat sur la taille réduite des structures construites par le lecteur.

Les expérimentations que nous relaterons au chapitre 8 s'intéresseront à ces "lacunes", en réalisant des observations sur les différentes catégories, pour chercher, d'une autre façon que les travaux sur la prosodie, les traces des différentes catégories de structures construites : celles s'organisant autour d'un syntagme nominal et celles s'organisant autour de la phrase.

Il faut également noter qu'à l'intérieur du syntagme flexionnel, on observe la flexion en elle-même, qui en est la tête, et son complément. Les marques du noyau flexionnel peuvent se décliner soit sous la forme d'un auxiliaire, soit sous la forme des marques de temps, d'accord ou de mode directement sur le verbe. En tant que tête, on peut supposer que la flexion commande à son complément et, en conséquence, pourrait être également projetée à l'arrière plan comme le propose la théorie de Koriat. Cependant, si Koriat a étudié l'oubli de lettres sur certains morphèmes enchâssés dans des mots (Koriat et al., 1991), il ne considérait que des morphèmes prépositionnels ou

¹⁶ En grande partie, en raison de leur très forte fréquence, qui permettait aux auteurs, en fonction des contextes dans lesquels étaient placés les mots-cibles de différencier les effets de fréquence, de structure ou de vision.

déterminants. En d'autres termes, il n'a encore jamais été observé d'oubli de lettres sur les morphèmes reflétant la tête de syntagme flexionnel, mais tout porte à croire que cette recherche, si les résultats étaient satisfaisants, renforcerait encore le modèle de Koriat.

Enfin, les sous-catégorisations et autres contraintes, dues aux grilles thématiques propres aux mots qui en possèdent, obligent le lecteur à les garder en mémoire pour vérifier que la phrase respecte bien ces contraintes. En conséquence, tous les mots possédant une sous-catégorisation ou une grille thématique ne devraient pas voir un oubli de lettres important, puisqu'ils ne peuvent pas être rejetés à l'arrière-plan.

Nous pouvons tenter d'illustrer ces deux hypothèses en reprenant la phrase (q) : *Le maçon va réparer le mur quand la famille trouvera l'argent*, dont la figure 14 présente le graphe syntagmatique. Nous pensons que tous les noyaux des structures X-barres doivent présenter plus d'oubli de lettres, sauf les noyaux des structures V-barres puisque leurs grilles thématiques et leurs sous-catégorisations les obligent à rester présents en mémoire (les derniers développements théoriques considèrent que les syntagmes nominaux possèdent une tête fonctionnelle). Ainsi, si on reprend l'arbre de la phrase, les mots ou morphèmes (dans l'ordre) *Le, va, réparer, quand, la, -ra, trouver, l* devraient avoir un taux d'oubli plus important mais on doit retirer de cette liste les deux verbes (*réparer* et *trouver*) à cause de leur grille thématique.

D'après les premières versions du modèle de Koriat, où seules des structures de taille réduite sont construites, on devrait trouver plus d'oublis de lettres seulement dans les mots fonctionnels *le* (qui correspond au déterminant du syntagme nominal *le maçon*), *la* (*la famille*) et *l* (*l'argent*). Comme la taille des structures s'est possiblement rallongée, on rajoutera le complémenteur *quand* dans la liste des mots dont les lettres seront probablement oubliées. Pour ce qui est des morphèmes flexionnels (*va* et *-ra*) on peut émettre l'hypothèse d'un fort taux d'oublis, dans la mesure où ils sont têtes de syntagmes. Il ne nous semble pas que ce type de recherche ait déjà été menée.

Nous tenterons d'apporter une réponse pour ce qui concerne spécifiquement les complémenteurs. Sont-ils oubliés comme le sont les déterminants ou les prépositions ? Dans l'affirmative, l'hypothèse selon laquelle le lecteur élabore des squelettes syntaxiques de l'ensemble de la phrase et non plus seulement des syntagmes nominaux ou prépositionnels s'en trouverait confortée.

Chapitre 3.

La vision et la lecture

3.1. Introduction

Les travaux que nous avons relatés précédemment, issus pourtant de différents domaines de recherches, ont en commun de reposer sur des expériences susceptibles de poser des questions de méthodologie. En particulier pour les expérimentations contrôlant les mouvements des yeux. Nous avons rappelé au cours des chapitres précédents la remarque de Rayner qui déclare que toutes les manipulations qui tentent de supprimer les processus travaillant à partir des informations parafovéales ou qui obligent les lecteurs à passer par un point de fixation particulier modifient les processus normaux de lecture (Rayner & Pollatsek, 1989).

Aussi, dans cette partie, nous présenterons des informations qui éclairent les processus de la vision, dans leurs deux composantes : vision fovéale et vision parafovéale.

3.2. Vision nette et vision dégradée

Chez l'Homme, la fovéa, petite dépression d'environ 1,2 mm de diamètre située au centre de la tache jaune ou macula, se caractérise par une grande finesse de son pouvoir de discrimination spatiale, ainsi que par son importance pour la vision des couleurs. Elle contient exclusivement des cônes, tous les éléments non récepteurs ainsi que les vaisseaux sanguins étant refoulés à sa périphérie, ce qui a l'avantage de laisser la lumière frapper directement les récepteurs. Leur densité est très élevée. Chaque cône est connecté à une cellule bipolaire. Au fur et à mesure qu'on s'éloigne de la fovéa, les cônes deviennent plus rares ; les bâtonnets progressivement prédominent. Ici, les prolongements dendritiques des cellules bipolaires font contact avec un nombre important de bâtonnets.

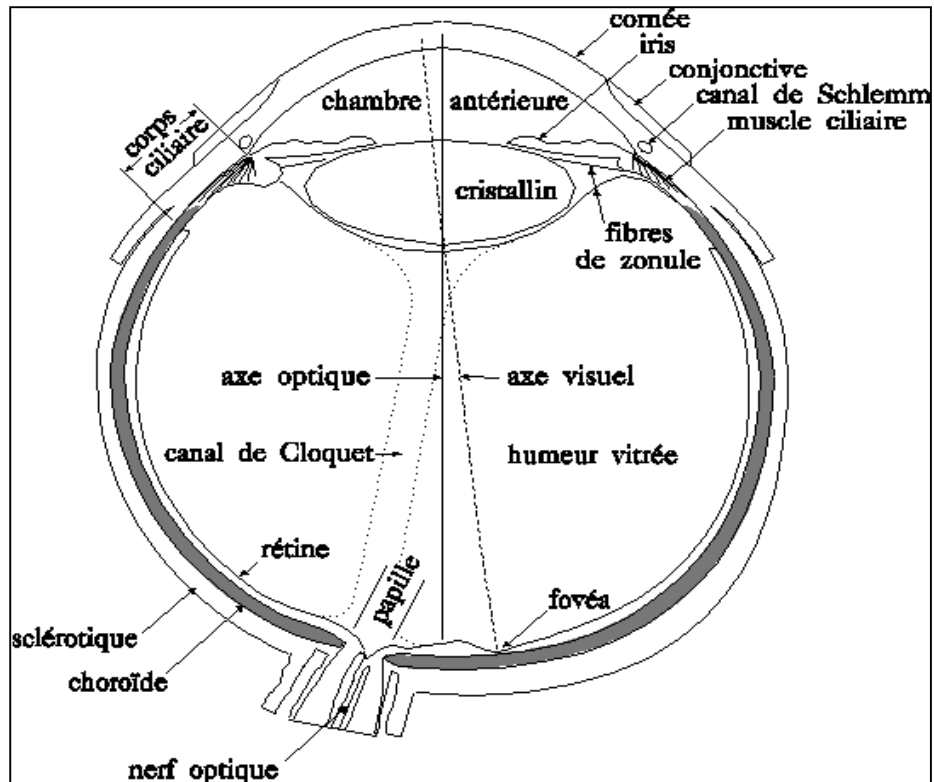


Figure 16 — Schéma de l'œil

La vision fovéale, donc la vision la plus nette, couvre environ 1° (degré) du champ visuel chez l'Homme. Le champ visuel, incluant la vision périphérique, totalise près de 160°. La dégradation de la netteté se fait progressivement, au rythme d'environ 50 % de dégradation par degré d'écartement.

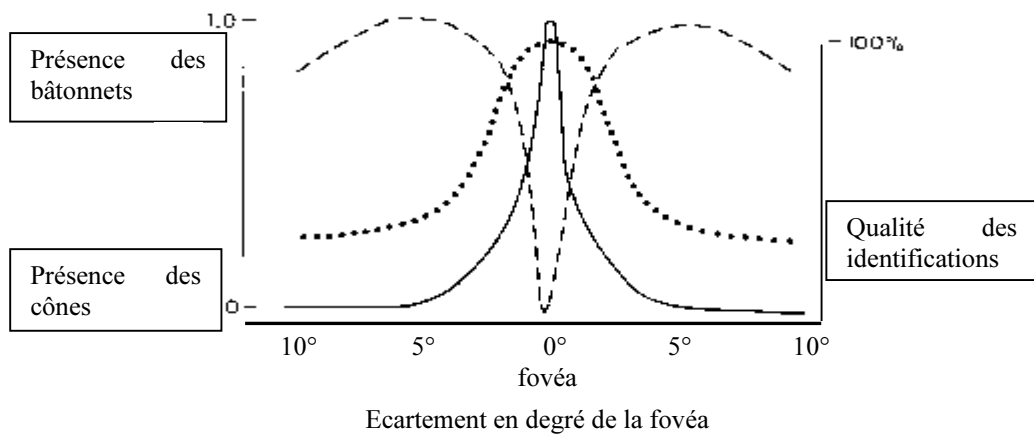


Figure 17 — Densité relative des cônes (ligne continue) et des bâtonnets (ligne en tirets), et la fidélité d'identification d'un mot-cible à une distance variable du centre de la rétine (ligne pointillée). (Rayner & Pollatsek, 1989)

Chez un lecteur expert, la taille de l'empan de lecture peut atteindre 10°. En conséquence, ce lecteur traite des informations très dégradées en bord d'empan, informations provenant de régions plus périphériques. Ces régions sont communément appelées parafovéales. Elles jouent sans conteste un rôle à chaque fixation de l'œil¹⁷.

3.3. La vision parafovéale et la lecture

Quelques équipes de recherche se sont particulièrement spécialisées dans l'étude de la prise parafovéale d'informations lors de la lecture. La question est ici d'étudier le rôle du mot situé à droite (donc dans la zone parafovéale) de celui fixé (qui se trouve, lui, dans la zone fovéale). Depuis 1981 et les travaux de Rayner, on sait que, si on retire des informations (par des systèmes de masques) de la zone parafovéale, la vitesse de lecture décline d'environ un tiers. Cette réduction de vitesse est due à l'accroissement des temps de fixation et à des saccades oculaires moins amples, donc à un plus grand nombre de fixations. Les informations sur les mots, fournies par la vision parafovéale, facilitent donc l'identification du mot lorsqu'il entre dans la zone fovéale (Inhoff & Rayner, 1986). On peut se demander par quels mécanismes :

- D'après Inhoff, il paraît improbable que la vue parafovéale active directement une attente sémantique qui accélérerait la reconnaissance du mot (Inhoff, 1982). Pour preuve, les bénéfices habituellement observés dus à la vision parafovéale disparaissent si, lors du déplacement de l'œil sur le mot suivant, on remplace celui-ci par un mot sémantiquement proche. Rappelons

¹⁷ Certains travaux plus anciens ((Rayner, 1984), cité par (Zagar, 1992)) ont séparé la vision en trois zones bien différenciées :

1. La zone fovéale.
2. La zone parafovéale, plus excentrée (6 à 12 caractères), où se traite principalement l'information donnée par les premières lettres du mot suivant.
3. La zone périphérique, encore plus distante (jusqu'à 20 caractères), dont le rôle se situe plus dans l'extraction de l'information sur la taille des mots pour la programmation des futures saccades.

Cependant, cette division en classes étanches ne semble pas correspondre à la réalité de la perception visuelle. Si elle perd en acuité au fur et à mesure de l'éloignement entre le lieu de réception et la fovéa, c'est parce que progressivement la nature des cellules change : on passe des cônes aux bâtonnets. Il nous semble que la figure 16 illustre bien l'aspect continu du phénomène. Dans la suite de ce travail, nous emploierons le terme parafovéal pour parler des informations provenant des zones comprises entre les bords de la fovéa jusqu'aux parties les plus périphériques de l'œil, en gardant bien à l'esprit que la dégradation du signal est progressive et continue.

que l'œil ne prend pas d'information pendant son déplacement (Berthoz & Petit, 1996), ce qui rend "invisible" la substitution.

- On a pu croire que la reconnaissance des mots se faisait de manière séquentielle, un mot après l'autre, ce modèle étant proposé par Taft & Forster. Il s'agirait ici d'un accès dans un fichier qui contiendrait des représentations des radicaux, classées par fréquence. La vision parafovéale aurait donc permis de pré-activer la bonne entrée dans ce « fichier », grâce à la vue anticipée du radical du mot suivant (Taft & Forster, 1976). Inhoff montra que cette hypothèse était fautive. Il rejeta aussi, l'hypothèse de l'influence de la première syllabe comme activation du mot à trouver. «La reconnaissance d'un mot lors de sa fixation n'est pas facilitée par la vision parafovéale antérieure de ses premières lettres, qu'elles forment une syllabe ou non. » (Inhoff, 1989).

- Des expérimentations plus récentes continuent dans la même veine. On notera en particulier l'article de Kennedy présentant deux nouvelles expériences pour étudier le degré d'influence temporelle d'un mot présenté parafovéalement sur le traitement du mot situé centralement (Kennedy, 2000). Ce que cherche à savoir cet auteur, c'est bien si on peut observer le temps pris par le traitement du mot situé en vision périphérique, en fonction de ses caractéristiques propres et abstraction faite du temps de traitement du mot fixé. Dans ces deux expériences, les sujets voient un ensemble de mots de cinq lettres pendant que les mots en vision parafovéale varient sur des critères de longueur, de fréquence et en fonction du type et de la fréquence des trois premières lettres du mot perçu en vision périphérique. Les résultats sont similaires à ceux obtenus par Inhoff onze ans plus tôt : ces auteurs considèrent que le début et la fin des mots contribuent à la reconnaissance périphérique des mots et que les bénéfices dus à la vision parafovéale sont fonction de la totalité de la morphologie du mot. En effet, on obtient de plus grands bénéfices en tenant compte des informations données par l'ensemble du mot périphérique qu'en cumulant les données issues du premier trigramme et de la fin du mot¹⁸.

¹⁸ Ces découvertes se rapprochent de l'effet-mot, qui fait qu'un mot est mieux perçu que l'ensemble des lettres qui le composent (Reichter, 1969).

- Ce type d'observation renforce les modèles de reconnaissance logographique, développés en leur temps par Adams (Adams, 1979) et McClelland (McClelland & Rumelhart, 1981). Ces modèles semblent offrir un point de départ pour la compréhension des bénéfices de la vision parafovéale dans la lecture normale. Par exemple, « toutes les lettres (ou les silhouettes des lettres) d'un mot accessible par la vue périphérique peuvent activer des représentations de lettres (et de possibles représentations lexicales) » (Inhoff, 1989). Ces modèles logographiques sont aujourd'hui très décriés, surtout par les modèles développementaux de la lecture. On les considère comme des phénomènes très primitifs (« ...les modalités primitives (logographiques) de traitement de l'information graphique ») (Sprenger-Charolles, 1992), appelés à être remplacés par des modalités grapho-phonologiques ou alphabétiques (Frith, 1986; Gombert, 1996). On associe, souvent de manière caricaturale, cette capacité logographique à certains types d'enseignement « qui prennent la lecture pour une tâche de mémorisation visuelle » (Stuart & Coltheart, 1988). Pourtant, elles semblent nécessairement à l'œuvre dans la reconnaissance par la vue périphérique, capacité « mathématiquement » indissociable d'une lecture experte.

Enfin, les recherches menées par Kennedy lui font considérer que les effets des zones parafovéales sur les zones fovéales, observés dans les différents travaux expérimentaux, sont incompatibles avec les modèles de lecture dans lesquels l'attention se déplace séquentiellement d'un mot à l'autre. Les données obtenues sont plus congruentes avec l'idée que les processus fovéaux et parafovéaux ont lieu en parallèle sur une région du texte plus large qu'un mot (Kennedy, 2000).

3.4. Quelques travaux en ophtalmologie

Le laboratoire de recherche sur la vision de l'Université du Minnesota a entrepris des études pour comprendre les conséquences, sur la lecture, de la perte ou de la diminution du champ visuel. Ces travaux retiennent l'attention, tant il apparaît que l'ensemble de l'œil participe à la lecture, et pas seulement

la fovéa. G. LEGGE et son équipe ont développé un modèle informatique d'observation et de simulation de la lecture, nommé M. Chips. Ils ont évalué les performances en lecture de M. Chips en fonction de la taille de l'empan visuel, c'est-à-dire du nombre (N) de caractères adjacents qui pouvaient être reconnus. La tâche de M. Chips était de se mouvoir dans un texte en un nombre minimum de saccades, tout en identifiant tous les mots. Deux sources d'informations étaient disponibles :

- Des données visuelles, N lettres identifiables clairement ; plus des données périphériques de moindre résolution, qui présentaient les lettres par des * et les espaces par des (_). Un peu comme des données de bas niveau...
- Des données lexicales, les mots possibles et leur fréquence. Un peu comme des données de haut niveau...

Les distributions des saccades oculaires furent mesurées pour des empan visuels variant de 1 à 11 lettres.

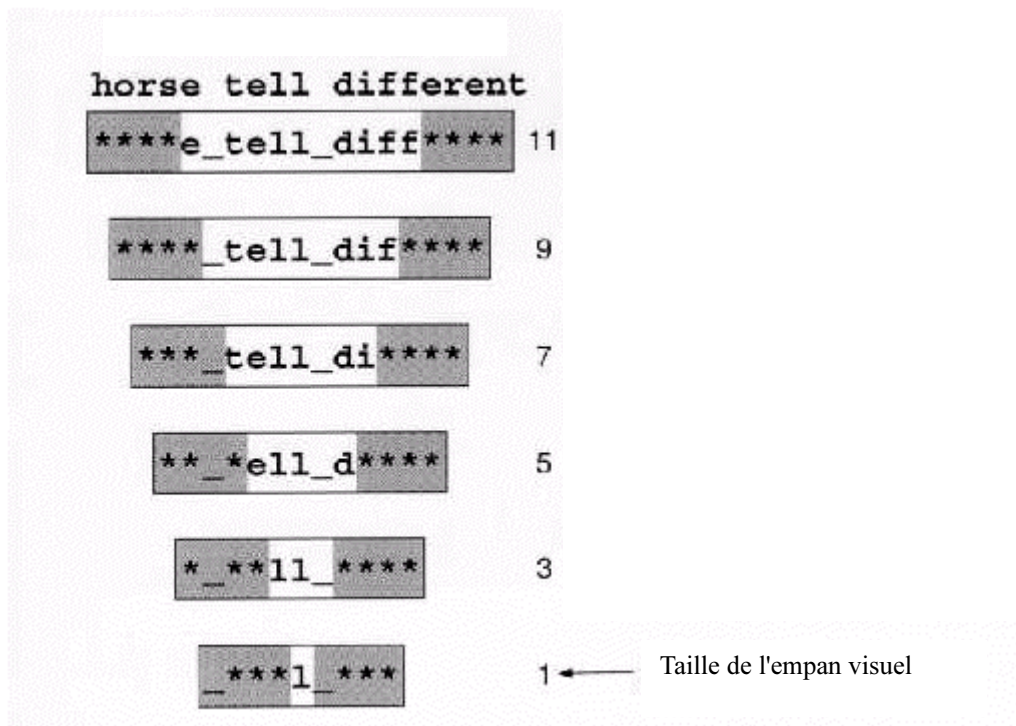


Figure 18 — Etude du champ visuel. Les croix représentent le masquage des lettres

Les deux schémas suivants montrent l'effet de la perte d'information due à une amputation d'une partie périphérique du champ visuel par comparaison avec un champ non perturbé.

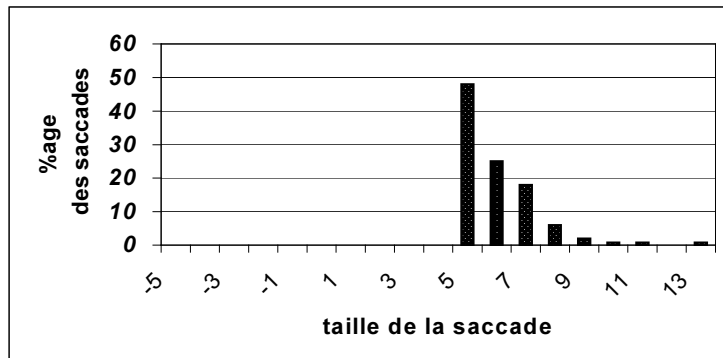


Figure 19 — Saccades oculaires sans perturbation du champ visuel

La figure 19 montre la distribution des saccades oculaires quand le lecteur n'est pas perturbé par des masques. On remarque qu'il n'y a pas de retour en arrière et que la taille des saccades atteint 13 signes

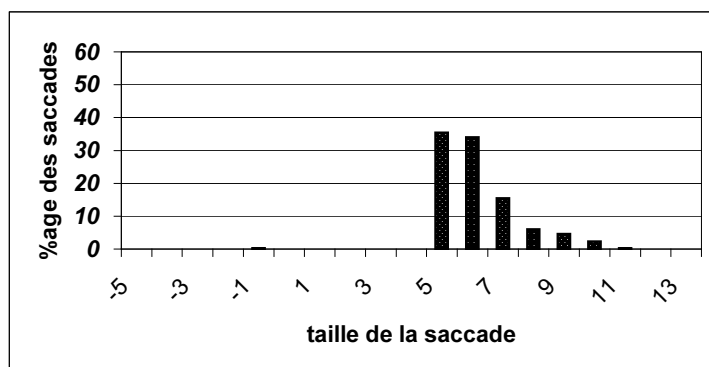


Figure 20 — Saccades oculaires après retrait du champ visuel gauche

En figure 20, on a retiré le champ gauche. On remarque une légère diminution de la taille moyenne des saccades.

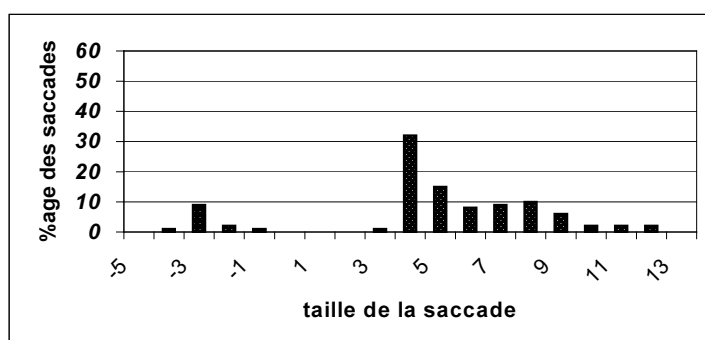


Figure 21 — Saccades oculaires après retrait du champ visuel droit

La suppression du champ périphérique droit cause une sévère diminution de la taille moyenne des saccades, un plus grand nombre de régressions et une plus grande variabilité de la taille moyenne des saccades.

Ainsi, si la vision d'un empan de texte se limite à la zone centrale du champ visuel, les performances diminuent sensiblement et les mouvements des yeux deviennent erratiques (on constate un grand éparpillement des saccades). Les informations données par les zones périphériques jouent donc un rôle déterminant dans la fluidité de la lecture.

3.5. La saccade oculaire

3.5.1. Généralités

Il est régulièrement fait état de l'étude des saccades oculaires dans la lecture. Mais ce phénomène fait aussi l'objet d'études physiologiques ; on l'observe dans les différentes aires corticales et dans le cerveau profond. Son étude est d'une trop grande complexité pour être véritablement traitée ici, néanmoins nous allons rappeler quelques bases.

Le mécanisme neuronal de production des saccades est contrôlé au moins autant par des neurones excitateurs que par des neurones inhibiteurs. Ainsi, « la décision de faire un mouvement des yeux, accompagné ou non d'un mouvement de la tête est prise à travers une hiérarchie de mécanismes parallèles excitateurs et inhibiteurs » (Berthoz & Petit, 1996). Il semble que l'évolution ait amené dans le cerveau des primates un bien plus grand nombre de mécanismes inhibiteurs, ce qui a été utilisé pour « bloquer l'exécution d'un mouvement, de façon à permettre la simulation interne des mouvements planifiés et à tester de façon endogène ses conséquences potentielles » (Berthoz & Petit, 1996). Ces inhibitions permettent de faire des prédictions et de « sélectionner les stratégies motrices ». De plus, quelques données montrent même, dans certaines conditions, une capacité du cerveau à anticiper déjà sur la saccade qui suivra la saccade à venir.

Ce qui est important, c'est cette capacité du cerveau à simuler les actions à effectuer, dont les saccades oculaires, et à les piloter en fonction du résultat qu'il pense en obtenir. On n'est plus du tout dans une conception mécanique de la saccade, comme une suite de mouvements servant à accumuler de l'information, on est bien dans un mouvement dont les conséquences ont déjà

été anticipées. La perception n'est plus seulement préparatoire à l'action, c'est d'abord l'action qui définit la perception dont elle a besoin. Berthoz a résumé ce nouveau modèle en écrivant que « le cerveau n'est pas une machine réactive, c'est une machine proactive qui projette sur le monde ses interrogations » (Berthoz & Petit, 1996).

3.5.2. Où fixer et quand partir ?

Un important domaine de recherche s'est développé depuis deux ou trois décennies concernant l'étude des mouvements des yeux chez les lecteurs experts (Rayner, 1998). La question qui sous-tend nombre de travaux est de distinguer, pour mieux en comprendre leur relation, le moment du déclenchement du mouvement des yeux et l'endroit où ils vont se poser. De nombreux travaux empiriques laissent envisager une relative indépendance entre le « *quand* » et le « *où* » (Rayner, Sereno, & Raney, 1996; Rayner, Fischer, & Pollatsek, 1998). La décision sur le moment auquel déclencher la saccade semble être principalement due à des contraintes cognitives, telles que la complexité du matériau linguistique à traiter pendant la fixation, alors que celle de l'endroit où la prochaine fixation doit se situer semble être influencée par des informations de bas-niveau, telles que l'environnement visuel perçu par la vision parafovéale.

Pour ce qui est de la durée de la fixation, deux grands phénomènes peuvent être signalés :

1. A taille égale, les lecteurs font des fixations plus longues sur les mots rares que sur les mots fréquents (Hyönä & Olson, 1995; Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Vitu, 1991).
2. Les mots contextuellement imprévisibles sont le siège de plus longues fixations que les mots anticipables (Ehrlich & Rayner, 1981; Balota, Pollatsek, & Rayner, 1985; Schustack, Ehrlich, & Rayner, 1987; Rayner & Well, 1996).

Quant à la décision de l'impact de la prochaine fixation, une observation de la littérature révèle l'importance du contexte visuel de la fixation en cours, traitée

par l'intermédiaire de la vision parafovéale. On peut relater deux grands types d'observations :

- Quand on enlève les espaces entre les mots, les saccades oculaires des lecteurs sont plus courtes (Pollatsek & Rayner, 1982; Morris, Rayner, & Pollatsek, 1990; Rayner et al., 1998).
- La longueur du mot situé à droite de la fixation influence le lieu où les yeux vont se poser sur lui, donc la taille de la saccade (O'Regan, 1979; O'Regan & Jacobs, 1992; O'Regan, 1992; Rayner & Pollatsek, 1989). Cette position d'arrivée est quasi systématique et se situe entre le début et le milieu du mot (McConkie, Kerr, Reddix, & Zola, 1988; O'Regan & Jacobs, 1992; O'Regan, 1992; Rayner, 1979; Vitu et al., 1990; Vitu, O'Regan, Inhoff, & Topolski, 1995). Elle a été nommée position préférée du regard (Rayner, 1979). Notons cependant que cette position est une moyenne et que, par exemple, environ 10% des points de fixations sont des espaces.

Ce champ de recherche n'est pas homogène : il oppose deux conceptions du fonctionnement de la saccade oculaire. Leurs divergences reposent avant tout sur la définition des contraintes qui commandent aux saccades. Pour la première, ce serait des contraintes purement oculomotrices et visuo-spatiales qui présideraient aux saccades, tandis que pour la seconde, différentes contraintes les affectent, comme, par exemple, le contexte linguistique dans lequel se trouve le lecteur. En fonction des protocoles expérimentaux, les recherches sont susceptibles d'apporter des résultats opposés sur des tâches similaires. En comparant les caractéristiques des mouvements des yeux selon qu'on est en lecture normale ou en recherche dans du texte modifié (remplacement de lettres) ou encore en simple recherche visuelle (rechercher une lettre dans un texte), deux conclusions très différentes sont apportées :

- l'équipe de Françoise Vitu observe que les caractéristiques générales du mouvement des yeux, comme la taille des saccades et la durée des fixations ne varient quasiment pas d'une condition expérimentale à une autre. Idem pour les caractéristiques plus locales tel que le taux de mots sautés, la zone d'arrivée dans les mots, la probabilité et la position des re-fixations. Elle en conclut que

les yeux sont capables de générer une stratégie de recherche indépendante des informations à traiter (des informations linguistiques) et généralise en postulant que ces stratégies purement oculomotrices seraient à la base des mouvements des yeux dans la lecture (Vitu et al., 1995).

- Rayner ajoute dans son protocole expérimental la prise en compte de la fréquence du mot fixé. A l'inverse de Vitu, il observe qu'en situation normale de lecture, les temps de fixations sont plus courts, les saccades plus amples, les sauts moins fréquents et les refixations (les retours en arrière) plus fréquentes. Cette différence de fréquence affecte la durée de fixation du mot cible pendant la lecture, mais pas pendant les recherches. Il en conclut que, contrairement aux observations apportées par Vitu, les mouvements des yeux ne seraient pas guidés par une stratégie globale et une tactique locale mais par la réponse immédiate aux contraintes locales (Rayner & Fischer, 1996).

3.6. Une fixation par mot ?

La plupart des expériences citées s'intéressent à la reconnaissance de mots isolés, dans des dispositifs expérimentaux nécessairement réducteurs. Il importe pour ces recherches que l'œil fasse une fixation sur chacun des mots présentés. Affirmer avec Gombert, maintenant que l'on sait enregistrer des mouvements extrêmement rapides, que l'œil se pose pratiquement sur tous les mots du texte (Gombert, 1996) est donc une extrapolation audacieuse des travaux de recherche en cours. Ces travaux sont à examiner avec attention.

Tout d'abord, on sait que le pourcentage de mots courts sautés (en situation de lecture de textes) dans la vision fovéale est relativement élevé.

Taille des mots	2	3	4	5	6	7-10
Pourcentages de mots sautés	75	58	42	35	20	10

Tableau 2 — Pourcentages de mots sautés en fonction de leur taille. D'après (Vitu et al., 1995)

Le tableau précédent montre que si les mots courts (souvent fonctionnels) sont très fréquemment sautés, les mots de plus de 4 lettres, pour la plupart à contenu sémantique, présentent un pourcentage de non-fixation encore important. Qui plus est, bien peu de ces observations intègrent la "vitesse de compréhension" du lecteur. Il semble toutefois qu'un accord existe sur la durée des deux événements repérables : en moyenne, 250 millièmes de seconde pour une fixation et 50 ms pour une saccade de déplacement (Vitu et al., 1990). Si on tient compte des retours en arrière et des passages à la ligne, on peut considérer qu'il y a, lors d'une lecture libre, environ 13 000 paires fixation+déplacement à l'heure. S'il était exact que chaque mot donne "pratiquement" lieu à une fixation, on ne dépasserait pas une vitesse de lecture de 13 000 mots/heure, à comparer à la vitesse de la parole ou à rapporter aux conditions expérimentales (par exemple en auto présentation segmentée, mot à mot). Pour autant, les bons lecteurs, sans recourir à "l'écrémage"¹⁹, comprennent aisément un texte ordinaire à plus de 20 000 mots/heure, donc en lecture intégrale. Il semble impossible d'affirmer aujourd'hui que l'œil s'arrête sur tous les mots.

Bien au contraire, une recherche conduite dès 1991 à l'INSERM U305 de Toulouse sur la coordination binoculaire chez de jeunes lecteurs apporte des précisions complémentaires. Ces chercheurs tiennent notamment compte de la vitesse de lecture des sujets. "L'amplitude de la saccade de progression détermine la plage de lecture. Plus celle-ci est étendue, plus le nombre de fixations est réduit et plus la lecture est rapide. Les lecteurs lents (vitesse de lecture inférieure à 9000 mots/heure) font des saccades de progression de faible amplitude par rapport aux lecteurs rapides (2°39' contre 4°23')." Cette amplitude est donnée ici en degrés d'arc mesurant l'angle visuel. Dans leur protocole de recherche, une ligne de texte est vue sous un angle de 21°10' et, par exemple, le mot "les" affiché à l'écran est vu sous un angle de 1°08'. Si une saccade moyenne chez un lecteur lent (2°39') dépasse déjà la longueur moyenne d'un mot (entre 5 et 6 caractères en français), c'est déjà plus du double chez un enfant de 10 ans, bon lecteur, soit en moyenne 3 mots, ce qui ne signifie pas non plus une amplitude régulière et mécanique des empan

¹⁹ Technique de lecture rapide qui consiste à prélever des informations dans un texte par survol ou "lecture en diagonale".

mais invalide complètement l'existence d'une fixation par mot. Notons que les meilleurs lecteurs que nous avons pu observer lisaient à des vitesses voisines de 60 000 mots/heure, soit environ 100 signes par seconde. A cette vitesse-là, ils utilisent des empan moyens de 25 signes soit environ 6 mots. De ce rapide calcul, nous pourrions déduire que la taille de l'empan de lecture pour un lecteur véritablement expert est d'environ 10°, soit approximativement la moitié d'une ligne (Pugh, Bassou, Granié, & Morucci, 1991).

Le recrutement, par les laboratoires de recherche, des sujets d'expérience se fait, le plus souvent, parmi les étudiants de l'université, sans jamais mesurer leur performance en lecture. Nos propres observations, portant sur des populations de professeurs, montrent une vitesse moyenne de lecture de l'ordre de 12 000 mots par heure. On verra plus loin que certains élèves de notre cohorte ont déjà des performances bien supérieures... Il est donc probable que les études de psychologie ne tiennent pas assez compte de ce qui se passe ordinairement chez un lecteur expert en situation réelle de lecture. Si l'empan de lecture augmente de façon aussi considérable, peut-on alors se dispenser de s'interroger sur les performances intrinsèques de l'œil ? Nous verrons un peu plus loin que l'examen des « experts perceptifs » enrichit notre propre connaissance de la perception.

3.7. La zone de traitement de la forme des mots

Depuis quelques années, l'équipe française de l'INSERM U334 s'efforce de préciser le rôle et la fonction d'une petite zone cérébrale qu'elle a nommée la Zone de Traitement de la Forme des Mots²⁰ (ZTMF). Quelques articles publiés précisent sa spatialité et son fonctionnement présumé (Cohen et al., 2000; Cohen et al., 2002; Dehaene, Le Clec'H, Poline, Le Bihan, & Cohen, 2002). Pour la neurologie, discipline scientifique des auteurs, il est important de déterminer avec exactitude la géographie de l'aire cérébrale concernée, ainsi que la temporalité de son activité. La zone concernée semble s'activer de manière ultra rapide après le stimulus visuel. Les auteurs ont montré que cette

²⁰ Visual Word Form Area

zone s'activait chez tous les adultes lettrés en réponse à la présentation visuelle de mots quelle que soit leur localisation sur la rétine et quelles que soient la typographie, la couleur ou la casse (majuscule vs. minuscule) utilisées. Les auteurs rapportent que cette zone semble être lésée chez les patients souffrant d'alexie pure, c'est-à-dire incapables de lire des mots, mais capables de les écrire et de les comprendre oralement.

Les expériences menées conduisent les auteurs à envisager que cette zone cérébrale est engagée au début de l'activation du réseau neuronal à l'œuvre dans le processus de lecture. Son rôle se situe juste après celui des aires visuelles et juste avant les aires plus spécifiquement engagées dans les traitements syntaxiques et/ou sémantiques. Les auteurs postulent que cette zone cérébrale a pour fonction principale de construire une représentation pré-lexicale des chaînes de caractères et de décider si cette chaîne respecte les règles orthographiques ou si elle n'est qu'une simple suite de lettres inorganisées (par rapport aux configurations réelles de lettres dans la langue du lecteur). Il est utile de préciser le protocole expérimental des auteurs : le sujet doit fixer un point central et le mot apparaît soit à gauche, soit à droite de ce point de fixation. Quand le mot est présenté à gauche du point de fixation, les lettres qui en sont le plus éloignées font un angle de 6° par rapport au centre, alors que les lettres les plus proches ne font qu'un angle de 2° . Si on se rapporte à la figure 17 du présent chapitre, on voit que les informations présentes à 6° du point de fixation sont déjà très dégradées (seulement 30% de la vision nette) ; on se trouve assurément en vision parafovéale. Ce qui revient à dire que la ZTMF traite des informations qui proviennent à la fois des zones fovéales et parafovéales²¹, et que c'est à partir de l'une et de l'autre zone qu'elle prendra sa décision sur l'existence orthographique de la chaîne projetée. Les auteurs n'ont pas explicitement envisagé cet état de la question, mais le domaine de recherche n'a guère plus de trois ou quatre ans, et ils enrichissent leurs articles de nouveaux éléments à tester. Nous ne rajoutons qu'un petit élément.

²¹ Ce qui est tout à fait congruent avec le fait que la ZTMF se situe après les aires visuelles.

Une question, épistémologiquement intéressante, se pose quant à l'émergence de cette zone et qui renvoie à l'innéité des processus langagiers. Comment se fait-il qu'une zone se soit spécialisée dans le traitement de la forme des mots écrits alors que, d'une part l'invention de l'écriture n'est qu'une conquête récente de l'humanité et que, d'autre part, sa maîtrise ne se manifeste pas avant quelques années chez le jeune enfant ? Les auteurs n'en sont encore qu'au stade des hypothèses et des analogies avec d'autres processus, mais il semble déjà que cette région est particulièrement sensible à la topologie des inputs visuels, à leur forme générale : sa position spécifique à l'intersection des zones visuelles et des zones du langage lui confère sa spécificité.

3.8. Les experts en perception visuelle

Il est intéressant de s'arrêter un instant sur le cas des "experts" dans des activités de perception, c'est-à-dire sur des personnes capables de percevoir ce que des novices ne remarquent pas : ce sont les observateurs des oiseaux, les radiologistes, les experts dans l'interprétation des photos aériennes, certains sportifs, les champions d'échecs, les aviateurs, etc²². On ne sait pas grand chose sur les mécanismes de perception mis en place par ces experts, même si ce qui les caractérise, c'est d'être très rapides, et très précis et, le plus souvent, de manière totalement inconsciente. Quand on leur demande comment ils procèdent, on obtient des réponses évasives, où il semblerait qu'il suffise d'ouvrir les yeux et de regarder. Néanmoins, les rares recherches disponibles suggèrent que les experts auraient appris à combiner ce vers où (et quoi) il faut diriger l'attention avec la connaissance des différents facteurs qui caractérisent la tâche perceptive à mener (Haber, 1966).

A l'appui de cette hypothèse, les travaux de Chase et Simon montrent que ce qui transparait chez un maître des échecs à la fois comme une analyse visuelle ultra-rapide des positions et une mémoire prodigieuse des situations ne se manifestait jamais quand les pièces étaient placées au hasard sur l'échiquier (Chase & Simon, 1973; Gobet & Simon, 1998). Pour ces auteurs, tout se passe

²² L'essentiel de cette partie est issu de la revue de la littérature faite en 1999 par (Pylyshyn, 1999)

comme si les champions d'échecs avaient appris un répertoire très large de patrons d'échiquier qui leur permet de classer et d'encoder tout nouvel échiquier, à partir du moment où ils respectent les règles de classification. Ce qui est donc particulier dans la vision experte des champions d'échecs, c'est ce système de classification qu'ils ont construit, qui leur permet de reconnaître instantanément tout nouvel échiquier comme plausible et de le mémoriser (probablement en l'intégrant au système de classification).

L'observation de certains sportifs de haut niveau offre aussi l'occasion d'étudier la perception visuelle experte. Il est banal de rappeler qu'une perception rapide et une réaction adaptée sont gages de réussite. Cependant, bien peu de travaux ont envisagé d'observer un grand nombre d'experts pour en extraire une généralisation des processus de traitement rapide de l'information (Starkes, Allard, Lindley, & O'Reilly, 1994; Abernethy, Neil, & Koning, 1994). Dans la plupart des cas, l'étude des différences entre novices et experts reste confinée dans des domaines spécifiques où les experts excellent. Malgré tout, on attribue généralement l'expertise à l'habileté d'anticiper les éléments clés de l'activité. Cette anticipation est basée sur l'observation des gestes de l'adversaire qui, à l'intérieur d'un cadre connu par l'expert, doivent orienter la balle de tel ou tel côté (par exemple pour la trajectoire du coup droit au tennis) (Abernethy, 1991). L'expertise visuelle en sport, comme l'expertise du joueur d'échecs, semble reposer en fait sur une expertise non-visuelle pour prévoir, attendre et identifier les seuls événements significatifs. Il s'agit donc d'une capacité acquise à orienter son attention (Nougier, Ripoll, & Stein, 1989; Castiello & Umiltà, 1992b; Castiello & Umiltà, 1992a; Greenfield, deWinstanley, Kilpatrick, & Kaye, 1994). Notamment, les travaux de Castiello précisent que les athlètes de haut niveau sont capables de séparer des cibles et de les manipuler simultanément dans des hémisphères séparés. Cet aspect de la perception visuelle (plurimodale ou unimodale) est l'enjeu de différents travaux contradictoires (McCormick, Klein, & Johnston, 1998) ; cependant, même dans le cas où on suppose que l'attention ne se situe que dans un seul hémisphère, il semble maintenant avéré que le cerveau est susceptible de traiter simultanément l'information de manière très performante dans au moins deux espaces différents.

Le niveau perceptif d'un expert diffère souvent de celui d'un novice par le fait que l'expert a appris où se situaient les informations critiques dans le patron général des stimulus. Dans ces conditions, l'expert peut diriger son attention sur les lieux ou les événements critiques. Biederman & Shiffrar ont réalisé une étude sur les personnes classant les jeunes poulets âgés de quelques jours selon leur sexe. Cette classification semble particulièrement difficile, tellement difficile d'ailleurs qu'il faut des années d'entraînement (consistant à répéter les essais) pour devenir un de ces rares experts. En les étudiant soigneusement, les auteurs ont trouvé que ce qui les distingue, c'est l'habileté qu'ils manifestent concernant d'une part le lieu de la fixation oculaire et d'autre part les éléments saillants qu'ils recherchent. La plupart des experts ne sont en rien conscients de ce qu'ils ont appris durant les années d'entraînement. Biederman et Shiffrar ont montré qu'en disant aux novices où se situaient les informations décisives pour leur activité de différenciation, on leur faisait gagner beaucoup de temps vers l'expertise (Biederman & Shiffrar, 1987). Ce que les experts ont appris tacitement par l'expérience et que les novices ont plus rapidement maîtrisé par la transmission des quelques éléments topiques, c'est comment amener le système visuel à la position spatiale exacte et quel type d'informations doit être encodé, ces deux aspects se trouvant en dehors du système visuel lui-même.

3.9. Apprentissage perceptif

Il existe une littérature assez abondante sur cette notion d'apprentissage perceptif²³. Dans une revue de la question, Goldstone caractérise cette notion par les changements à long terme d'un système perceptif, ce qui lui permet d'augmenter la qualité de ses réponses à l'environnement (Goldstone, 1998). Quatre grands mécanismes sous-tendent cet apprentissage perceptif, qui améliorent la perception à l'intérieur d'une tâche :

1. Une plus grande attention apportée aux formes discriminantes.
2. Un développement de récepteurs sensoriels spécialisés dans des sous-partie du stimulus.

²³ « Perceptual learning » est le terme consacré dans la littérature anglo-saxonne.

3. Une capacité à séparer psychologiquement des stimuli jusqu'alors indistinguables.
4. Une habileté à détecter une construction unique représentant une configuration complexe, en lieu et place d'une multitude de parties juxtaposées.

Dans les expériences menées sur l'expertise visuelle spatiale (Poggio, Fahle, & Edelman, 1992; Fahle, Edelman, & Poggio, 1995), l'observation de l'apprentissage perceptif²⁴ montre qu'il ne nécessite que relativement peu d'essais pour être opérationnel. Cependant, les effets de l'apprentissage ne se transfèrent pas entre différentes tâches requérant de l'hyperacuité visuelle.

Dans une tâche visuelle de détection de certaines figures (par exemple de courtes barres obliques dans un environnement de barres verticales), on mesure la précision, la vitesse et l'insensibilité à la distraction. Ahissar et Hochstein ont étudié l'amélioration par l'entraînement de cette pratique. Comme souvent, ils constatent l'amélioration par la pratique, mais ce qui semble important, c'est que l'amélioration ne se produit que lorsque l'ensemble de la tâche est entraîné, lorsque le rapport entre les cibles est possible : si on entraîne séparément à la détection de deux formes (donc indépendamment de la discrimination de l'une avec l'autre), on n'observe que peu d'amélioration dans la tâche de départ. Même pour des tâches qui pourraient paraître ne nécessiter que peu de charges cognitives, ces observations confirment le rôle primordial joué par le modèle dans lequel s'intègrent les observations : ce qu'on entraîne, c'est la détection « systémique » des modalités visuelles qui invalident ou corroborent les traitements cognitifs engagés.

3.10. Le traitement inconscient des informations visuelles

Différents domaines de recherches font plus ou moins explicitement appel au traitement inconscient des données visuelles pendant la lecture. On peut les regrouper en trois catégories :

²⁴ Dans plusieurs tâches de discrimination spatiale, on observe que le système visuel humain est capable de traiter des relations spatiales avec une précision de l'ordre d'une fraction d'un diamètre de photo-récepteur (Fahle et al., 1995)

- La première travaille sur la simultanéité des informations perçues dans le champ visuel. On citera, parce qu'elle est spectaculaire, l'expérimentation menée par Andreewsky : on projette au tachyroscope des couples de mots espacés l'un de l'autre et suffisamment rapidement pour que le lecteur n'ait le temps d'en lire qu'un. On demande au sujet de dire à haute voix le mot vu. Parmi les couples de mots, on trouve, par exemple, [FILS, AIGUILLE] et [FILS, PERES]. Lorsque le mot vu est FILS, on observe s'il va être, de manière significative, prononcé [fil] ou [fis] en fonction de l'autre mot du couple (AIGUILLES ou PERE) qui, lui, n'a pas été vu, tout au moins dont le lecteur n'a pas conscience. (Andreewsky, 1980). Pour Andreewsky, l'accès au lexique est, dans ce cas, tributaire des propriétés sémantiques de mots non-identifiés consciemment. Ce traitement inconscient de différentes zones du champ visuel est tout à fait congruent avec ce que nous rappelions des experts au paragraphe précédent.

- Une deuxième catégorie d'expérimentation repose sur ce qu'on nomme la technique d'amorçage et qui a été utilisée très fréquemment pour l'étude des processus automatiques et inconscients. Le paradigme expérimental classique repose sur la présentation d'une chaîne de caractères en minuscule pendant un temps très court (environ 30-60 ms), suivie d'un masque (par exemple #####) suivie de la cible en majuscule. Ce qui permet de parler d'amorçage, c'est le temps d'identification plus court d'un mot quand l'amorce, pourtant non vue consciemment, entretient un rapport avec lui. L'effet maximum est obtenu quand la cible et l'amorce sont identiques, mais on observe également des effets quand le mot diffère d'une lettre ou avec des amorces faisant varier les aspects morphologiques des mots. La littérature est contradictoire sur le fait de savoir s'il existe des effets d'amorçages dans les épreuves de décision lexicale (décision sur l'appartenance de la chaîne présentée à l'ensemble des mots) et dans les épreuves de dénomination de mot (Bowers, Vigliocco, & Haan, 1998; Perea & Gotor, 1996; Perea & Gotor, 1997). Malgré tout, il semble qu'il existe un accès inconscient au lexique par les amorçages (Ferrand, Grainger, & Segui, 1994; Draine & Greenwald, 1998; Luck, Vohel, & Shapiro, 1996). Sereno montre comment une amorce sémantique peut aussi être activée par

une phrase entière (Sereno & Rayner, 1992), et rappelle des études antérieures allant dans ce sens (Schuberth & Eimas, 1977).

- La troisième catégorie de traitement inconscient se réfère à des expériences dont les premières publications sont encore récentes. Le propos n'est plus cette fois de préciser les modalités d'*accès* aux informations, mais de savoir si ces informations peuvent être *manipulées* de manière inconsciente. Les travaux récents de l'équipe de Stanilas Dehaene montrent que, non seulement, les informations peuvent être perçues inconsciemment et permettre un accès à d'autres informations, mais encore qu'elles sont susceptibles de servir de support à des manipulations, des constructions sémantiques inconscientes (Dehaene et al., 1998; Naccache & Dehaene, 2001). Ainsi, à partir de leur paradigme expérimental observant le rôle d'un amorçage sur la classification de nombres en deux ensembles (supérieur ou inférieur à 5), ils ont montré que plus l'amorce était numériquement proche de la cible, plus l'effet était fort, illustrant par-là non seulement un effet de proximité sémantique entre amorce et cible, mais aussi une nécessaire manipulation de l'amorce.

3.11. Conclusion

L'exposé de ces différents domaines de recherches illustre bien la complexité des processus qui sous-tendent la perception visuelle. Bien des entrées sont manifestement justifiées pour l'analyser et, pour chacune d'entre elles, un débat quant aux observations et à leurs analyses est en cours. Cependant, nous pouvons essayer d'extraire quelques idées forces qui aident à mieux comprendre les contraintes et les effets de ce système perceptif sur la lecture :

1. L'information se prend à l'aide de deux zones : fovéale et parafovéale. Ces deux zones ont des caractéristiques différentes, la zone fovéale, réduite dans sa taille, apporte des informations très détaillées, alors que la zone parafovéale semble performante dans l'exploitation de zones floues, plus en rapport avec les formes globales de mots sans distinction précise des lettres.

2. Dans le cas du traitement d'un matériau écrit, ces deux zones activent une petite zone cérébrale située avant les aires du langage, laquelle travaille la forme du mot et l'adéquation de la chaîne avec les règles de construction orthographique.
3. Sans information parafovéale, la lecture de phrases devient plus erratique, on observe davantage de retours en arrière, comme si les informations n'avaient pas de cadres où s'intégrer.
4. Les experts perceptifs sont le plus souvent inconscients des processus qui fondent leur expertise. Il semble qu'ils soient capables d'organiser les informations sur une grille interprétative construite par l'expérience. Elle leur permet également de savoir vers quels endroits diriger l'attention, voire les différents centres d'attention.
5. Cette expertise visuelle est susceptible de s'entraîner pour améliorer plus rapidement l'habileté des sujets.
6. L'inconscience observée chez les experts se retrouve dans des situations normales de lecture où le lecteur peut extraire de l'information d'inputs non perçus.

Pour revenir quelque peu au propos général de ce travail sur le rôle de la syntaxe dans les processus de lecture, nous voudrions terminer cette partie par le rapide rappel des tous derniers travaux de Shapiro et Caramazza. Ils se demandent comment les connaissances sur les différentes catégories des mots sont stockées dans le cerveau et, surtout, comment le cerveau y accède (Shapiro & Caramazza, 2003). Reprenant de nombreux travaux de neuropsychologie, d'électrophysiologie et d'imagerie cérébrale, ils suggèrent que l'information sur la catégorie grammaticale des mots pourrait être représentée indépendamment de celle sur sa signification, et qu'elle serait associée au niveau du traitement de la forme des mots et des informations morphologiques. Dans ces conditions, si l'information sur la catégorie est située au niveau de la forme des mots, tout porte à croire qu'elle est perceptible par la vision parafovéale et, de ce fait, de manière inconsciente.

Nous voudrions également insister sur le fait que l'intention semble bien être la base de l'action. Les neurophysiologistes les plus avertis considèrent le cycle perception-action comme indissociable. « L'exploration du monde visuel par un sujet serait définie par des schèmes anticipatoires définis comme des plans pour l'action perceptive. » (Berthoz & Petit, 1996). Cette grande place accordée à l'anticipation semble bien souvent exclue des recherches sur la lecture, déconnectant ainsi la perception de toute anticipation.

Nous tenterons de préciser au cours du chapitre 10 de la deuxième partie de ce travail l'importance de la vision parafovéale dans l'établissement des cadres structuraux tels que les envisage Koriat. Ce modèle structural semble bien être situé à la croisée de l'usage efficace de la perception parafovéale et des connaissances que le lecteur a du fonctionnement de la syntaxe de la langue écrite, connaissances qui lui permettent de créer des schèmes syntaxiques anticipatoires de la phrase qu'il parcourt (l'équivalent des « grilles perceptives » des experts perceptifs).

Chapitre 4.

Traitement syntaxique et compréhension de phrase.

Le modèle de lecture structural que nous avons présenté au premier chapitre, et qui est au centre de nos préoccupations dans ce travail, ne peut s'examiner sans l'apport des différents modèles qui tentent d'expliquer les processus psychologiques. C'est un domaine très « controversial », ce qui est sans doute le signe d'une grande vitalité. Néanmoins, il nous semble nécessaire de démêler les liens qu'entretient le modèle de Koriat avec les descriptions des différents processus cérébraux, tout en rappelant que Koriat lui-même ne s'en est soucié que fort tardivement : la première mention d'un éventuel rapprochement entre ses propositions et un des modèles que nous allons présenter ne date que de 2000 (Musseler et al., 2000).

4.1. Les grands modèles explicatifs

Le langage n'est pas simplement une collection de noms et de significations conservés en mémoire. Depuis fort longtemps, des chercheurs comme John Hughlings Jackson²⁵ ont émis l'hypothèse que la nature, l'essence du langage se situait dans la production de propositions. Il semble bien que, plus que la capacité à produire des sons articulés pouvant devenir des mots, c'est la possibilité d'organiser ces mots à l'aide de règles syntaxiques qui est spécifique du cerveau humain. Comme la linguistique structurale l'a

²⁵ J.H. Jackson (1835-1911), est considéré comme le fondateur de la neurologie britannique. Il ne concevait pas l'existence de "centres" fixes mais défendait l'idée d'une représentation " plurielle " d'une même fonction cérébrale. Ainsi, contrairement à la doctrine localisationniste standard qui postule l'existence de plusieurs "centres" du langage assurant chacun une fonction spécifique, Jackson conçoit le fonctionnement cérébral de manière globale et hiérarchique. Son insistance sur la dimension évolutive de ces fonctions est aujourd'hui largement prise en compte dans le cadre de la psychologie évolutionniste.

abondamment montré, la syntaxe ne se limite pas aux règles de remise en ordre de mots, elle révèle la façon dont les propositions peuvent prendre des sens différents en fonction du contexte où elles se trouvent.

Comment et où la syntaxe s'intègre-t-elle dans les processus de compréhension du langage ? Différents modèles théoriques ont été construits qui tentent, chacun à leur manière, de répondre à cette question en s'appuyant sur des données empiriques. L'analyseur²⁶ syntaxique n'est certes pas le seul à être à l'œuvre durant la compréhension de phrases. D'autres sources d'informations sont activées telles que les informations pragmatiques, lexicales ou sémantiques. Toutefois, dans cette partie, nous ne nous attarderons pas sur les modalités des traitements non-syntaxiques, sauf si elles viennent à interagir avec les processus syntaxiques.

Tous les modèles de compréhension supposent l'existence d'un « processeur syntaxique », mais ils diffèrent sur le moment de sa mise en œuvre, sur ses rapports avec les autres « processeurs » et sur l'exhaustivité de ses traitements : construit-il une analyse structurale détaillée ou seulement une description des relations existant entre différents éléments de la phrase ?

Pour tenter une explication, forcément abstraite, de ces différents modèles, nous devons emprunter à Fodor la notion de système périphérique²⁷. Pour cet auteur, un système périphérique est chargé de « représenter le monde à la pensée », laquelle est considérée comme le système central, celui vers qui les systèmes périphériques vont déverser de l'information. Un tel système est chargé d'une tâche bien précise et très spécialisée, à l'intérieur d'un seul domaine, qui lui est propre. Les systèmes périphériques existeraient pour les sens habituels, auquel Fodor rajoute le langage. Si on peut parler d'un système périphérique de la vision ou du langage, on envisage également l'existence de sous-systèmes comme, par exemple, ceux de la reconnaissance des formes ou

²⁶ Il s'agit du « programme mental qui analyse la structure de la phrase au cours de la compréhension du langage » (Pinker, 1999). Nommé « parser » dans la littérature anglophone.

²⁷ L'appellation *système périphérique*, largement employée par Fodor, est souvent remplacée par la notion de *module*. Nous préférons employer *système périphérique* car le *module* est trop connoté (modulaire, modules autonomes,...). Les systèmes périphériques ne sous-entendent pas ce type de fonctionnements, sauf, bien sûr, chez Fodor qui considère que les systèmes périphériques sont modulaires.

de la reconnaissance des couleurs qui participent du système de la vision. Pour le langage, plusieurs sous-systèmes sont opérants, comme (mais ce n'est pas exhaustif) ceux de la syntaxe ou de la reconnaissance des items lexicaux. Les choses se compliquent dès qu'on tente de comprendre comment ces systèmes périphériques fonctionnent, à partir de quelles données ils travaillent, quelles sorties ils génèrent et pour qui, quel est leur degré d'autonomie dans le traitement... Un certain nombre des caractéristiques de ces systèmes ne posent pas de problèmes fondamentaux : par exemple, dire que les systèmes sont rapides ou qu'ils sont associés à une architecture neuronale fixe n'ouvre pas sur de grandes controverses²⁸. Par contre, certains aspects divisent sévèrement la communauté scientifique. L'affirmation de Fodor selon laquelle « les systèmes périphériques sont informationnellement cloisonnés²⁹ » (Fodor, 1986) en fait partie. Ce que veut dire Fodor, c'est d'une part que le fonctionnement des systèmes périphériques est insensible aux processus allant de haut en bas (top-down), donc que le processeur central ne peut pas influencer sur la façon de travailler des systèmes périphériques et, d'autre part, que les systèmes (ou sous-systèmes) périphériques ne peuvent pas communiquer entre eux : ils ne font que fournir de l'information au processeur central. Fodor réfute l'idée que les mécanismes d'anticipations ou de croyances puissent influencer sur le travail de certains systèmes périphériques. En quelque sorte, dire que les systèmes sont cloisonnés revient à dire qu'ils contiennent toutes les informations dont ils ont besoin pour mener à bien les opérations qu'on leur demande. Les informations dont ils disposent seraient tellement délimitées que Fodor qualifie ces systèmes d'obtus. Ils ne prennent appui que sur une petite partie de l'ensemble de l'information qui pourrait être nécessaire pour résoudre l'ensemble de la tâche. Cependant, le peu d'indices qu'ils ont à traiter et le fait de ne s'occuper que d'indices de bas-niveaux les rend, pour Fodor, particulièrement rapides.

Bien entendu, ces positions sont très contestées et de nombreux travaux proposent une vision différente du fonctionnement des différents systèmes périphériques, du côté d'une interaction entre les systèmes, de ce qu'on peut

²⁸ Nous reviendrons sur ces deux aspects dans le chapitre 5.

²⁹ On parle parfois d'encapsulation.

appeler leur pénétrabilité. Il serait à la fois possible que les niveaux cognitifs supérieurs apportent de l'information non spécifique à chacun des systèmes périphériques, mais aussi qu'en cours de traitement, l'information circule entre eux. Les expériences sur l'influence du contexte dans l'accès lexical sont exemplaires de cette alternative. Dans la tâche de décision lexicale où après avoir entendu la phrase amorce : « Parce qu'il avait faim, il est entré dans la boulangerie pour acheter une *baguette*. » on montre classiquement qu'un sujet répond plus rapidement sur le mot *baguette* que s'il était présenté isolément ou dans un contexte neutre. Les modèles interactifs expliquent cette rapidité par la pénétrabilité du système périphérique lexical qui est susceptible d'intégrer une information non spécifique : le contexte d'apparition du mot. Notons bien que, dans ce cas, le contexte est propre à la compréhension que le sujet fait de la phrase et, en cela, est donc une information envoyée par les systèmes supérieurs : on se rapproche des conduites d'anticipation. A l'inverse, les systèmes postulant une impénétrabilité des systèmes périphériques font valoir une autre interprétation : ce qui donne une plus grande rapidité dans le cas de l'exemple, ce n'est pas le contexte de la phrase, mais le fait de l'association avec la forme lexicale *boulangerie*, comme on peut l'avoir dans de simples tâches d'amorçage par présentation de mots isolés. Dans cette optique, la présentation de *boulangerie* active un réseau lexical dont un autre nœud est le mot *baguette*. Cette activation explique la diminution observée du temps nécessaire de traitement (Morton, 1969). Dans ce cas, toutes les informations traitées restent internes au système du lexique mental, ce qui nous ramène au cloisonnement des systèmes périphériques.

Une autre caractéristique des systèmes périphériques, c'est qu'ils n'ont pas à se soucier de prendre une quelconque décision sur la tâche en cours : ils laissent ce travail au processeur central, qui, lui, se charge d'intégrer les différents résultats proposés par les différents systèmes périphériques. La question se pose alors de la simultanéité temporelle du fonctionnement des systèmes périphériques. Cette question est particulièrement sensible quand on considère le problème de la mémoire de travail.

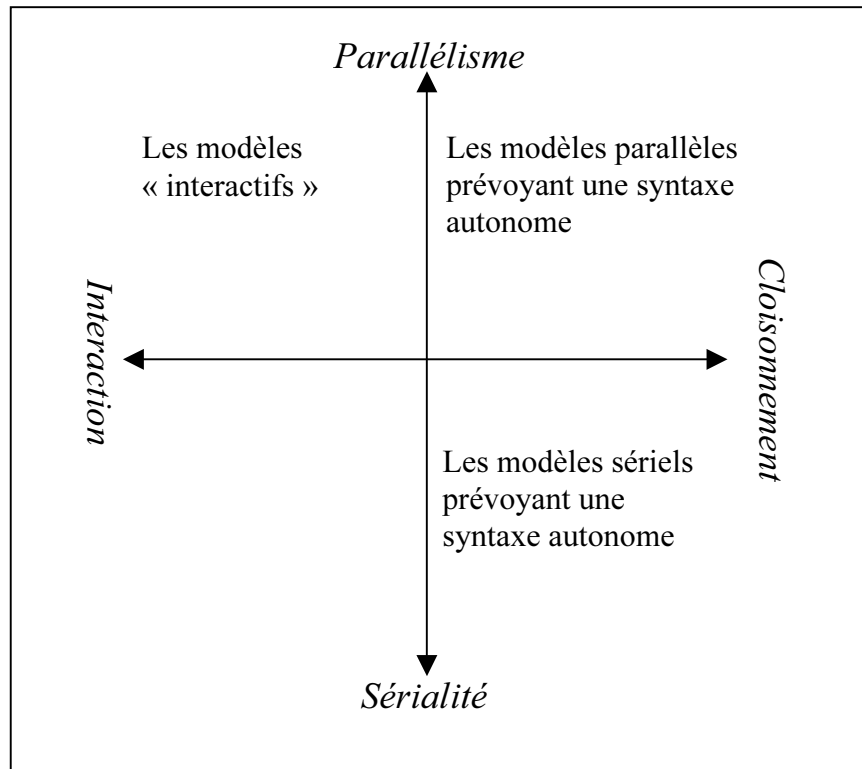


Figure 22 — Positionnement des 3 modèles qui tentent de décrire le processeur syntaxique.

Les modèles que nous allons brièvement présenter pour la lecture se réfèrent aux deux grands paradigmes du fonctionnement cérébral et se structurent autour de deux grands axes : d'une part le cloisonnement informationnel des systèmes périphériques ou leur interaction et, d'autre part, la sérialité temporelle de fonctionnement des systèmes ou leur parallélisme. Rappelons que les modèles interactifs impliquent une communication permanente entre les différents processeurs qui sont donc en constant échange d'information en réception, comme en émission (MacDonald, Pearlmutter, & Seidenberg, 1994; McClelland & Rumelhart, 1981). Pour ce qui est de la seconde opposition, les modèles parallèles postulent l'existence d'une interaction seulement après que les aspects sémantiques et syntaxiques aient fourni leurs données dans un interpréteur central responsable de l'interprétation, alors que les modèles sériels proposent une priorité à l'analyse syntaxique qui influence l'analyse sémantique plus tardive, mais cette influence ne peut être réciproque (Frazier & Rayner, 1982; Mecklinger, Schriefers, Steinhauer, & Friederici, 1995). Les modèles sériels impliquent que chaque type de traitement linguistique soit traité par une unité autonome (ils ne peuvent donc être que cloisonnés) et

surtout qu'ils fonctionnent séquentiellement depuis le traitement des informations de bas niveau jusqu'aux informations de plus haut niveau. Les informations passent d'un système à un autre, chacun des systèmes recevant donc une partie de ses informations des sorties des traitements inférieurs.

Bien sûr, il faut considérer ces oppositions comme autant de continuums où les modèles explicatifs prennent place, même si l'on verra que certains modèles se situent à une extrémité de ces axes (figure 22). Bien évidemment, les quatre critères construisant les deux axes du plan sont susceptibles de catégoriser beaucoup d'activités cérébrales.

4.2. Les modèles sériels prévoyant une syntaxe autonome

On est là dans des schémas explicatifs du traitement syntaxique strictement encapsulés et sériels. D'une part, les modules syntaxiques, pragmatiques, thématiques, etc., ne peuvent pas s'influencer mutuellement et d'autre part, ils calculent chacun une seule sortie qui sera transmise au système central et aux modules suivants. Par exemple, le traitement sémantique ne peut pas, par définition, agir sur le processus syntaxique, sauf à la sortie du processeur syntaxique, mais alors seulement par le biais de l'analyseur général. Si celui-ci suspecte une contradiction ou une anomalie entre les informations élaborées par l'analyseur syntaxique et par l'analyseur sémantique (ou pragmatique,...) une ré-analyse de la phrase sera enclenchée pendant laquelle de nouveaux attachements seront construits, en tenant compte des premiers essais. D'autre part, comme nous l'avons expliqué plus haut, le processeur syntaxique ne traite que de l'information syntaxique puisqu'il est informationnellement cloisonné.

Le modèle le plus célèbre et sans doute le plus usité s'intégrant à ce paradigme explicatif est le modèle de *garden-path*³⁰ initié par Frazier et ses collaborateurs. Depuis environ deux décennies, ce modèle illustre la façon dont les théories linguistiques trouvent leur place dans la recherche des mécanismes de traitements des phrases. Très schématiquement, ce modèle postule que le processeur linguistique humain construit des structures

³⁰ Venant de l'expression *to be led in garden-path* : être mené en bateau.

arborescentes similaires aux représentations syntaxiques issues de la théorie X-bar. Deux règles principales commandent ce principe : *l'attachement minimal* (Frazier), qui élimine tous les nœuds inutiles et la *clôture tardive* qui attache chaque nouvel item lexical dans la structure syntaxique en cours d'élaboration. En fait, ces deux règles prennent leur place en cas d'ambiguïté dans la construction de la phrase. Ainsi, si une ambiguïté syntaxique est rencontrée, le processeur explore la représentation la plus simple, soit en ne créant pas de nœud inutile qu'on chercherait à garder en mémoire le temps de résoudre l'ambiguïté (attachement minimal), soit en intégrant les items qui génèrent l'ambiguïté à la structure qu'il avait commencée à construire (clôture tardive).

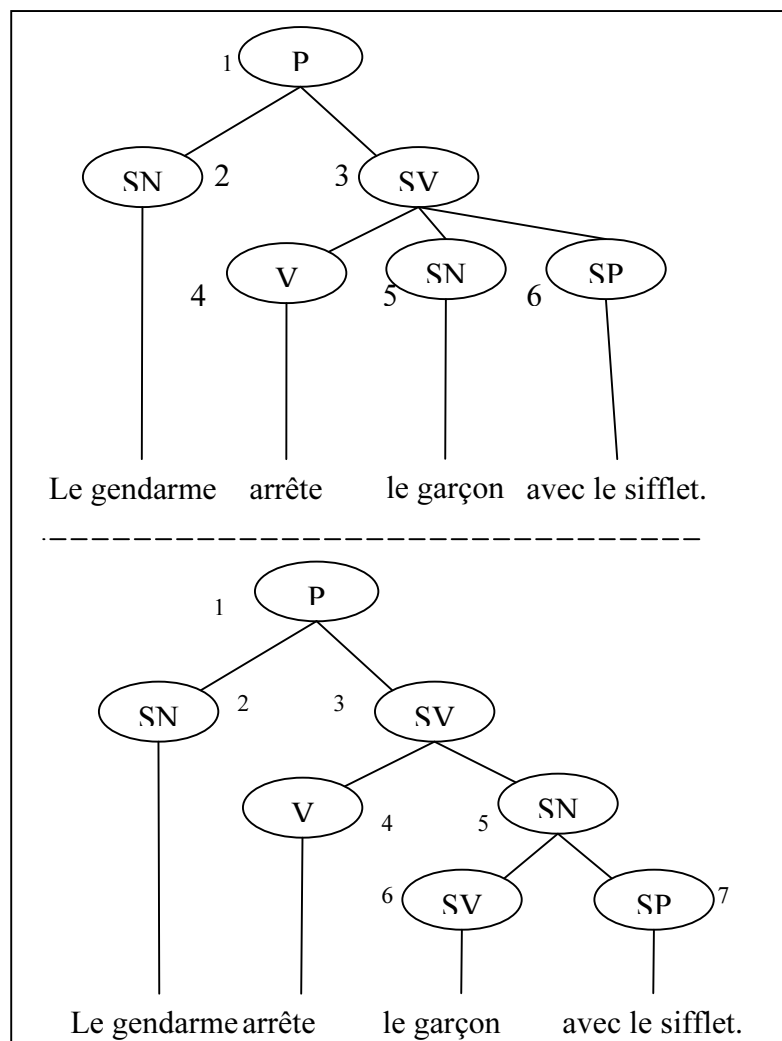


Figure 23 — Attachement non-minimal (7 nœuds) et minimal (6 nœuds) d'un syntagme prépositionnel ambigu

La figure 23 illustre le phénomène de l'attachement minimal : d'après le modèle de *garden-path*, le processeur syntaxique génèrera la forme à 6 nœuds

puisque le modèle cherche à générer la représentation syntaxique la plus simple.

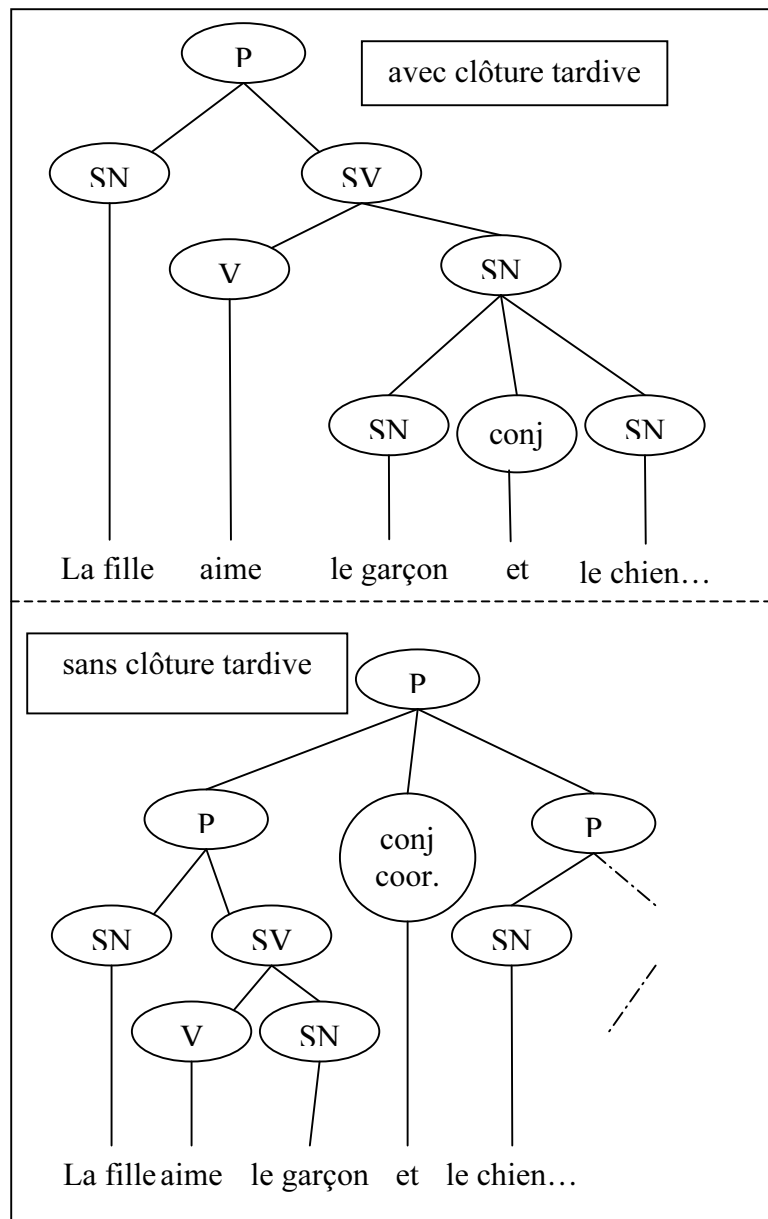


Figure 24 — Analyses d'une phrase selon le principe de clôture tardive

La figure 24, quant à elle, illustre le principe de clôture tardive dans l'analyse de la phrase « La fille aime le garçon et le chien aime son maître. » L'analyse correcte est donnée par la partie basse de la figure, la partie haute représentant ce que devrait faire un lecteur conduit sur une fausse piste à cause de la construction pas à pas de la syntaxe. Jusqu'au syntagme nominal *le chien*, il construit une structure syntaxique correspondant à la partie gauche, a priori

plus simple. La rencontre du deuxième *aime* va le conduire à recommencer depuis le début son analyse où il modifiera les attachements pour les rendre compatibles avec les nouvelles informations.

Si les informations non-syntaxiques ne sont pas utilisées pendant la période initiale d'analyse, ce modèle ne stipule pas qu'elles ne le seront pas à un autre moment. Un tel modèle doit prévoir quelles sont les autres informations disponibles et comment elles s'intègrent les unes aux autres. Frazier a ainsi proposé un processeur thématique (Frazier & Rayner, 1982) qui opère à la sortie du processeur syntaxique. Il propose, si nécessaire, des résultats d'analyses complémentaires qui seraient sémantiquement ou contextuellement plus précises. Cependant, ce n'est qu'à cette étape d'une éventuelle ré-analyse que les informations seraient alors consultées. Mais uniquement comme la marque d'une incohérence, comme une clé supplémentaire qui sera fournie au processeur syntaxique lequel, dans le cours de la ré-analyse, consultera uniquement des données d'ordre syntaxique puisqu'il est fonctionnellement cloisonné (Ferreira & Clifton, 1986).

Ce modèle de *garden-path* a d'abord été confronté à l'observation des mouvements oculaires du lecteur. En effet, selon les travaux de Frazier et de Rayner, les durées de fixation seraient plus longues sur la partie de la phrase levant l'ambiguïté lorsqu'elle existe (Frazier & Rayner, 1982) : le lecteur met plus de temps à lire les phrases avec un attachement non-minimal que les phrases à attachement minimal. Les lecteurs procèdent également à des régressions oculaires vers les zones ambiguës, pour pouvoir procéder à une nouvelle analyse de la phrase.

Toutefois, ces expériences ne sont pas sans susciter de vives controverses. Tout d'abord, le principe d'attachement minimal ne semble pas être possible dans toutes les langues, contrairement aux énoncés de Frazier. D'autre part, l'indépendance totale du processeur syntaxique est une conception du fonctionnement cérébral particulièrement contesté, comme nous le verrons un peu plus tard. Malgré tout, la volonté affirmée par ces auteurs de relier les travaux de la linguistique contemporaine à ceux de la psychologie

expérimentale classique paraît pour le moins digne d'intérêt et partiellement fertile pour les travaux à venir.

4.3. Les modèles parallèles prévoyant une syntaxe autonome

Ces modèles, respectant toujours l'encapsulation des traitements syntaxiques, sémantiques, thématiques, etc., introduisent la notion de parallélisme des processus. L'interaction entre les différents traitements aura lieu une fois que chaque processus spécialisé aura rendu son verdict, le processeur central sélectionnant une donnée à partir des sorties que les différents processeurs spécialisés auront données. Le parallélisme existe dans la mesure où les processeurs sémantiques, thématiques, syntaxiques, etc. travaillent en même temps, de manière parallèle, mais sans interagir entre eux, au moins dans l'étape d'analyse initiale.

4.3.1. Le modèle référentiel

Le paradigme, sans doute le plus connu à l'intérieur de cette catégorie, est le *modèle référentiel* (Crain & Steedman, 1985). Ce modèle stipule que des informations référentielles (construites soit par le contexte linguistique précédent, soit par la connaissance du discours dans lequel s'insère la production langagière que le lecteur analyse), sont susceptibles de jouer un rôle dans la sélection de l'analyse syntaxique la plus pertinente. D'une certaine façon, le contexte référentiel informe et prévient le lecteur des éventuelles mauvaises constructions syntaxiques qu'il pourrait être amené à faire. Dans les phrases présentées plus haut (« Le gendarme arrête le garçon avec un sifflet »), le syntagme prépositionnel « avec un sifflet » présente une ambiguïté : soit il précise l'instrument avec lequel le gendarme fait stopper le garçon (attachement au verbe), soit il précise le garçon que le gendarme arrête (attachement au nom). Alors que pour le modèle de *garden-path*, la résolution de cette ambiguïté ne peut avoir lieu qu'à l'aide d'informations purement syntaxiques, Crain et Steedman considèrent que sa résolution sera faite en référence à l'éventuel contexte précédant l'apparition de la phrase ou du

discours dans lequel elle se situe. Jusqu'à récemment, le chercheur ne pouvait décider qu'en observant les temps de lecture des phrases présentant une ambiguïté dans des contextes soit neutres soit favorisant une interprétation. En effet, si le contexte est en accord avec une interprétation, alors on doit observer des temps de lecture plus courts. Les résultats présentés dans les travaux illustrant le modèle référentiel (Crain & Steedman, 1985; Altman & Steedman, 1988) montrent que l'effet de *garden-path* (priorité à la structure la plus simple) dans les contextes favorisant telle ou telle construction syntaxique disparaît, alors qu'il est toujours présent quand le contexte est neutre. C'est là l'origine de la notion de *parcimonie* qui suppose qu'une construction syntaxique sera soutenue par un contexte référentiel qui la porte. Les auteurs se sont livrés à un réexamen des principales expérimentations mises en place pour illustrer le modèle de *garden-path*, en contrôlant systématiquement le contexte référentiel dans lequel les phrases apparaissaient. En quelque sorte, le processeur syntaxique fournit des informations à l'analyseur contextuel qui peut lui répondre de continuer puisqu'il y a congruence avec le contexte sémantique ou de stopper en cas d'incohérence. Cette alternative a lieu de manière séquentielle, sur chacun des mots rencontrés. Dans ce modèle, le processeur syntaxique reçoit une information supplémentaire lui intimant l'ordre d'arrêter ou de continuer son analyse. Cette commande lui vient du contexte dans lequel apparaît la phrase à analyser, donc par le traitement d'un processeur référentiel. Par-là, on aborde une certaine interactivité des processus cérébraux, puisque le processeur syntaxique est contraint par des informations extérieures et préalables. Cependant, cette interactivité reste limitée : le processeur syntaxique n'est pas directement affecté par le matériau qui est en train d'être analysé. Il ne l'est que par ce qui est déjà passé dans le texte et éventuellement par la connaissance du monde, linguistique et extra-linguistique qu'a le lecteur.

4.3.2. Le modèle de représentations lexicales

Shapiro et ses collaborateurs ont montré que quand un verbe était rencontré dans une phrase, toutes les structures que ce verbe est susceptible de générer

sont activées (Shapiro, Brookins, Gordon, & Nagel, 1991; Shapiro, Zurif, & Grimshaw, 1987; Shapiro, Zurif, & Grimshaw, 1989; Rubin, Newhoff, Peach, & Shapiro, 1996). Cette activation rend primordial le rôle du verbe dès la phase primitive de l'analyse syntaxique, puisque le verbe peut indiquer à l'analyseur les différentes structures qu'il est susceptible de rencontrer. Ainsi le verbe [partir] peut s'employer dans plusieurs contextes structurels :

- seul : Je pars.
- [partir + où] : Je pars à la mer.
- [partir + avec] : Je pars avec ma famille.
- Une combinaison [partir + où + avec], avec les deux ordres possibles de construction.

Cette liste n'est pas exhaustive. De surcroît, la présence de [partir] configure le processeur thématique avec les différents rôles thématiques qu'on peut indiquer pour le sujet et les différents compléments. Les travaux de Shapiro ont montré que, dans une tâche de décision lexicale, la difficulté de traitement des propositions augmente selon le nombre de structures possibles ouvertes par les verbes. De plus, ces auteurs insistent sur la nécessité de prendre en compte les différences entre les individus pour comprendre l'analyse effectuée : en effet, ils considèrent que, parmi les structures activées par le verbe, c'est celle qui sera la préférée de l'individu qui sera considérée comme structure possible. Ce n'est que si ce choix présente des incohérences avec le reste de la phrase qu'une ré-analyse démarrera. En conséquence, ces auteurs considèrent que l'analyseur syntaxique est gouverné par les entrées lexicales. Bien entendu, ce modèle présente de grandes incompatibilités avec celui du *garden-path*, du moins dans ces variantes les plus extrêmes qui ne prévoient aucune immixtion de l'expérience individuelle dans la construction syntaxique, sauf en ce qui concerne les variations de la mémoire à court terme. Nous reviendrons sur cet aspect un peu plus tard.

4.4. Les modèles interactifs et parallèles

Ils supposent que les différentes sources d'informations présentes dans la tâche cognitive en cours partagent leurs données à tout moment pour élaborer une interprétation de la phrase. De fait, l'analyseur syntaxique est lui aussi en liaison permanente avec les autres processeurs pour la réalisation de sa tâche, à savoir la construction des attachements syntaxiques de la phrase. Cet aspect implique le parallélisme des processeurs, puisque chacun doit travailler pour pouvoir échanger de l'information avec les autres. En conséquence, des contraintes comme le rôle thématique d'un mot donné ou ses propriétés sémantiques seraient également des sources importantes d'information pour la construction de la structure et du sens d'une phrase (MacDonald et al., 1994; Trueswell, Tanenhaus, & Garnsey, 1994). Les tenants de ces modèles se sont attaqué aux modèles encapsulés en se focalisant sur la recherche des caractéristiques sémantiques et/ou thématiques agissant sur la résolution des ambiguïtés syntaxiques. Par exemple, les travaux de Taraban et MacClelland, observent les temps de lecture de quatre types de phrases, présentant la même attaque mais contenant des violations susceptibles de porter sur trois aspects différents : le rôle thématique, le nom et l'attachement syntaxique.

Attaque de la phrase : Le concierge lave l'aire de stockage		
Type	Fin de la phrase	Violation
1	avec le balai.	Aucune
2	avec le solvant.	Du nom
3	avec le directeur.	Du nom et du rôle thématique
4	avec la odeur.	Du nom, du rôle thématique et de l'attachement syntaxique.

Tableau 3 — Catégories de phrases présentées dans l'expérience de Taraban et MacClelland (1988, 1990).

Dans la phrase de type 1, les trois attentes sont respectées ; dans la phrase 2 l'attachement au verbe est respecté comme l'est le rôle thématique d'instrument, mais le nom est peu probable ; dans la troisième phrase, seul l'attachement au verbe est respecté, alors que ni le rôle

thématique, ni la probabilité du nom ne le sont. La quatrième phrase, quant à elle, ne respecte aucune attente. Les auteurs observent que les temps de lecture pour les phrases de type 3 et 4 sont identiques ce qui indiquerait une absence d'effet du site d'attachement, puisque c'est la seule différence entre le type 3 et le type 4. Les temps de lecture sont plus importants pour les phrases de type 3 que de type 2, suggérant un effet du rôle thématique dans la résolution d'ambiguïtés. De même, les temps de lecture sont plus importants pour les phrases de type 2 que de type 1, suggérant une influence de la plausibilité du nom. D'autre part, dans d'autres expériences, Taraban et MacClelland observent une augmentation du temps de lecture des syntagmes prépositionnels quand ils sont peu probables (mais possibles) avec les attentes créées par l'objet du verbe.

En conséquence, les modèles parallèles, comme le sont les modèles probabilistes à résolutions de contraintes, rejettent le fait que l'analyse initiale de la phrase ne s'appuie que sur des informations syntaxiques. Des analyses concurrentes ont lieu en parallèle, et sont pondérées selon différents critères. Les effets de *garden-path* n'apparaissent que quand la construction dont l'évaluation est la plus élevée est incompatible avec la suite de la phrase. Par exemple, la construction en objet direct est celle qui est *a priori* préférée sur le critère de sa très grande fréquence. Dans un modèle probabiliste, ce sera donc une contrainte très forte. Des effets sémantiques pourront se voir dotés d'une pondération plus importante si, localement, on rencontre une structure (par exemple induite par la grille thématique d'un verbe) qui en possède une plus forte. De plus, ces modèles ne prévoient pas de différences de nature entre la première analyse de la phrase et les éventuelles ré-analyses postérieures qui se produisent en cas de construction erronée. Dans une ré-analyse, les différentes constructions possibles sont ré-évaluées en modifiant les pondérations des différentes alternatives de construction.

On voit donc que l'introduction d'éléments de différentes natures (et non plus uniquement syntaxiques) dans les critères qui permettent de construire la représentation de la phrase éloigne ces modèles de la notion d'encapsulage, et les rendent interactifs par la compétition à laquelle se livrent les différents

composants de l'analyse. Cependant, il est difficile de distinguer expérimentalement, entre les modèles purement sériels et ces modèles parallèles surtout quand les modèles parallèles intègrent des critères structurels dans les alternatives pondérées (Gorrell, 1989). Finalement, la seule chose qu'il semble possible d'expérimenter consiste à déterminer si des informations sémantiques sont susceptibles de conditionner les processus initiaux de construction syntaxique. Ceci est évidemment rejeté par les modèles à syntaxe autonome. Dans le cas où la réponse est affirmative, les chercheurs se demandent quel est le poids de ces informations dans l'établissement de la structure, par rapport, par exemple, à des informations purement syntaxiques (Lipka, 2002).

4.5. La question de la mémoire de travail

Pour valider les modèles que nous venons d'évoquer, la majorité des expérimentations se partagent entre deux grands protocoles :

- l'observation du mouvement des yeux, avec en particulier celle des retours en arrière et des temps de fixations. Ces protocoles sont principalement le fait des partisans des modèles *garden-path* qui prétendent que ce type d'observation permet de distinguer les deux types d'analyse (initiale et réanalyse).
- les techniques d'auto-présentation segmentée³¹ avec une observation des temps d'analyse que les lecteurs s'accordent en fonction du lieu des difficultés. Ces techniques sont plus souvent utilisées par les tenants des modèles plus interactifs.

Ainsi s'ajoutent aux désaccords théoriques des divergences méthodologiques. Pour tenter de passer outre, une approche divergente a été mise en place qui passe par l'observation de la quantité de mémoire de travail des individus,

³¹ Les techniques d'APS se déroulent devant un écran d'ordinateur. Le lecteur commande l'apparition des segments de textes qui ont été découpés au préalable par l'expérimentateur en fonction de ses hypothèses : ce peut être des mots, des syntagmes, des propositions... On enregistre le temps que met le sujet entre deux appuis sur la touche. Ce temps d'exposition est utilisé comme indicateur de l'activité cognitive du lecteur (Zagar, 1992).

laquelle pourrait expliquer les problèmes de compréhension de certaines phrases.

Deux grandes approches sous-tendent les rapports entre mémoire de travail verbal et compréhension de phrase. On peut schématiquement définir le concept de mémoire de travail comme étant un système à court terme dans lequel des fragments d'informations sont simultanément stockés et manipulés pour l'accomplissement d'une tâche (Caplan & Waters, 1999).

- La première approche, initiée par Just et Carpenter, suppose que la difficulté de compréhension est liée à la capacité de la mémoire de travail (Just & Carpenter, 1992; Just, Carpenter, & Keller, 1996; Miyake, Carpenter, & Just, 1995; Miyake et al., 1995). Dans le cas d'une réduction de cette capacité, les processus langagiers vont se ralentir et la compréhension va se détériorer. Cette réduction de capacité peut se produire quand :

- a) La complexité syntaxique de la phrase est élevée.
- b) La taille de la mémoire individuelle de travail est réduite.
- c) Une tâche concurrente est en cours dans la mémoire de travail.

Cette théorie considère donc que la mémoire de travail sert à la fois pour les processus en cours et pour les stockages temporaires d'information de différents ordres. Une de ses conséquences est sans doute la remise en cause d'un cloisonnement informationnel du traitement syntaxique : la capacité plus importante de mémoire de travail³² possédée par certains individus leur permettrait une interaction entre informations syntaxiques et pragmatiques, de sorte que leur processeur syntaxique ne saurait être considéré comme

³² Cette taille est le plus souvent mesurée à partir de la tâche de Daneman et Carpenter (Daneman & Carpenter, 1980). Elle est régulièrement adaptée dans différentes langues ou pour différents âges. On en trouve un exemple en français dans l'ouvrage « Apprentissage de la lecture et compréhension d'énoncés » page 185-186 :

« Des séries de deux à cinq phrases sont présentées successivement à l'enfant, chaque série comportant 3 essais. Pour chaque essai, les "n" phrases qui le composent sont présentées à la suite l'une de l'autre. L'enfant doit retenir le dernier mot de chaque phrase de l'essai et les rappeler dans l'ordre, à l'issue de la présentation de toutes les phrases de l'essai (...) Deux questions sont posées à la fin du rappel sur une des phrases de la série. Quand tous les mots sont rappelés dans l'ordre exact de présentation et que les réponses aux deux questions sont correctes, l'essai est considéré comme réussi. La tâche s'interrompt quand tous les essais d'une série sont échoués. La taille de l'empan [de mémoire de travail] se calcule de la manière suivante : l'empan est soit la taille de la dernière série pour laquelle au moins deux essais sont réussis, soit la taille de la dernière série pour laquelle un seul essai est réussi – 0,5. » (Lecocq et al., 1996)

encapsulé. De plus, la plus grande capacité de mémoire de travail de certains individus leur permettrait de maintenir activée simultanément plusieurs interprétations dans les phrases comportant des ambiguïtés syntaxiques. Le modèle de Just, prenant en compte la capacité de la mémoire de travail suppose donc que, devant une phrase ambiguë, les différentes constructions syntaxiques possibles sont construites et restent activées dans la mémoire de travail jusqu'à la résolution de l'ambiguïté par la suite de la phrase. Cependant, la construction ayant la plus forte probabilité, parce qu'elle est plus simple ou parce qu'elle est plus fréquente, est privilégiée dans l'analyse. En cas d'erreur dans la construction, les sujets à capacité de mémoire de travail élevée devraient donc pouvoir réactiver efficacement les autres constructions syntaxiques construites, alors que les sujets à capacité plus faible les auraient déjà oubliées, causant par là une compréhension moindre ou des temps de traitement plus longs. C'est ce qu'ont montré les travaux de King et Just (King & Just, 1991) (Just & Carpenter, 1992).

- La deuxième approche est celle de ressources du traitement langagier séparées (Caplan & Waters, 1999; Caplan & Waters, 1999; Waters, Caplan, & Yampolsky, 2003). Selon cette théorie, il existe au moins deux ressources spécialisées : une première servant dans les processus tels que la construction syntaxique et l'attribution des rôles thématiques et une deuxième servant dans les tâches de contrôle verbal par exemple lorsqu'il s'agit de maintenir activées en mémoire des informations sémantiques. Les processus se servant du premier type de ressource seraient activés obligatoirement après la présentation d'une phrase. Le traitement d'une phrase complexe est supposé faire une demande plus importante à ces ressources qu'une phrase à structure simple. En conséquence, cette théorie postule que le processeur syntaxique est indépendant à la fois de la présence de tâches concurrentes et de la mesure de l'empan de mémoire de travail, alors que les travaux de Just postulent quant à eux l'inverse. Aussi, Caplan et ses associés ont étudié l'éventuelle spécialisation d'une mémoire de travail syntaxique en faisant varier les sources d'information : les différences selon les individus entre mémoire de travail et efficacité à traiter des phrases complexes, les processus syntaxiques chez les patients ayant une faible mémoire de travail, ou encore l'étude de patients

aphasiques. Pour ces auteurs, les résultats expérimentaux les amènent à conclure à une spécialisation de la mémoire de travail dévolue au langage pour la construction des structures syntaxiques de la phrase et pour l'utilisation de cette mémoire de travail dans l'édification du sens de la phrase. Cette mémoire de travail serait différente de celle utilisant le sens de la phrase dans d'autres fonctions. En conséquence, ces auteurs présentent une théorie de la division de la mémoire de travail langagière.

4.6. Conclusion

Koriat, un des plus importants concepteurs du modèle structural de lecture, précise expressément que le modèle qu'il a initié ne peut en aucune façon être associé à l'existence d'un analyseur syntaxique encapsulé tel que nous avons pu le décrire plus haut (Musseler et al., 2000). Il considère que les résultats présentés dans les différents articles supportant le modèle structural de lecture s'expliqueraient davantage par un modèle probabiliste à résolution de contraintes (MacDonald et al., 1994; Trueswell et al., 1994; MacDonald et al., 1994). Ce qui pose le plus de problème dans les modèles s'appuyant sur la notion de cloisonnement informationnel, c'est sans aucun doute le refus d'envisager que des informations de natures différentes soient à la source de décisions de nature unique : ces modèles ne conçoivent pas que le contexte puisse influencer sur une construction locale. Il faut comprendre contexte aussi bien au sens large, c'est à dire la connaissance de la langue en général et la capacité à prévoir l'utilisation de structures bien particulières en fonction du type de discours dans lequel elles apparaissent³³, qu'au sens de contraintes beaucoup plus locales. Les modèles à résolutions de contraintes permettent à ces apports multiples d'interagir (avec des pondérations différentes selon les cas) dans chacune des tâches que doit résoudre le lecteur. Et donc en particulier dans celle de construire le cadre structurel de la phrase.

Les travaux menés jusqu'ici par Koriat reposent plus spécifiquement sur le rôle des mots fonctionnels dans la création de la structure de la phrase. Dans le

³³ Le livre de Bronckart, *le fonctionnement des discours* en donne un exemple éloquent (Bronckart et al., 1985)

chapitre que nous avons consacré à la linguistique comme dans celui-ci, nous avons vu l'importance des grilles thématiques et des sous-catégorisations dans la structure des phrases³⁴. Rien n'empêche de faire l'hypothèse que ces informations sont également utilisées pour anticiper et construire leur cadre syntaxique. Si, dans le cas des mots à rôle syntaxique, leur placement à l'arrière-plan cognitif une fois le squelette de la phrase construit, permet d'observer un plus fort oubli des lettres les constituant, il reste à imaginer les protocoles expérimentaux pour mesurer le rôle des sous-catégorisations et des grilles thématiques dans la construction de l'architecture de la phrase. Car les sous-catégorisations impliquent les verbes et les verbes portent également une information de sens qui les empêche d'être relégués à l'arrière-plan.

Le seul impératif du modèle de Koriat est que l'analyse de la structure syntaxique précède impérativement l'analyse sémantique. Et les observations faites jusqu'à maintenant insistent sur la grande utilité des mots fonctionnels dans cette analyse.

³⁴ Nous avons essayé de développer cette partie précédemment au chapitre 2

Chapitre 5. Les observations neurophysiologiques.

5.1. Les lésions cérébrales

Pendant des décennies, l'étude des zones cérébrales à l'œuvre dans les traitements langagiers a été fondée sur la mise en rapport de troubles décrits chez des malades et de la dissection de leur cerveau après leur mort. En 1861, Paul Broca dissèque le cerveau d'un aphasique que le personnel de l'hôpital avait surnommé Tan, la seule syllabe qu'il pouvait prononcer. Broca découvrit un kyste qui produisait une lésion dans l'hémisphère gauche, comme pour les autres cas d'aphasiques dont il put, par la suite, disséquer le cerveau. Plus précisément, le cerveau de Tan était lésé dans les régions situées juste au-dessus de la scissure de Sylvius.

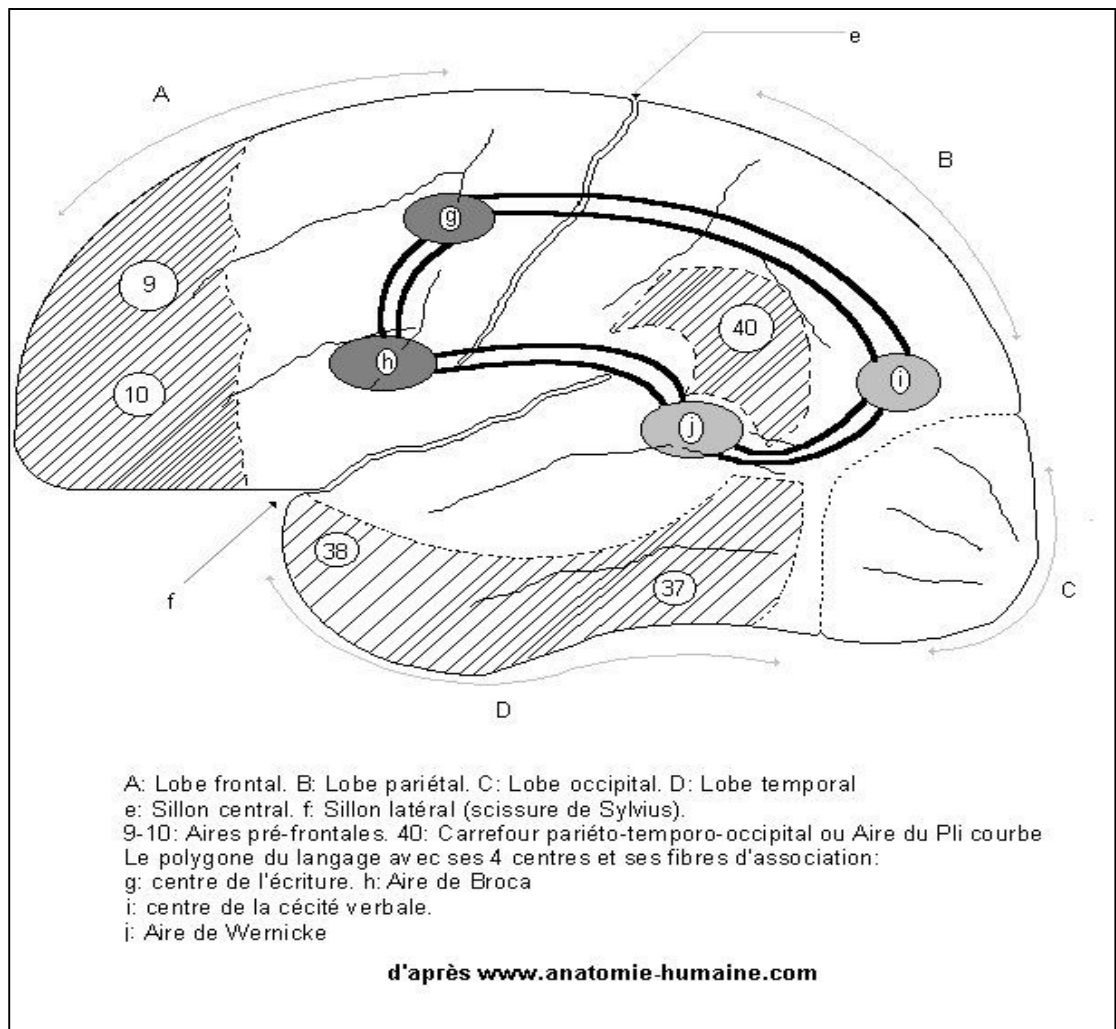


Figure 25 — Aires associatives principales et polygone du langage.

Cette région se nomme maintenant l'aire de Broca et est reconnue entretenir des rapports étroits avec le fonctionnement du langage. Les patients qui ont une atteinte de cette zone souffrent de ce qu'on appelle une aphasie de Broca. Il en est ainsi de M. Ford qui eut une attaque à l'âge de 39 ans suivie de sévères troubles du langage : il omettait les terminaisons en *-ed* et *-s* et les mots grammaticaux comme *the* ou *or*. Il comprenait bien les questions comme

« *Est-ce qu'une pierre flotte sur l'eau ?* »

mais pas celle nécessitant un traitement grammatical comme

« *Le lion a été tué par le tigre ; lequel est mort ?* ».

Hormis ces troubles grammaticaux, les neurologues décrivaient M. Ford comme alerte et attentif et estimaient que toutes les fonctions intellectuelles qui n'étaient pas étroitement liées au langage étaient préservées (cette description est issue de (Pinker, 1999)).

L'aire de Broca touche la zone cérébrale consacrée au mouvement des mâchoires, des lèvres et de la langue. De ce fait, les neurologues ont cru pendant des décennies que l'aire de Broca était engagée dans la production du langage. D'après Pinker, cette aire est plutôt impliquée dans le traitement de la grammaire en général. Il est évident qu'un défaut dans la grammaire affecte très visiblement la production langagière, alors que la compréhension peut rester performante grâce aux signes extralinguistiques qui entourent la parole et/ou à cause de la connaissance pragmatique du monde : la phrase

« *Le chien a mordu l'homme* »

peut être comprise si l'on sait qu'un chien mord plus souvent un homme que l'inverse. En revanche, des phrases comme

« *La fille que la nourrice pousse est grande* »

nécessite, pour être comprise un traitement syntaxique complexe, que les aphasiques grammaticaux de Broca ne peuvent mener à bien.

Nous ne résumerons pas ici un siècle et demi de recherches et d'observations sur les différentes aphasies, d'autant que la réalité est toujours plus complexe que ce que les neurologues croient en avoir saisi. D'autres aires sont

impliquées dans les aphasies telle celle de Wernicke (l'aphasie de Wernicke donne un flot de paroles plus ou moins grammaticales comprenant beaucoup de néologismes et de substitutions lexicales), et les aires voisines de l'aire de Broca ou encore la substance blanche qui la connecte avec d'autres aires cérébrales.

Notons toutefois puisqu'il se rapporte directement à la lecture, ce cas très intéressant décrit par Andreewsky d'alexie agrammaticale. Ce malade, à l'instar de M. Ford, ne peut lire que les mots lexicaux comme les noms ou les verbes et est incapable de lire les mots grammaticaux comme les articles ou les conjonctions. Ainsi, pour la phrase

« la dame va chez le coiffeur »

il lira *dame va coiffeur*. Si on demande au malade de lire isolément les mots de la phrase

« le car ralentit car le moteur chauffe »

il lira *car, ralentit car, moteur chauffe*. Si on lui demande de lire la phrase, il lira

car ralentit moteur chauffe

omettant le second *car*. Alors que la reconnaissance consciente et la dénomination des mots grammaticaux sont impossibles chez ces alexiques agrammaticaux, le processus de lecture montre qu'il existe ce que Andreewsky appelle un filtre syntaxique qui procède à un traitement différencié des mots en fonction de leur nature : ce filtre est non seulement inconscient mais aussi appliqué préalablement à la dénomination du mot (Andreewsky & Seron, 1975; Andreewsky, 1980).

Les lésions cérébrales ont donc joué un grand rôle dans la découverte des mécanismes cérébraux. Certains patients atteints d'une aphasie de Broca, outre leurs difficultés dans l'usage des mots, peuvent avoir du mal à former des propositions bien construites ou des phrases grammaticalement correctes (Changeux, 2002), surtout si elles comportent des propositions subordonnées. D'autres malades, atteints d'aphasie de Wernicke sont capables de produire un ersatz de langage, même s'il est dépourvu de toute syntaxe. On observe aussi des malades ne pouvant plus ranger les mots dans l'ordre correct de la phrase,

et d'autres qui, par exemple, terminent les mots de manière anormale. Ce qu'il faut donc retenir, c'est *qu'il existe des voies neurales impliquées de manière sélective dans le traitement syntaxique* (Changeux, 2002).

5.2. Apports des nouvelles techniques d'imagerie cérébrale

Malgré tout, assez peu de malades peuvent être observés et une attaque cérébrale ne présente que rarement une seule zone touchée à laquelle on pourrait par la suite affecter un rôle particulier. Les techniques d'imagerie cérébrale permettent, depuis une vingtaine d'années, d'observer les cerveaux normaux en fonctionnement et prennent ainsi le relais de l'observation des lésions cérébrales. Elles permettent d'observer, en trois dimensions, les zones activées par une activité et de comprendre les réseaux qui se créent lors de cette réalisation.

On distingue aujourd'hui trois grandes techniques d'imagerie cérébrale.

5.2.1. L'IRMf : IRM fonctionnelle

Lors de l'examen par IRM (Imagerie par Résonance Magnétique, également appelée Résonance Magnétique Nucléaire (RMN)), le patient est placé à l'intérieur d'un tunnel où règne un champ magnétique très intense (de même nature que celui qui oriente l'aiguille d'une boussole, mais dont l'intensité est environ 10000 à 15000 fois plus forte). Une fois placé dans ce champ magnétique, l'ensemble des atomes d'hydrogène (principalement ceux des molécules d'eau) qui composent notre organisme se comporteront comme des milliards de boussoles, et s'orienteront tous dans le même sens. Durant de très brefs intervalles, la machine enverra différents signaux (dits signaux de radio fréquence) afin de perturber l'orientation de ces atomes. Ces milliards d'atomes "perturbés" reviendront à leur "position d'origine" dès que possible. En revenant à leur état "initial", ces atomes, en fonction de leur propriété et de leur voisinage, émettront à leur tour certains signaux électromagnétiques. Ce seront ces derniers qui seront à l'origine des images générées par l'IRM. En d'autres termes, et d'une manière un peu simpliste, l'IRM est en mesure de

fournir des images à partir de la densité des tissus en atomes d'hydrogène (densité protonique), donc d'établir une cartographie de la distribution anatomique de l'eau et de la graisse dans les organes. En modifiant certains paramètres d'acquisition des images, il sera possible de refléter "les variations de l'apport sanguin". Ces variations de débit dans le cerveau étant reliées à l'activité des neurones, il devient possible de les "voir" fonctionner. Le domaine de l'IRMf s'intéresse exclusivement au cerveau, et reste du domaine de la recherche en neurologie, en neuro-psychologie et en neuro-psychiatrie. Les images IRMf sont des cartographies de la distribution anatomique des variations du débit sanguin (augmentation ou diminution) au niveau de certaines régions cérébrales, chez un sujet que l'on soumet à une tâche sensitive ou motrice mettant en jeu les populations neuronales de ces régions. Elles sont donc réservées actuellement à l'exploration cérébrale.

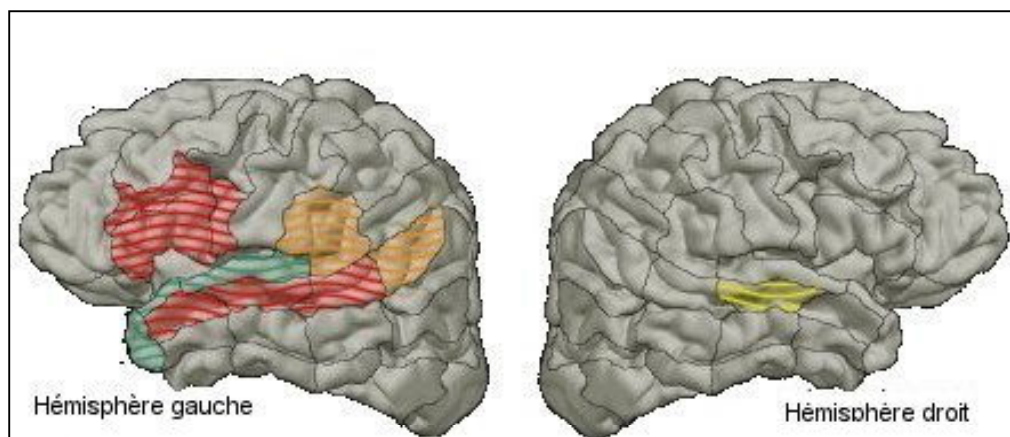


Figure 26 — Régions activées lors de la lecture de phrases en anglais et de la lecture de pseudo-mots. Technique d'IRMf. D'après (Bavaliere et al., 1997)

5.2.2. La Tomographie par Emission de Positons

La Tomographie par Emission de Positons (TEP) est une technique d'imagerie médicale fonctionnelle, également appelée métabolique par opposition aux techniques d'imagerie médicale conventionnelle, dites anatomiques (Radiologie, Scanner, Imagerie par Résonance Magnétique). Elle est fondée sur l'administration d'une substance radioactive émettrice de photons gamma, dont on suit la distribution dans l'organisme avec un détecteur approprié : le

PETSCAN. En TEP, les isotopes utilisés sont de demi-vie³⁵ très brève (de quelques minutes à quelques heures), leur schéma de désintégration est simple : lors de leur désintégration radioactive, ils émettront un positon (l'anti particule de l'électron). Ce positon après un parcours très bref dans la matière (1 à 3 millimètres) rencontrera sa particule sœur, l'électron. De cette rencontre matière-antimatière, résultera une réaction d'annihilation, les deux particules disparaîtront en donnant naissance à deux photons. Ces deux photons seront émis en direction diamétralement opposée (180°) et emporteront avec eux une énergie constante. Le PETSCAN ne détectera ainsi pas directement le positon, mais les deux photons émis lors de son annihilation. Le principe du petscan repose sur la détection simultanée (détection en coïncidence) de ces DEUX photons.

Ce phénomène d'annihilation se répétera plusieurs milliers de fois par seconde : le PETSCAN en enregistrant simultanément ces deux photons tentera de connaître l'origine et le lieu de cette désintégration. Une fois les informations recueillies par l'ensemble des détecteurs, l'ordinateur sera en mesure de reconstruire une série de coupes anatomiques du corps (tomographie), faisant apparaître sous différents contrastes les lieux où a été détecté le maximum d'annihilations.

5.2.3. Les potentiels évoqués

Les potentiels évoqués sont de potentiels cérébraux à l'intérieur de l'activité électrique spontanée du cerveau. Un traitement informatique permet de les soustraire du « bruit de fond » de l'activité électrique normale du cerveau. Cette focalisation sur certains signaux électriques dont la provenance est bien localisée permet d'étudier en détail des événements jusqu'alors noyés dans l'ensemble des informations électriques. Ils apparaissent en fonction des événements que le cerveau traite et sont visualisés le plus souvent par des courbes ayant en abscisse le temps et en ordonnée l'intensité de l'activation électrique. Ils sont caractérisés par leur polarité (positive ou négative), leur

³⁵ Intervalle de temps au bout duquel la moitié des atomes de l'élément s'est désintégrée.

latence (c'est-à-dire par leur relation temporelle au stimulus), leur distribution crânienne (en fonction de la tâche, ils apparaissent à certains endroits fortement, légèrement ou pas du tout).

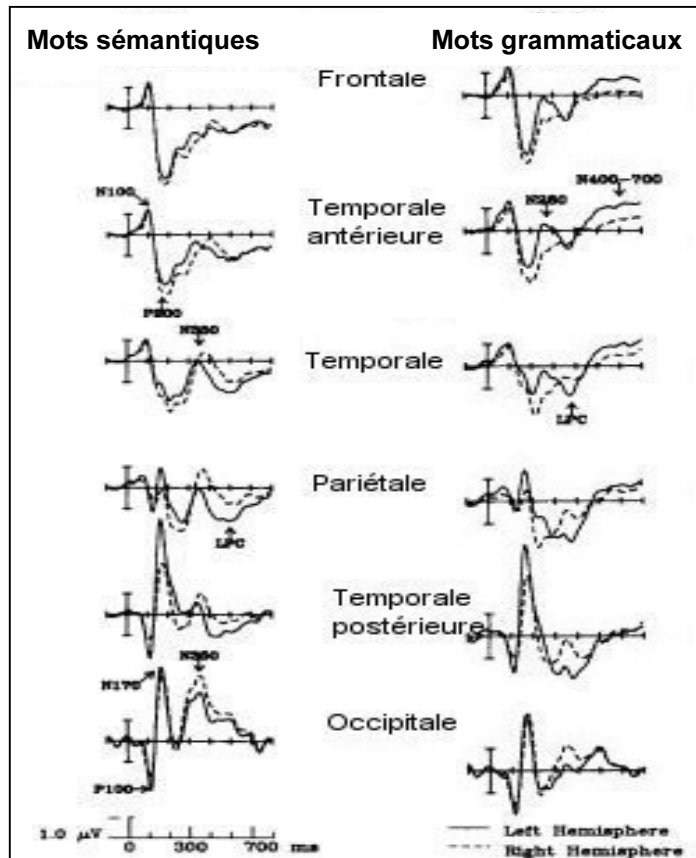


Figure 27 — Potentiels évoqués enregistrés durant la lecture de mots de différentes catégories. D'après (Neville, Mills, & Lawson, 1992)

La technique des potentiels évoqués est de plus en plus utilisée dans les recherches psycholinguistiques puisque les résultats sont continus, non invasifs et surtout parce qu'elle offre une résolution temporelle de l'ordre de la milliseconde. De plus, la nature multidimensionnelle des potentiels évoqués (polarité, latence et distribution) permet de qualifier très précisément différents processus. Ils ont été utilisés de manière de plus en plus systématique dans la dernière décennie pour différencier les processus langagiers dans les domaines de la syntaxe et de la sémantique. La figure 27 montre les potentiels évoqués enregistrés à la surface de la boîte crânienne de sujets lisant des mots à contenu sémantique (verbes, noms et adjectifs) et des mots grammaticaux (articles, conjonctions). La ligne verticale montre l'apparition du mot. Les différentes

courbes permettent d'observer les activités électriques de différentes parties du cerveau en fonction du type d'événement qu'il doit traiter. On peut ainsi observer pour la région temporale antérieure une onde positive pour les mots grammaticaux qui n'existe pas pour les mots sémantiques. A l'inverse, une onde spécifique apparaît dans les aires occipitales pour les mots sémantiques. Ce qu'on observe donc, c'est une activité électrique spécifique à certains stimuli dont les caractéristiques temporelles sont très précises.

5.3. Quelques résultats

Les techniques plus récentes d'imagerie cérébrale apportent bien des précisions aux observations antérieures à partir des lésions cérébrales.

- Par exemple, les travaux des chercheurs américains Dapretto et Bookheimer montrent la dissociation entre syntaxe et sémantique dans la compréhension de phrases (Dapretto & Bookheimer, 1999). Ils ont utilisé des phrases-test où le sujet doit décider s'il y a des différences de sens entre des phrases pourtant très proches. Pour que les différences soient de nature sémantique, on substitue à un des mots de la phrase soit un synonyme, soit un autre mot avec une signification très éloignée. Pour les différences d'ordre syntaxique, on utilise soit la voie passive ou la voie active, soit un ordre des mots différent. Les images cérébrales obtenues par les auteurs montrent une activité cérébrale différente selon le type d'information traité : dans une partie de l'aire de Broca pour les informations syntaxiques et dans la circonvolution frontale antérieure gauche pour les informations d'ordre sémantique. Mais les activités langagières impliquent bien plus de zones cérébrales que les seules aires de Broca ou de Wernicke. Les expériences mesurant les débits sanguins cérébraux de sujets confrontés à des environnements linguistiques très différents le montrent sans équivoque (IRMf). Si on expose des sujets ne parlant que la langue française à des histoires en langue Tamoul, qu'ils ne connaissent donc pas, les images cérébrales ne décèleront une activité que dans les aires auditives primaires et secondaires : le sujet n'a qu'entendu les sons d'une langue qu'il ne comprend pas. Si on expose les sujets à une liste de mots qu'ils

peuvent comprendre, on observe alors l'activation supplémentaire de la circonvolution frontale inférieure gauche. Si les sujets sont confrontés à des phrases comportant des pseudos-mots ou des mots sémantiquement anormaux, on observe, en plus des aires auditives une activation des pôles temporaux. Enfin, l'exposition à une histoire en français, donc enfin compréhensible, offre les images cérébrales les plus spectaculaires : le cerveau du sujet comprenant mobilise le plus grand nombre d'aires cérébrales, et en particulier la région préfrontale gauche (Mazoyer et al., 1997).

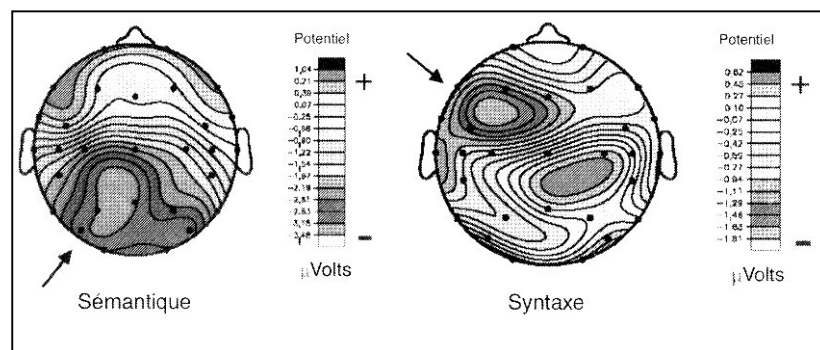


Figure 28 — Aires cérébrales activées lors de traitements sémantiques et syntaxiques

- Changeux considère que l'activation de ces aires préfrontales témoigne de « l'entrée en activité de neurones qui sont capables de relier entre elles de multiples unités de traitements disséminées dans l'ensemble du cortex cérébral et d'inhiber les autres » (Changeux, 2002). Cette mise en relation de territoires cérébraux ne peut se faire, toujours selon Changeux, que par des « anticipations qui permettent de saisir ensemble de nombreuses zones cérébrales ». On revient par là aux deux racines latines du mot "compréhension" puisque l'activité cérébrale de compréhension ne peut s'envisager sans cette activation simultanée de zones topographiquement distinctes, compréhension qui ne peut plus se cantonner dans une activité séquentielle « d'accumulation de mots et de phrases sans ligne de pensée définie ».

- Depuis une vingtaine d'années, les observations cérébrales au moyen des potentiels évoqués sont de plus en plus utilisées pour étudier les différents composants du traitement du langage. Cette technique, nous l'avons vu, est particulièrement utile pour observer les décalages temporels d'activation entre

différentes zones cérébrales. Elle permet de suivre l'activité cérébrale milli-seconde par milli-seconde et d'observer les structures et les interactions des sous-processus à l'œuvre dans les tâches complexes. Parmi les premiers travaux réalisés à l'aide de cette technique, on notera ceux de l'équipe de Bentin. Ils ont montré que les aspects sémantiques du traitement du langage correspondent à une vague d'onde négative (negative-going waveform) qui a lieu environ 400 millisecondes après l'affichage d'un mot, (d'où son nom, « le composant N400 »). Les violations de la cohérence sémantique sont significativement associées à de larges amplitudes de N400. Cette zone semble être le marqueur le plus valide des processus d'intégration sémantique (Bentin, 1987). Les travaux sur les processus cérébraux en rapport avec les traitements sémantiques sont bien sûr de première importance, et ont permis de préciser les aires concernées. Par exemple, Halgren et ses collaborateurs, ont fait lire des phrases se terminant par des mots possibles ou impossibles sur un plan sémantique, en observant les lecteurs à l'aide d'encéphalographies magnétiques³⁶ (Halgren et al., 2002). Comparées à celles conclues par un mot sémantiquement possible, les phrases terminées incorrectement ont activé un large champ magnétique sur l'hémisphère gauche à environ 450 ms. Ces travaux ont permis la mise en évidence d'une multitude de zones proches de celles activant N400, activées par les processus sémantiques : au pic de N400, correspond une activité de l'hémisphère gauche comprenant le cortex temporal antérieur, perisylvien, orbital frontopolaire et dorsolatéral. On voit donc se dessiner un réseau neuronal géographiquement proche et réagissant dans les mêmes intervalles temporels qui semble particulièrement actif dans les traitements sémantiques.

- D'autres études utilisant le même procédé ont cherché à observer différentes zones cérébrales au cours de tâches langagières. On note ainsi que les processus syntaxiques, eux, sont corrélés avec deux zones cérébrales différentes de N400. La première, P600³⁷, est corrélée très fortement avec les anomalies syntaxiques présentées au lecteur, et se manifeste par une onde

³⁶ Technique d'observation cérébrale proche de l'observation des potentiels évoqués, mais s'appliquant aux champs magnétiques et non plus aux champs électriques.

³⁷ aussi nommée vague positive syntaxique (SPS : syntactic positive shift)

positive avec un maximum centro-pariétal au bout d'environ 600 ms ; la deuxième zone réagit de manière beaucoup plus rapide (100 ms) et ne réagit qu'aux fautes de structure syntaxique. Les travaux de Friederici ont porté sur l'interaction qui peut exister entre ces différentes zones (Friederici, Steinhauer, & Frisch, 1999). Le protocole expérimental comportait quatre types de phrase : phrases correctes ; phrases avec une faute sémantique ; phrase avec une faute syntaxique ; phrases avec les deux fautes³⁸. Les contraintes de l'observation par potentiels évoqués obligent à présenter les mots un par un de façon à être certain de prendre les clichés correspondant à un mot. Outre les résultats habituels (N400 activé sur les violations sémantiques ; les deux autres sur les violations syntaxiques), tout l'intérêt de cette expérience réside dans l'étude des phrases comportant une double violation : on observe alors que l'activation de N400 n'a pas lieu (pas plus que P600), comme si l'activation antérieure de la zone à 100 ms qui, rappelons-le, réagit uniquement aux violations syntaxiques bloquait tous les traitements ultérieurs. Dans le même paradigme, les travaux de Gunter se sont focalisés, à partir de trois expériences différentes, sur l'observation des violations syntaxiques et sémantiques (Gunter, Stowe, & Mulder, 1997). Les potentiels évoqués ont également montré que les violations sémantiques produisent une réponse N400, tandis que les violations syntaxiques apportent une réponse négative très rapide (150 et 300 ms) puis une réponse P600. Aucune interaction entre les potentiels précoces syntaxiques et ceux obtenus par N400 n'a été observée, ce qui évoque encore une fois l'indépendance des traitements. L'ordre temporel des mots dans la phrase semble être traité à la fois par les processus précoces (observés par les potentiels à 100 ms) et par les processus sémantiques (N400), le tout étant clôturé par une analyse syntaxique *a posteriori*. Cette double analyse syntaxique se retrouve dans d'autres travaux comme, par exemple, ceux de

38

Phrase correcte	Das Haus wurde bald gebaut. La maison était bientôt construite.
Violation sémantique	Der Priester wurde bald gebaut. Le prêtre était bientôt construit.
Violation syntaxique	Das Haus wurde vom gebaut. La maison était par le construite.
Violation sémantique et syntaxique	Der Priester wurde vom gebaut. Le prêtre était par le construite.

D'après (Friederici et al., 1999).

Hagoort, portant sur des phrases présentées en néerlandais (Hagoort, Wassenaar, & Brown, 2003) ou de Friederici (Friederici, Hahne, & Mecklinger, 1996). Toujours en enregistrant les activités cérébrales à l'aide des potentiels évoqués, les auteurs ont présenté aux sujets des phrases correctes, d'autres comportant des erreurs sur une catégorie syntaxique requise ou encore avec des ambiguïtés sur une catégorie syntaxique. Ces phrases sont présentées soit oralement soit par écrit. Pour les phrases présentées oralement, on observe une onde négative antérieure gauche pour les violations syntaxiques mais pas pour les ambiguïtés. Pour les deux catégories, on observe une onde positive centro-pariétale. Quand les phrases sont présentées à l'écrit, les mêmes aires cérébrales sont activées au même moment que précédemment. Ces données suggèrent que l'analyseur syntaxique obéit à un modèle de fonctionnement à deux étapes : une première étape, très précoce, où la structure est construite et une deuxième où prennent place les différents rôles thématiques et si nécessaire, une ré-analyse syntaxique. La première étape de l'analyseur syntaxique semble liée à des ondes négatives dans les zones antérieures gauches du cerveau, et la deuxième est corrélée avec des ondes positives plus tardives dans les zones centro-pariétales. Friederici et son équipe ont précisé, quelques années plus tard, les différentes zones concernées (Friederici, Ruschemeyer, Hahne, & Fiebach, 2003) : tout indique que les processus sémantiques et syntaxiques sont portés par un réseau neuronal temporo-frontal, avec des aires distinctes spécialisées dans les processus syntaxiques et sémantiques.

5.4. Conclusion

Les travaux de recherche, extrêmement nombreux, semblent aller très majoritairement dans la même direction, à savoir l'autonomie des processus syntaxiques et sémantiques, autonomie observée à la fois par les localisations cérébrales et les temporalités différentes. Il semble donc que les observations à l'aide de l'imagerie cérébrale corroborent les modèles explicatifs que nous avons décrits plus haut, notamment cette observation de la double activation cérébrale en liaison avec la syntaxe, à l'œuvre dans l'établissement de la

structure de la phrase pour la première et dans la réanalyse pour la seconde. Toutefois, un certain nombre de remarques méthodologiques s'imposent.

Dans les expériences d'imageries cérébrales, on a recours la plupart du temps à une présentation mot à mot (automatique ou gérée par le sujet) des phrases lues, souvent avec un délai important entre deux mots de façon à pouvoir affecter un événement cérébral (une image) à un mot. D'autres types d'expériences demandent un jugement sur la grammaticalité d'une phrase mais peu de travaux supposent que l'effectuation de ce jugement est susceptible d'apparaître dans les phénomènes cérébraux observés, alors que les chercheurs qui le supposent trouvent que la tâche à accomplir influe grandement sur les images obtenues (Friederici, Opitz, & von Cramon, 2000). D'autres encore demandent de répondre à une question après la présentation du stimulus, obligeant les sujets à une opération supplémentaire incluant la mémoire à court-terme (Kang, Constable, Gore, & Avrutin, 1999). Il semble bien évident que ces contraintes expérimentales nous éloignent d'une situation de lecture fluide où le sujet n'a d'autres contraintes que de comprendre le texte. Certes, ces remarques n'invalident pas les conclusions proposées par les différents auteurs, mais il nous faut garder en mémoire que les protocoles d'observation des mécanismes à l'œuvre dans les activités cérébrales complexes influent sur les processus eux-mêmes, donc sur ce qui est observé.

Sans doute plus gênante encore est la présentation des mots eux-mêmes. Dans la lecture normale d'un texte, l'œil est à même de traiter un grand nombre d'informations, pas toujours de manière consciente. Les présentations en APS, ou mot à mot, empêchent de fait l'œil de percevoir les informations. Par ce biais, on n'est plus en train d'étudier la lecture mais de recueillir les effets d'un protocole expérimental mis en place pour étudier la lecture. Encore une fois, ces remarques rejoignent celles de Rayner quand il déclare que toutes les manipulations qui tentent de supprimer les processus travaillant à partir des informations parafovéales ou qui obligent les lecteurs à passer par un point de fixation particulier modifient les processus normaux de lecture (Rayner & Pollatsek, 1989). Cependant, et même si les contraintes de présentation modifient notablement la situation de lecture en interdisant la vision

parafovéale, dont on sait par ailleurs toute l'importance, l'ensemble des résultats convergent pour confirmer la présence du traitement syntaxique dans la temporalité des traitements langagiers.

Les observations de la neuro-imagerie interrogent à leur manière les modèles du fonctionnement cérébral que nous avons exposés au chapitre précédent. De prime abord, on pourrait inférer de la présence temporelle des processus syntaxiques et de leur localisation cérébrale qu'ils sont purement syntaxiques, au sens où ils ne travailleraient qu'à partir d'informations syntaxiques. Dans ce cas, nous opterions pour des modèles informationnellement cloisonnés et sériels. Toutefois, même si ces observations issues de la neuro-imagerie permettent sans conteste de préciser que certaines zones cérébrales sont particulièrement actives lors du traitement de la syntaxe, (lors de l'établissement du cadre structural de la phrase parcourue, cf. l'onde N100), d'autres travaux d'imagerie cérébrale témoignent que ces mêmes zones sont également actives (dans les mêmes limites temporelles) lors de tâches non linguistiques mais supposant l'établissement, la création d'une structure en général (Hoen & Dominey, 2000). En conséquence, les zones qui semblent être à l'œuvre dans cette création structurale sont susceptibles de travailler à partir d'informations de types très différents, linguistiques (à l'intérieur desquels on peut différencier au moins sémantique et syntaxique) ou pas. Dans ce cas, nous rejoindrions les descriptions des modèles interactifs. Ce que nous rapportions de Changeux au début de ce chapitre va dans ce sens, quand il insistait sur la nécessaire mise en relation de territoires cérébraux qui ne peut se faire que par des « anticipations » qui permettent de « saisir ensemble » de nombreuses zones cérébrales. On est donc bien loin de la notion d'encapsulation puisque le matériau à partir duquel ce réseau spécialisé dans les réponses syntaxiques travaille, ne semble pas être de nature exclusivement syntaxique (Hoen et al., 2003). Ce que nous suggèrent alors ces données, c'est plutôt l'existence d'une capacité, inhérente au système cérébral, à créer très rapidement un cadre structurel dans lequel des informations vont s'intégrer par la suite. Ce type de fonctionnement est brillamment décrit par Berthoz quand il s'intéresse à ce qu'il nomme la machine proactive (Berthoz, 1997). On le voit, les informations recueillies par l'imagerie cérébrale sont cohérentes avec le

modèle de Koriat. Quand un lecteur commence une nouvelle phrase, sa première activité, c'est de construire le cadre structurel qui sera rempli, dans un second temps, par des informations d'ordre sémantique. Il est donc tout à fait normal d'observer que les zones cérébrales s'activant les premières sont celles habituellement décrites pour participer en général à l'élaboration d'une structure et, dans ce cas, au repérage d'informations syntaxiques.

Deuxième partie :
Approches expérimentales

Chapitre 6. Hypothèses

Plusieurs hypothèses demeurent concurrentes dans l'élucidation du rôle de la syntaxe dans les processus de lecture. Si le modèle de Koriat, élaboré pour rendre compte de l'effet d'oubli de lettres, semble faire largement écho aux observations du fonctionnement cérébral, dans la mesure où il prend appui sur la préséance de l'analyse syntaxique, il reste un grand nombre de questions en suspens.

Les hypothèses que nous allons présenter et les dispositifs expérimentaux mis en place pour tenter d'y répondre sont les fruits de l'état de la question que nous avons présentée précédemment. Ce n'est qu'après les avoir testés que nous engagerons un travail expérimental de nature pédagogique et qui fera l'objet de la troisième partie de ce travail.

Nous retiendrons donc quatre hypothèses dans la mesure où les réponses qui leur seront données peuvent influencer sur les options d'un enseignement renouvelé de la lecture.

6.1. Première hypothèse

Le modèle de Koriat suppose que l'inégal oubli de lettres est généré par le rôle des mots dans l'organisation de la phrase. A l'inverse, sous l'influence d'Alice Healy, cet effet s'expliquerait au niveau de l'accès au mot. Dans leurs publications les plus récentes, ces deux auteurs suggèrent que les deux causes sont susceptibles d'intervenir dans l'effet observé d'oubli de lettres. Dans ces conditions, il nous semble nécessaire d'approfondir l'étude de l'influence respective des deux modèles et de leur rapport.

Ces aspects ne sont pas sans incidence sur la pédagogie de la lecture. Toutes les variantes du modèle de Healy insistent sur l'importance de la familiarité du mot dans sa reconnaissance. Cette familiarité s'acquiert par une fréquentation

régulière des textes et la rencontre, en contexte, des différents items lexicaux. Si la fréquence est l'élément le plus important, l'oubli de lettres viendra donc comme une conséquence de la pratique de la lecture. Si, au contraire, la construction des cadres syntaxiques était l'élément le plus prégnant, on pourrait alors se demander s'il ne serait pas utile de développer une conscience syntaxique chez les apprentis lecteurs.

Nous étudierons l'influence et le rapport de la fréquence et du rôle syntaxique des mots dans l'effet d'oubli de lettres grâce à deux modalités expérimentales :

- En mettant les deux effets en concurrence dans un même modèle explicatif.
- En observant les effets du désordre dans un texte, puisque les mots y gardent leur fréquence mais y perdent leur rôle.

La deuxième expérimentation du chapitre 8 testera l'hypothèse nulle n°1 : le taux d'oubli de lettre dans les mots dépend plus de la fréquence des mots que de leur rôle syntaxique dans la phrase.

6.2. Deuxième hypothèse

La deuxième question est celle de la taille des unités syntaxiques construites lors de la lecture. Si Koriat a débuté ses travaux en envisageant que ces unités devaient être de taille assez réduite (de l'ordre d'un syntagme nominal), des travaux plus récents (Musseler et al., 2000) ont montré qu'elles pouvaient être plus longues. La question est d'importance puisqu'elle éclaire la taille des unités que les processus cérébraux sont susceptibles d'anticiper, c'est-à-dire, comme l'explique Berthoz, la taille des unités que le cerveau peut combiner et avec lesquelles il peut « jouer » à l'avance. Autrement dit, le lecteur est-il tributaire d'une prise d'informations séquentielle et d'une construction syntaxique progressive de la phrase qu'il parcourt ou, au contraire, peut-il construire a priori des cadres syntaxiques étendus dont il va, par la suite, vérifier la pertinence ?

La présentation des travaux sur la syntaxe nous a permis de mieux cerner la notion de constituants et de projections maximales de têtes (fonctionnelles ou non) commandant ce constituant. On a vu également que ces constituants ne se limitaient pas, d'un point de vue théorique, à ces projections maximales de tête lexicale : le fonctionnement est le même sur des unités plus vastes comme une phrase, à travers le concept de syntagme flexionnel, lequel peut être le complément d'un complémenteur.

Nous rappelions plus haut l'importance de ces mots fonctionnels dans l'architecture de la théorie syntaxique. Pour Pinker, ce sont des « fragments de grammaire cristallisée » dans la mesure où ils commandent et déterminent des syntagmes plus grands dans lesquels s'insèrent les syntagmes nominaux, verbaux et adjectivaux. Pinker considère qu'ils fournissent un échafaudage pour la phrase (Pinker, 1999). En conséquence, est-il possible qu'on observe une différence d'oubli entre ces fragments cristallisés, dans la mesure où ils ne « servent » pas de la même façon l'échafaudage de la phrase ? Les mots qui introduisent les syntagmes permettent d'anticiper une taille des unités phrastiques qu'ils commandent : un article introduit nécessairement un syntagme nominal dont la forme simple reste assez courte ; une préposition introduit un syntagme prépositionnel dont la taille standard reste également réduite même si elle est plus importante que celle du simple syntagme nominal ; enfin la liaison de différentes propositions au sein d'une même phrase par la subordination ou la coordination révèle une organisation de la structure de la phrase à plus grande échelle.

Cette recherche de différences d'oubli au sein de la classe fermée des mots fonctionnels rejoint une préoccupation de Koriat (Koriat & Greenberg, 1996b). D'après le modèle de lecture structurale, le fait d'oublier des lettres dans un article montre que le lecteur s'en est servi pour construire l'unité syntagmatique qui y correspond ; s'il oublie une lettre dans une préposition c'est parce qu'il s'en est servi pour construire un syntagme prépositionnel et s'il oublie des lettres dans des conjonctions de subordination, c'est qu'il s'en est servi pour construire une longue unité subordonnée propre aux phrases complexes. Ainsi, l'observation de différences d'oubli à l'intérieur de la

catégorie des mots à rôle syntaxique, pourrait renseigner sur la taille des unités construites par le lecteur, donc sur le processus de lecture lui-même.

La première et la deuxième expérimentation du chapitre 8 testeront donc l'hypothèse nulle n°2 : la nature grammaticale des mots fonctionnels n'a aucun effet sur le taux d'oubli de lettres.

6.3. Troisième hypothèse

Nous situons cette hypothèse à la croisée des chapitres sur la vision et sur le modèle de Koriat. Le chapitre sur la vision nous a permis de mieux cerner les différentes informations visuelles avec lesquelles le cerveau est amené à travailler. Nous rappelions d'une part l'importance de la vision parafovéale dans les processus de lecture et d'autre part la grande part des processus inconscients dans la perception, tout en insistant sur le fait relativement nouveau, que ces informations inconscientes servent de supports à des manipulations cognitives (Dehaene et al., 1998; Naccache & Dehaene, 2001).

Mais la question du rôle des informations parafovéales dans l'établissement des cadres syntaxiques reste entière. Le lecteur serait-il capable de repérer la structure d'une phrase inconsciemment perçue ? Koriat a émis cette hypothèse (Koriat & Greenberg, 1996b) mais n'a jamais mis en place de protocoles pour la vérifier. Un lecteur peut-il construire, au moyen d'informations obtenues par sa vision périphérique, une hypothèse sur le cadre syntaxique de la phrase ? Plus précisément, peut-il mettre en relation une perception inconsciente avec la connaissance qu'il a du fonctionnement de la langue, ici sur un plan purement syntaxique ? Cette construction instantanée et inconsciente d'un cadre syntaxique permettrait d'éclairer le modèle de lecture structurale de Koriat. En effet, deux grands schémas peuvent expliquer la construction des cadres syntaxiques :

- La construction se fait à partir du croisement de ce que le lecteur a déjà rencontré dans le texte et des connaissances qu'il a sur la syntaxe de la langue.

Le lecteur serait ici dans une exploitation de l'amont du texte et dans ce cas, les informations parafovéales sont inutiles.

- La construction se fait à partir du croisement des informations perçues parafovéalement et de leur mise en relation avec les connaissances que le lecteur a sur la syntaxe. Cette explication repose avant tout sur les différents travaux que nous avons présentés au cours du chapitre 3 sur le rôle de la vision parafovéale dans la lecture.

Si le premier schéma est le seul avéré, alors le programme pédagogique devrait s'appuyer sur des inférences à partir du connu pour construire le squelette de ce qui va être lu. Si le deuxième l'est également, alors le programme pédagogique devra chercher aussi à exploiter au mieux les informations parafovéales et leur utilisation dans l'établissement du cadre syntaxique.

L'expérimentation relatée au chapitre 10 testera l'hypothèse nulle n°3 : les informations parafovéales n'interviennent pas pour prédire la structure syntaxique.

6.4. Quatrième hypothèse

Comme nous l'avons remarqué à plusieurs reprises dans la première partie, la quasi totalité des recherches portant sur l'observation des processus syntaxiques chez les lecteurs ne différencie pas les sujets en fonction de leur niveau de lecture, que ce soit par la vitesse de lecture ou par la compréhension. On peut trouver dans la littérature des mises en rapport de différentes habiletés (donc aussi syntaxiques) et du niveau de lecture (Abrahamsen & Shelton, 1989; Bentin, Deutsch, & Liberman, 1990; Gaux & Gombert, 1999). Cependant, la littérature reste peu abondante pour ce qui est de l'observation de lecteurs normaux et adultes. Un des rares exemples montre que les bons compreneurs sont plus performants que les compreneurs moyens dans une tâche de jugement syntaxique qui leur demande de décider si deux mots présentés ont la même fonction grammaticale dans une phrase, mais qu'il n'y a pas de différence entre eux devant une tâche de jugement sémantique (Cupples

& Holmes, 1992). Cependant, les tâches de jugements syntaxiques sont des prises de décision *a posteriori* sur le rôle de mots et ne sont donc en rien assimilables à la construction, au cours de la lecture, de la structure syntaxique de la phrase.

Nous pouvons supposer qu'un meilleur lecteur élabore des cadres syntaxiques de manière plus performante qu'un lecteur moins habile. D'après le modèle de Koriat, une fois le cadre construit, le lecteur le garnit d'informations sémantiques sur lesquelles se focalise alors son attention. Le lecteur, dont la construction structurale est plus laborieuse, dans la mesure où elle est moins "immédiate", devrait mener de front (de manières séquentielles) les deux processus (syntaxiques et sémantiques) et donc consacrer moins d'attention au sens de ce qu'il lit. Un élément de réponse pourrait être avancé s'il était montré à travers une épreuve d'oubli de lettres, que, meilleur lecteur on est, plus on oublie les lettres dans les mots qui organisent la syntaxe de la phrase, parce que la structure aurait déjà été construite et que tout le travail de lecture porterait alors sur l'intégration des informations portant le sens de la phrase.

Si le lien entre habileté dans la compréhension et différentiel d'oubli plus important entre les mots à rôle syntaxique et les mots à rôle sémantique était confirmé, alors il deviendrait indispensable, pour le pédagogue, de vérifier si, en retour, un entraînement spécifiquement conçu pour développer l'habileté dans la construction première des cadres syntaxiques bénéficie au lecteur, en terme de compréhension et/ou de vitesse de lecture.

L'expérimentation relatée au chapitre 9 testera l'hypothèse nulle n°4 : la différence d'oubli de lettres entre les mots organisant la syntaxe et les autres n'est pas liée au niveau de compréhension et/de vitesse des lecteurs.

Le travail sur cette quatrième hypothèse nécessite d'évaluer le niveau de lecture d'une population expérimentale. C'est par cette évaluation que commence notre deuxième partie.

6.4.1. Récapitulatif

	Hypothèse nulle	Dispositif d'étude
1	Le taux d'oubli de lettres dans les mots dépend plus de la fréquence des mots que de leur rôle syntaxique dans la phrase.	Chapitre 8 : expérimentation 2
2	La nature grammaticale des mots fonctionnels n'a aucun effet sur le taux d'oubli de lettres.	Chapitre 8 : expérimentation 1 & 2
3	Les informations parafovéales n'interviennent pas pour prédire la structure syntaxique.	Chapitre 10 expérimentation 1
4	La différence d'oubli de lettres entre les mots organisant la syntaxe et les autres n'est pas liée au niveau de compréhension et/de vitesse des lecteurs.	Chapitres 7 & 9.

Chapitre 7.

Les performances en lecture de la population expérimentale

Dans les chapitres suivants, afin de tester nos différentes hypothèses, nous utiliserons les résultats d'une seule population précisément définie dans différentes épreuves de lecture. Soixante-quatre adultes, tous enseignants, donc au moins titulaires du baccalauréat, vont passer l'ensemble des épreuves que nous allons décrire dans cette deuxième partie. L'âge moyen est de 40,7 ans, avec un maximum de 58 ans et un benjamin de 26 ans. L'âge médian est de 41 ans. La distribution en âge de la population est normale (Test de Kolmogorov-Smirnov $d = 0,071$, $p > 0,20$).

On dénombre 16 hommes pour 48 femmes. Cette répartition 25% / 75% ne saurait étonner dans le monde enseignant où la surreprésentation féminine est la règle. On note que les hommes sont légèrement plus âgés que les femmes avec respectivement une moyenne d'âge de 42,2 contre 39 ans, différence à la limite de la significativité : l'analyse de variance montre une probabilité d'erreur de l'ordre de 12 % ($F(1, 59) = 3,0878$, $p = ,08407$).

7.1. Deux épreuves de lecture

Comme annoncé, nous mettrons en rapport la capacité des lecteurs à construire des cadres syntaxiques avec leur niveau de lecture. Pour évaluer leur niveau en lecture, les 64 individus composant notre population d'étude ont passé 2 épreuves de manière à couvrir des situations de lecture différentes et représentatives de ce qu'on attend d'un lecteur performant.

7.1.1. Epreuve n°1 : Lecture de textes courts

La première épreuve tente d'évaluer la performance dans une forme de lecture très courante, sans doute à l'œuvre dans plus de 60 % des situations ordinaires, celle où il s'agit de prendre connaissance simplement de l'explicite d'un texte, ce qui correspond à ce que l'ex Direction des Études et Prospective du Ministère de l'Éducation Nationale décrivait comme une "compétence approfondie" (Vugalic, 1996).

7.1.1.1. Déroulement

L'épreuve se déroule sur ordinateur dont la résolution d'écran est contrôlée (800*600). Chacun des neuf textes s'affiche, le sujet indique, en pressant une touche du clavier, qu'il en a terminé la lecture et répond alors à une question. Les textes sont diversifiés entre presse, documentaire et fiction et sont d'une taille similaire d'environ 20 lignes et de même niveau de difficulté aussi bien pour le lexique employé que pour la complexité des phrases. L'épreuve s'arrête lorsque 2 réponses correctes sont données à la suite. Les questions portent sur des points explicitement présents dans le texte et sont systématiquement introduites par la formule « *Le texte parle :* » suivie de trois propositions parmi lesquelles une seule est correcte. Les documents (textes et questions) sont présentés dans l'annexe 1.1.1.

7.1.1.2. Variables primaires

La compréhension

On enregistre le numéro du texte auquel le sujet s'est arrêté. Un score 2 signifie qu'on a satisfait à la condition (2 bonnes réponses de suite) dès le deuxième texte. Ce nombre de textes lus est ensuite transformé en un nombre de points s'échelonnant de 0 (pour 9 textes lus, c'est-à-dire quand on n'a jamais réussi à répondre correctement deux fois de suite) à 100 (quand on a

bien répondu aux deux premiers textes présentés)³⁹. Le tableau 4 présente les statistiques descriptives de cette compréhension.

Moyenne	Médiane	Ecart-type	Minimum	Maximum
78,86	71	21,68	0	100

Tableau 4 — Statistiques descriptives de la compréhension. Epreuves de lecture n°1.

La vitesse de lecture

Elle est calculée en faisant la moyenne de la vitesse de chacun des textes lus. Le tableau 5 présente les résultats principaux pour cette variable.

Moyenne	Médiane	Ecart-Type	Minimum	Maximum
18202	17714	5600,58	11759	40128

Tableau 5 — Statistiques descriptives de la vitesse de lecture. Epreuves de lecture n°1.

Plus de 18000 mots/heure de moyenne pour cette population, c'est une performance qui correspond à des empan moyens de 8 à 9 signes lors de chaque fixation. La vitesse la plus faible est à peine supérieure à celle du mot à mot et la vitesse la plus grande correspond, en lecture intégrale, à des empan moyens proches de 4 mots. Notons que le lecteur le plus rapide satisfait néanmoins à la condition des 2 réponses correctes consécutives dès le 4^{ème} texte, ce qui lui vaut un score de compréhension de 71 points qui est égal à la médiane.

7.1.2. Epreuve n°2 : Lecture d'un texte long et plurisémiq

Une seconde épreuve fait travailler sur l'implicite du texte, ce que l'ex DEP dénommait "*compétence remarquable*" et dont semblerait disposer moins de 20 % des élèves entrant en sixième. Il s'agit de franchir ce que dit le texte pour atteindre l'intention de l'auteur et apprécier les moyens qu'il emploie.

³⁹ La transformation linéaire se fait au moyen de la formule
 $Points = (-100/7) * (Nombre\ de\ textes\ lus) + 128,5714$.
Cette formule correspond à la courbe droite passant par les points (9 textes lus, 0 point) et (2 textes lus, 100points).

7.1.2.1. *Déroulement*

Un texte de fiction de Gianni Rodari, long de 1526 mots, est présenté sur un écran d'ordinateur ; ce texte permet de nombreuses interprétations, en partie par l'usage que fait l'auteur de différents épilogues. Le sujet peut parcourir à sa guise les 9 pages écran pendant le temps qu'il estime nécessaire. Ensuite, il répond à 12 questions par un système de QCM, le texte n'étant alors plus consultable. Un barème a été établi par un groupe de juges formés d'enseignants et de bibliothécaires pour décrire des degrés d'interprétation et ne pas s'enfermer dans le tout ou rien. Ainsi, toutes les réponses proposées sont possibles mais certaines témoignent d'un niveau plus approfondi de compréhension.

Le degré de compréhension est donc calculé en s'efforçant de correspondre aux définitions de la lecture experte d'un texte littéraire, se rapprochant ainsi de ce qu'il est permis d'attendre à l'entrée des classes de premier cycle de l'enseignement secondaire. Ce score de compréhension peut s'échelonner de 15 à 94 points.

Les documents sont présentés dans l'annexe 1.1.2.

7.1.2.2. *Variables primaires*

La vitesse de lecture en mots/heure

Moyenne	Médiane	Écart-type	Minimum	Maximum
20393	19607	5701	10258	35346

Tableau 6 — Statistiques descriptives de la vitesse de lecture. Epreuves de lecture n°2.

La vitesse est ici légèrement supérieure à celle des textes courts, ce qui témoigne en faveur de l'adaptation des stratégies à la tâche. Toutefois, les individus ayant des vitesses lentes restent lents quoiqu'il arrive, sans doute par impossibilité technique de faire autrement.

La compréhension

Le score est la somme des questions 1 à 11 à laquelle s'ajoutent les 2/3 de la somme des 5 questions classées à la question 12. Le maximum possible est 94 (voir l'annexe 1.1.2.2 pour une description détaillée du calcul des points).

Moyenne	Médiane	Ecart-type	Minimum	Maximum
76,68	79,50	9,77992	40,00000	91,00000

Tableau 7 — Statistiques descriptives de la compréhension. Epreuves de lecture n°2.

7.2. Commentaires sur les variables primaires

Ces épreuves ont maintenant été passées par des sujets de différents âges à l'occasion de travaux antérieurs (Foucambert, 1997; Foucambert, 2000). Le tableau 8 présente les différents résultats issus de ces travaux :

Références	Effectif	Niveau	Epreuve n°1		Epreuve n°2	
			Compr.	Vitesse	Compr.	Vitesse
Foucambert, 1997	59	CM2		14940	54,22	11835
Foucambert, 2000	85	6 ^{ème}		14574	57,41	14331
Etude présente	64	Adulte	78	18202	76,68	20393

Tableau 8 — Récapitulatif des variables issues des mêmes épreuves selon les âges.

La compréhension de l'épreuve n°1 des épreuves passées en CM2 et en 6^{ème} n'a pas été rapportée car elle avait été calculée d'une manière légèrement différente. Si les élèves de CM2 et de sixième (testés dans les deux cas au mois de juin), ont des vitesses et des compréhensions comparables, les résultats des adultes montrent une avancée significative⁴⁰ que ce soit en vitesse ou en compréhension. Malgré cette élévation bien compréhensible du niveau de lecture chez des adultes lettrés, on note malgré tout une assez forte dispersion interne, puisque les vitesses de lecture varient d'un facteur 3 entre les plus

⁴⁰ Le calcul des T de Student montre une différence significative à $p < 0,001$ entre les adultes et les enfants, que ce soit pour la compréhension ou pour la vitesse de lecture.

rapides et les plus lents, alors que les compréhensions vont du simple au double pour l'épreuve n°2.

Nous précisons déjà lors du travail publié en 2000 l'absence de corrélation entre les vitesses de lecture et les compréhensions. Les données relevées sur notre population ne montrent une corrélation significative ni à l'intérieur de l'épreuve n°1 ($r = 0,06$, $p < 0,65$), ni à l'intérieur de la deuxième épreuve ($r = -0,05$, $p < 0,7$). Ce résultat est cohérent avec un certain nombre de travaux antérieurs (Underwood, Hubbard, & Wilkinson, 1990; Calef, Pieper, & Coffey, 1999).

7.3. Construction des indices du niveau de lecture

Les deux épreuves différentes utilisées pour construire la performance de lecture nous semblent révélatrices de ce qui attend un lecteur dans sa pratique quotidienne :

- parcourir des textes relativement simples pour en extraire de l'explicite
- conduire une lecture savante sur un texte.

Deux questions se posent pour extraire de ces quatre variables primaires des informations pertinentes. La première est de savoir comment rendre complètement indépendantes les vitesses et les compréhensions de chacun des individus. Cette séparation s'impose car nous chercherons, dans la suite de ce travail, à mettre en rapport la vitesse et la compréhension avec l'oubli de lettres. Le deuxième est de s'assurer que les phénomènes illustrés par les valeurs numériques sont suffisamment homogènes. Pour répondre à ces deux impératifs, il nous a semblé qu'une Analyse en Composante Principale était une méthode appropriée. Plutôt que de simplement additionner les résultats obtenus par les sujets, nous avons construit différents indices à partir des axes d'une Analyse en Composantes Principales. Ce type d'analyse factorielle est en effet particulièrement utile quand on veut synthétiser plusieurs variables à partir des liaisons qu'elles entretiennent. Les facteurs construits par l'analyse en composantes principales sont obtenus par des "combinaisons linéaires" des

variables d'origine. L'analyse en composantes principales calcule d'abord la droite pour laquelle la variance des individus est maximale et qui est appelée la première composante principale⁴¹. Ensuite, une deuxième composante principale est recherchée, sous deux conditions :

- avoir une corrélation linéaire nulle avec la première
- avoir, à son tour, la plus grande variance.

Ainsi, l'analyse en composantes principales en éliminant successivement les corrélations multiples entre les résultats aux différentes épreuves permet de décomposer chaque épreuve en fonction des liens multiples que son résultat entretient avec les autres. Cette analyse aboutit à remplacer ici les résultats des 2 épreuves par 4 composantes plus ou moins simultanément à l'œuvre dans chaque résultat.

7.3.1. Résultats de l'ACP

Les quatre variables primaires issues des deux épreuves sont les variables actives qui permettent de construire 4 facteurs.

Axe	Valeur propre	Pourcentage de la variance expliquée	Valeur propre cumulée	Pourcentage cumulé
1	1,790168	44,75420	1,790168	44,7542
2	1,071344	26,78359	2,861512	71,5378
3	0,945825	23,64562	3,807337	95,1834
4	0,192663	4,81659	4,000000	100,0000

Tableau 9 — Résultats de l'ACP : Variance expliquée et valeurs propres de chacun des axes.

Les trois premiers axes de l'analyse rendent compte, à eux seuls, de plus de 95% de la variance totale ; la figure 29 les présente.

⁴¹ Cette explication du fonctionnement de l'analyse en composantes principales doit beaucoup à (Philippeau, 1986)

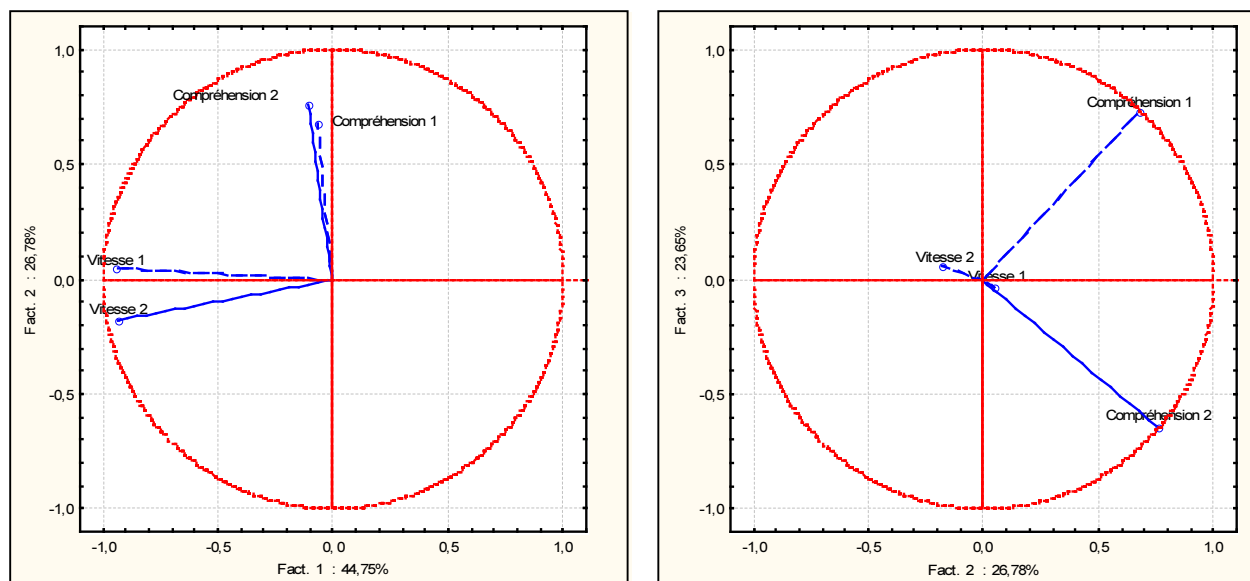


Figure 29 — Facteur 1, 2 et 3 de l'ACP sur les épreuves de lecture (Vitesse 1 et Compréhension 1 pour l'épreuve 1 ; Vitesse 2 et Compréhension 2 pour l'épreuve 2)

On remarque que le premier axe est construit par les deux vitesses de lecture alors que le deuxième l'est par les deux scores de compréhension. Le troisième axe, quant à lui, oppose les compréhensions issues des deux épreuves. En conséquence, l'espace proposé par les trois premiers axes représente bien la compétence générale en lecture, en tenant compte aussi bien des phénomènes qui unissent les vitesses de lecture (axe 1) que de ceux qui rapprochent les processus à l'œuvre dans les deux compréhensions (axe 2). Le troisième axe différencie les processus de compréhension entre ceux assez factuels d'un texte simple et ceux plus profonds d'un texte à plusieurs niveaux de lecture (axe 3) ; entre la compréhension en lecture d'information (documentaire et presse) et en lecture littéraire, entre saisie de l'explicite et traitement de l'implicite. La compréhension est donc évaluée à l'aide de deux composantes de poids assez proches (26 et 23 % de variance exprimée) : l'axe 2 exprime ce qu'elles ont en commun et permet de parler d'un niveau général de compréhension tandis que l'axe 3 exprime ce qui les différencie selon la nature des textes qu'il s'agit de comprendre. Cette dissociation rejoint la complexité décrite par Fayol au sujet des processus de compréhension : « L'immense complexité des processus, où tout se modifie simultanément et en interaction :

les signifiants, les signifiés, les procédures, les capacités de contrôle, etc... » (Fayol, 1992).

Dans leur ensemble, les deux phénomènes vitesse et compréhension ont un poids similaire, puisque l'axe 1 explique 44 % de la variance alors que les deux axes représentant la compréhension en expliquent un peu plus de 50%.

L'axe 1 représentant massivement la vitesse de lecture, la coordonnée des individus sur cet axe sera réutilisée pour définir leur vitesse de lecture dans les analyses ultérieures. De la même façon et pour des raisons similaires, les coordonnées des individus sur l'axe 2 et sur l'axe 3 définiront respectivement leur niveau et leur style de compréhension. Dorénavant, nous appellerons la compréhension construite par l'axe 2 de l'ACP la compréhension ^U (pour Union) et celle construite par l'axe 3 la compréhension ^D (pour Différence).

7.3.2. Description des trois indices retenus

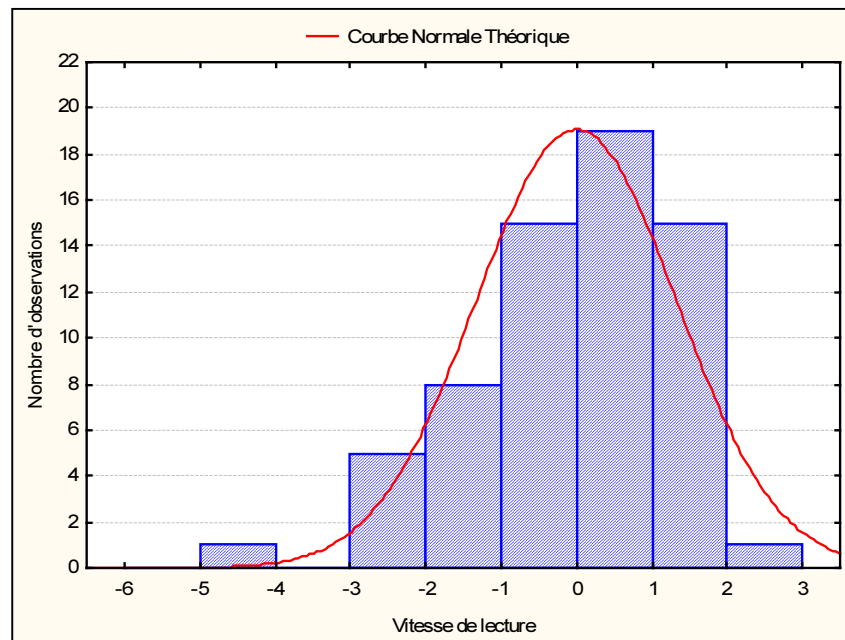


Figure 30 — Distribution de la vitesse de lecture.

Le calcul du test de Kolmogorov-Smirnov montre que la vitesse (axe 1 de l'ACP) possède une distribution normale ($d = 0,09401$, $p > 0,20$)

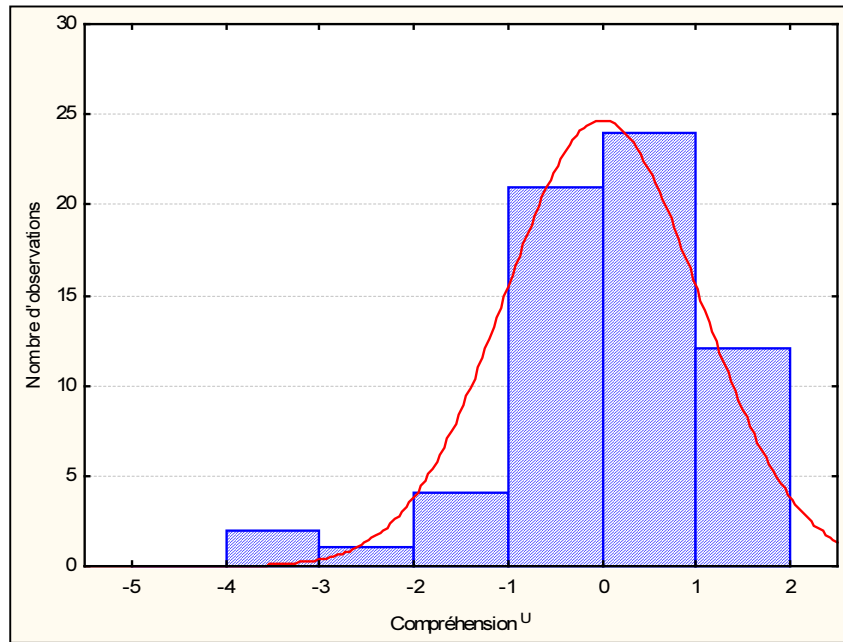


Figure 31 — Distribution de la compréhension^U (Axe 2).

Le calcul du test de Kolmogorov-Smirnov montre que la compréhension^U (axe 2 de l'ACP) possède une distribution normale ($d = 0,08499$, $p > 0,20$)

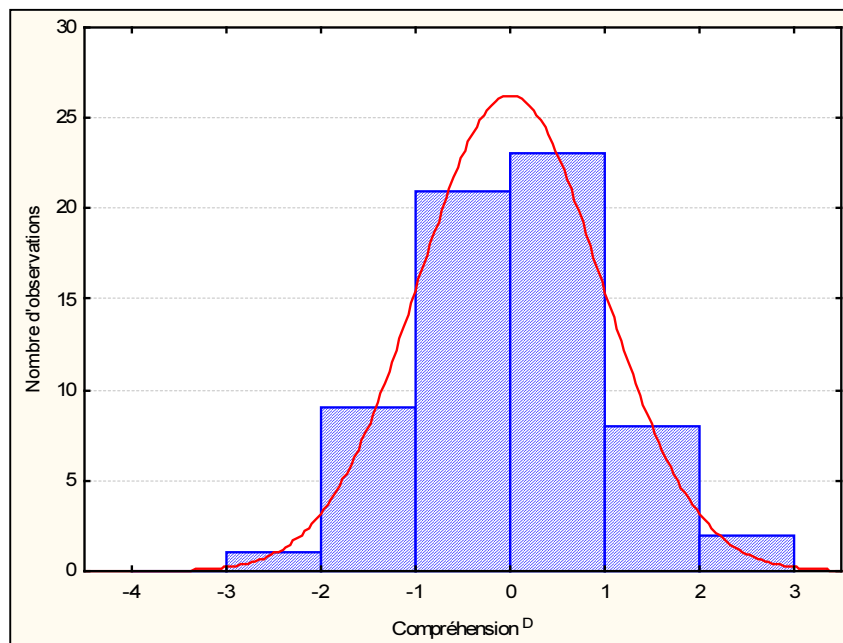


Figure 32 — Distribution de la compréhension^D (Axe 3).

Le calcul du test de Kolmogorov-Smirnov montre que la compréhension^D (axe 3 de l'ACP) possède une distribution normale ($d = 0,04608$, $p > .20$)

Chapitre 8.

L'oubli de lettres : évaluations des fréquences, rôles et natures des mots

Ce chapitre présente deux procédures expérimentales qui ont pour vocation de tester les hypothèses nulles que nous énoncions au cours du chapitre 6. L'hypothèse nulle n°1 prétend que le taux d'oubli de lettre dans les mots dépend plus de la fréquence des mots que de leur rôle syntaxique dans la phrase. L'hypothèse nulle n°2, quant à elle, postule que la nature grammaticale des mots fonctionnels n'a aucun effet sur le taux d'oubli de lettres.

8.1. Expérience 1

Plusieurs expérimentations ont montré l'intérêt d'utiliser des phrases ou des groupes de mots en désordre pour mieux comprendre les oublis de lettres (Drewnowsky, 1978; Drewnowsky & Healy, 1977; Koriat et al., 1991). Nous en reprendrons ici l'idée générale, mais en appliquant le désordre à l'ensemble du texte. Si le modèle de Koriat s'applique ici, on peut faire l'hypothèse que l'effet d'oublis de lettres va s'amoinrir puisque le texte désordonné devient une simple suite de mots dans laquelle toute syntaxe a disparu. Cette première expérience a pour objectif principal de répliquer l'effet d'oubli de lettres en français et de mettre à l'épreuve le modèle de Koriat à l'aide de cette langue.

8.1.1. Population

Les 64 individus adultes décrits précédemment (cf chapitre 7) ont participé de manière volontaire à cette série d'expérimentations. Leur langue maternelle est le français et leur vue est corrigée en cas de besoin.

8.1.2. Matériel

Deux supports vont être présentés aux sujets à quatre jours d'intervalle. Le premier jour, la lettre à barrer est le *u* sur un texte de 421 mots. Le deuxième « texte » proposé quatre jours plus tard est constitué des mots du premier texte, mais qu'un programme informatique a mélangé en utilisant un algorithme de

génération de nombres aléatoires. Les marques fortes de ponctuation ont été maintenues et un retour sur le texte a repositionné les majuscules après elles.

Les textes ont été présentés de manière identique, dans un double interlignage, dans la fonte Times New Roman et dans une taille de 12 points. Au total, 218 mots contenant la lettre *u* (109 mots dans chaque texte) vont être observés⁴².

8.1.3. Procédure

Avant la première épreuve, les participants ont eu un temps d'apprentissage de la tâche, pendant lequel ils ont barré, sur le texte de la consigne, une lettre qui n'a plus été recherchée par la suite. La consigne de lire le texte pour le comprendre leur était donnée, de le lire à leur vitesse et, comme toujours dans cette tâche, de ne pas faire de retours en arrière s'ils avaient la sensation d'avoir oublié une lettre. Quatre jours plus tard, les consignes furent redonnées oralement juste avant l'épreuve ; aucune explication de ce qui était recherché n'a été apportée. La lettre –u– à barrer est rappelée au haut de chaque texte. L'ordre de présentation a été choisi pour alourdir encore la charge de la preuve dans le cas où le souvenir du texte ordonné se projetterait malgré tout sur ses éléments en désordre.

8.1.4. Résultats

La figure 33 montre les pourcentages moyens d'omission des lettres dans les mots en fonction à la fois des deux présentations (ordonné vs mélangé) et des critères de rôle du mot (sémantique vs fonctionnel). Cette variable a été construite à partir de la nature (donc aussi du rôle) des mots dans le texte ordonné : les noms, verbes, adjectifs et adverbes sont considérés comme ayant un rôle sémantique, c'est-à-dire qu'ils transportent de l'information sur un personnage ou une action du texte. Les autres mots ont été considérés comme ayant un rôle syntaxique, dans la mesure où ils organisent les mots à rôle sémantique entre eux et ne portent pas d'informations sémantiques propres : ce sont les conjonctions, les déterminants, les pronoms relatifs, les prépositions,...

⁴² Voir l'annexe 1.2.1

Une analyse de covariance a été conduite sur ces résultats en introduisant comme variables indépendantes les deux variables « Rôle du mot » et « Type de présentation » (Ordonné vs Mélangé).

Les résultats univariés montrent :

- *Un effet du type de présentation* ($F(1, 239) = 30,43, p < 0,001$) avec une moyenne d'omission de 37% pour le texte ordonné contre 21% pour le texte mélangé.
- *Un effet de la nature du mot* ($F(1, 239) = 36,34, p < 0,001$) avec une moyenne d'omission de 37% pour les mots à rôle syntaxique contre 20% pour les mots à rôle sémantique.

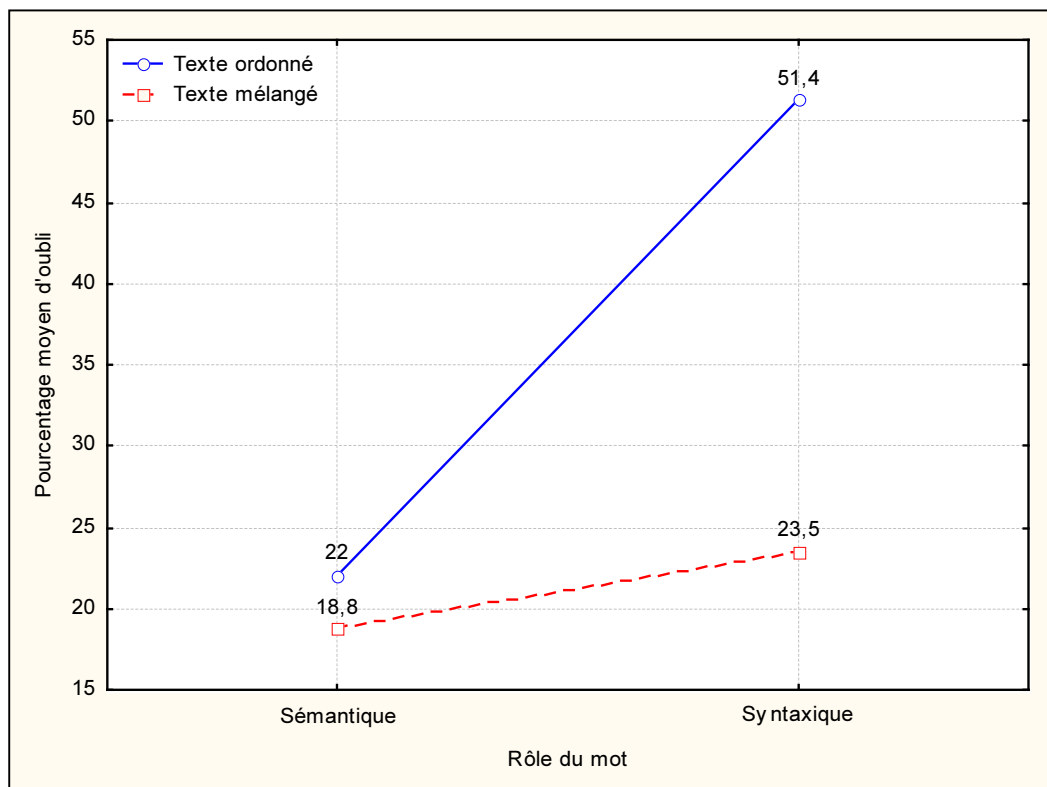


Figure 33 — Pourcentages d'oubli de lettres dans les mots en fonction du rôle du mot et du type de texte. Expérience 1.

L'étude des interactions apporte des résultats intéressants. Une interaction significative a lieu entre le type de présentation du texte et le rôle du mot ($F(1,239) = 19,06 ; p < 0,001$). La figure 33 illustre donc plusieurs phénomènes :

- Le pourcentage d'omission est sensiblement le même pour les mots ayant un rôle sémantique que le texte soit ordonné ou non.
- Dans le texte mélangé, il n'y a pas de différence significative entre les omissions sur les mots ayant un rôle syntaxique et les mots ayant un rôle sémantique.
- La seule différence significative est dans l'oubli des lettres des mots fonctionnels, organisant la syntaxe de la phrase, puisque leur pourcentage d'omissions est supérieur de près de 100 % dans le texte ordonné. Les tests Post Hoc (HSD de Tukey pour des échantillons de tailles différentes) montre une différence de moyenne très significative quand on est dans le texte ordonné (Sémantique vs Fonctionnel $p < 0,0001$).

8.1.5. Discussion

Ces résultats confirment l'importance de la syntaxe dans l'oubli de lettres. La désorganisation du texte, en faisant perdre la structure des propositions et l'ordre naturel des mots à l'intérieur des différents syntagmes, ne permet plus au lecteur d'établir des cadres syntaxiques comme il peut le faire dans un texte normalement construit. Néanmoins, il peut arriver que le placement aléatoire des mots ait, par hasard, recréé quelques groupes fonctionnels, ce qui explique, que, même dans le texte mélangé, on observe une légère différence (4,5 %) d'oublis entre les mots à rôle syntaxique et à rôle sémantique, absolument non significative.

Mais ce qui est essentiel dans cette expérience, c'est que dans le texte en désordre les mots fonctionnels existent tout autant que dans le texte ordonné. Mais ils n'ont plus de rôle syntaxique. Et dans ce cas, ils ne sont pas traités de manière significativement différente de tous les autres mots. Ce n'est donc pas quelque chose qui leur appartiendrait dans leur apparence, leur nature ou leur familiarité (leur fréquence) qui crée un différentiel d'oubli de lettres, mais bien, et seulement, la possibilité de leur reconnaître un rôle.

8.2. Expérience 2

Toutes les expérimentations que nous avons évoquées sur l'effet d'oubli de lettres (cf. chapitre 1) ont été menées sur des corpus relativement réduits de mots puisque, d'une manière générale, les auteurs comparent deux items l'un à l'autre en choisissant les couples de manière à différencier au mieux les effets de fréquence ou de structure. Notre démarche est différente : nous voulons étudier les contraintes qui pèsent sur cet oubli de lettres dans les conditions les plus proches d'une lecture "ordinaire" d'un texte.

Avec cette deuxième expérimentation, nous allons tester les première et deuxième hypothèses posées au chapitre précédent. Rappelons que la première hypothèse propose d'examiner le poids respectif des deux processus concurrents dans la littérature pour expliquer l'effet d'oubli de lettres, à savoir le rôle syntaxique et la familiarité avec le mot, illustrée par sa fréquence.

Pour ce qui est de la deuxième hypothèse, elle repose sur l'idée que différents types de cadres syntaxiques doivent être construits par le lecteur, dans la mesure où la construction syntaxique d'une phrase possède une arborescence complexe, avec différents niveaux de branchements. Ce que nous recherchons, ce sont les éléments en relation avec des structures de taille importante (de l'échelle de la phrase) et ceux relatifs à des syntagmes fonctionnels de taille plus réduite. Cette construction plurielle de structures diverses devrait s'observer chez le lecteur par des taux d'oubli de lettres différents suivant le type de structures introduit par le mot fonctionnel observé. En conséquence :

- Si les structures anticipées par le lecteur sont uniquement de taille réduite, à l'échelle par exemple d'un syntagme nominal⁴³, on ne devrait pas trouver d'oubli massif sur les mots organisant la liaison entre deux syntagmes flexionnels, à savoir les complémenteurs (conjonctions, pronoms relatifs).

⁴³ Dont on a vu que la théorie syntaxique considère que la tête n'est plus le nom mais le déterminant. Cependant, pour faciliter la lecture, on continue à le nommer syntagme nominal et non syntagme déterminant.

- Au contraire, si la construction se déroule à la fois au niveau des constituants et des syntagmes flexionnels, on devrait obtenir des taux d'oublis plus élevés pour les mots fonctionnels commandant ces deux catégories.

Nous prendrons également en compte dans l'analyse d'autres phénomènes dont nous avons vu qu'ils pouvaient influencer l'oubli de lettres comme, par exemple, la position de la lettre dans le mot, du mot dans la page, etc...

8.2.1. Population

Les mêmes 64 individus que dans la première expérience ont participé à cette série d'expérimentations.

8.2.2. Matériel et procédure

Quatre textes supports sont présentés sur quatre jours aux sujets (un par jour). Le premier jour, la lettre à barrer est le *u* sur un texte de 421 mots (correspondant au premier texte de l'expérience 1). Le deuxième texte comprend 445 mots, et la lettre à barrer est le *d* ; le troisième texte, long de 592 mots a comme objet de recherche la lettre *r*. Le dernier jour, il s'agit de la lettre *t* sur un texte de 586 mots⁴⁴. Les textes sont présentés de manière identique, dans un double interlignage, dans la fonte Times New Roman avec une taille de 12 points. Sur les 2044 mots de l'ensemble des textes, on en compte 617 contenant les différentes lettres à barrer⁴⁵.

La présentation des textes est identique à celle de la première expérience.

8.2.3. Résultats

Pour chacun des textes, le même décryptage des données a eu lieu. Les variables suivantes ont été construites :

⁴⁴ Voir l'annexe 1.2.1

⁴⁵ Parmi ces 617 mots, 23 ne possèdent pas de fréquence dans la base Frantext (principalement des noms propres comme Europe, Amérique,...). Nous leur donnerons la fréquence moyenne des mots de la même catégorie grammaticale.

a. *la position de la lettre sur la ligne*. La ligne étant découpée en quarts (gauche, gauche-milieu, milieu-droite, droite), on obtient une variable à quatre modalités.

b. *la position de la lettre dans la hauteur de la page*, en quatre classes, du haut vers le bas de la page.

Les variables *a* et *b* définissent une cible sur la page, afin de rechercher d'éventuels phénomènes de localisation « géographique ». En particulier, la variable *a* tente de vérifier l'hypothèse selon laquelle les mots situés plutôt à droite de la ligne vont subir un taux d'oubli de la lettre plus important, étant préalablement perçus par la vision parafovéale à l'inverse des mots situés dans le quart gauche de la ligne.

c. *La position de la lettre dans le mot*. Comme nous ne suivons pas le déplacement oculaire pendant l'épreuve, il est nécessaire de chercher des indices relatifs aux contraintes visuelles du mot où la lettre figure. En nous inspirant des travaux sur la position optimale de l'œil pendant la lecture (O'Regan, 1992), nous avons découpé chacun des mots en tiers pour y noter la position de la lettre à barrer. La position optimale se situant approximativement vers la fin du premier tiers, on peut s'attendre à ce que les lettres figurant autour de cet emplacement soient moins oubliées que les suivantes, si tant est que les lecteurs fixent successivement chacun des mots.

d. *la taille du mot* où se trouve la lettre à barrer. Ce sera une variable continue, exprimant, en nombre de lettres, la longueur du mot.

e. *la position du mot dans la phrase*. Cette variable sera aussi donnée en quarts. La question sous-jacente consiste à décider si les mots situés en fin de phrases comportent plus d'oublis de lettres que ceux situés en début.

f. *la longueur de la phrase*, en quatre modalités. Ici, on cherche à savoir si le fait que le mot se trouve dans des phrases de longueur différente influe sur l'oubli de lettres.

- g. *la fréquence du mot*. Elle est calculée en fonction des valeurs extraites pour chacun des mots de la base FRANTEXT (New, Pallier, Ferrand, & Matos, 2001).
- h. *la lettre à barrer*. Dans un grand nombre de travaux cités au chapitre 1, on trouve des différences de taux d'oublis de lettres en fonction du choix de la lettre.
- i. *le rôle du mot dans la phrase*. Variable à deux modalités (rôle syntaxique vs. rôle sémantique).

Dans les analyses statistiques qui vont suivre, on prendra également en compte les interactions entre les variables (*c* et *d*) écrivant le mot (taille du mot x position de la lettre dans le mot), celles entre les variables (*a* et *b*) décrivant la position du mot dans le texte (haut/bas * gauche/droite), et celles entre les variables (*e* et *f*) décrivant le mot dans la phrase (longueur de la phrase x position du mot dans la phrase).

Première analyse

La première analyse à laquelle nous procéderons a pour objectif d'élucider la dichotomie fréquence/rôle des mots dans l'oubli de lettres. Pour mieux évaluer les poids respectifs des différents facteurs introduits dans le modèle explicatif de l'oubli de lettres, une analyse en régression multiple a été exécutée, en introduisant dans les modèles respectifs l'ensemble des variables décrites précédemment.

L'objectif étant ici d'affiner le modèle, nous éliminerons du premier modèle les mots à fort résidu (dont la valeur-absolue du résidu calculée dans le modèle M1 est supérieure à 1,5 fois l'écart-type de tous les résidus). Le tableau 10 présente les résultats généraux du modèle calculés en tenant compte de toutes les observations (M1) et ceux calculés avec un effectif amputé de moins de 15% des observations (M2).

Modèle	N	Corrélation	Pourcentage de variance expliquée	Significativité
M1 : avec tous les mots	617	0,63	0,39	<0,0001
M2 : après élimination des forts résidus	533	0,72	0,52	<0,0001

Tableau 10 — Résultats généraux de la régression cherchant à expliquer le pourcentage d'oubli de lettres. Modèles M1 et M2. Expérience 2.

La part de variance de l'oubli de lettres expliquée par les variables indépendantes (importante puisque plus de 50% du phénomène ($R^2 = 0,52$) sont expliqués dans le modèle M2) et le niveau de la corrélation multiple ($R = 0,72$), justifient l'examen approfondi des contributions de certaines des variables à l'effet d'oublis de lettres.

Effet	Niveau	Paramètre	Err. Type	t	p	bêta
Rôle	Fonctionnel	9,12	0,914	9,98	<0,0001	0,446
Lettre	t	7,7	0,855	9	<0,0001	0,390
Fréquence		0,0008	0,0001	6,19	<0,0001	0,295
Lettre	r	-3,01	0,857	-3,51	<0,0001	-0,145

Tableau 11 — Analyse de régression multiple : modèle M2. Paramètres des variables ou modalités présentant un bêta significatif à 0,001. Expérience 2.

Le tableau 11 présente les coefficients de régression, les coefficients standardisés, et la significativité de ces coefficients, pour les variables ou modalités présentant un coefficient bêta significatif à 0,0001⁴⁶. On remarque toute l'importance de la variable *Rôle du mot*, puisque les coefficients standardisés des deux modalités sont les plus fortes (en valeur absolue). La fréquence ne vient qu'en troisième position, avec un bêta de 0,3.

Deuxième analyse

Le propos de cette deuxième analyse est de tester notre deuxième hypothèse qui concerne l'éventuelle hétérogénéité des oublis de lettres à l'intérieur des

⁴⁶ Le tableau complet est en annexe 3.2.1

mots ayant un rôle syntaxique. Pour l'étudier, aux huit variables (a \Rightarrow h) et aux trois interactions décrites plus haut, nous substituons à la variable *i* (rôle du mot) la variable *j* qui en est un développement :

j. *la nature du mot*. Les natures possibles des mots où figure une lettre à barrer sont les adjectifs, adverbes, articles, noms, prépositions, verbes, conjonctions de coordination, conjonctions de subordination et enfin les pronoms relatifs. Pour deux raisons, nous regroupons les trois dernières natures (conjonctions de coordination, conjonctions de subordination et pronoms relatifs) dans une seule rubrique que nous nommerons *complémenteur* :

- Ces mots représentent les complémenteurs dont nous avons parlé au chapitre 2. Ils permettent une liaison entre deux syntagmes flexionnels et, à ce titre, l'oubli de lettres serait révélateur d'une construction d'un cadre au niveau de la phrase.
- Plus secondairement, en raison de leur faible effectif.

Une analyse de covariance a été effectuée en introduisant dans le modèle l'ensemble des 9 variables décrites et des trois interactions. Le tableau 12 présente les résultats univariés de chacune des variables.

	Degré de liberté	F	p<
Position de la lettre sur la ligne (var. a)	3	1,29	0,28
Position de la lettre dans la hauteur de la page (var. b)	3	0,56	0,64
Position de la lettre dans le mot (var. c)	2	0,41	0,66
Taille du mot (var. d)	1	5,38	0,02
Position du mot dans la phrase (var. e)	3	1,33	0,26
Longueur de la phrase (var f.)	3	1,13	0,34
Fréquence (var. g)	1	33,44	0,001
Lettre (var. h)	3	17,75	0,0001
Nature (var j.)	6	13,8	0,0001
a * b	9	1,94	0,05
c * d	2	2,47	0,09
e * f	9	0,78	0,64
Ord.Orig.	1	158,56	0,0001

Tableau 12 — Analyse de covariance. Résultats univariés de chaque variable explicative. Les contributions significatives sont en gras.

- On observe que la lettre à barrer agit de manière importante ($F(3, 548) = 17,7532, p < 0,0001$) avec davantage d'oublis sur les deux lettres hautes. Nous n'avons pas d'explication de ce phénomène. Cependant des données similaires existent dans la littérature, comme, par exemple, la différence saisissante entre le pourcentage d'omission de la lettre *t* et de la lettre *u* dans le mot *outré*⁴⁷ (Saint-Aubin & Poirier, 1997), ou encore le plus fort taux d'oublis de la lettre *b* comparativement aux lettres *l*, *s* ou *m* dans la langue hébraïque (Koriat & Greenberg, 1996a).
- La nature du mot est très fortement significative dans l'effet d'oublis de lettres ($F(6, 571) = 13,79, p < 0,0001$).
- Du côté des deux variables continues, notons, toutes choses égales par ailleurs, un fort effet de la fréquence du mot sur l'oubli de lettres ($F(1, 571) = 33,44, p < 0,0001$) avec un plus fort taux d'oublis pour les mots les plus fréquents.
- Un effet significatif bien que moindre pour la taille des mots ($F(1, 571) = 5,37, p < 0,021$) où plus le mot est long, plus la lettre a de chance d'être oubliée.
- Une seule interaction est significative : celle entre la position de la lettre sur la ligne et la position de la lettre dans la hauteur de la page *haut/bas* \times *gauche/droite* ($F(9, 571) = 1,9443, p < 0,05$), avec une tendance à oublier plus souvent une lettre placée dans la partie basse/droite de la feuille que dans sa partie haute/droite.

Le tableau 13 donne les résultats pour chaque modalité des variables catégorielles (on trouvera dans l'annexe 3.2.2. les pourcentages d'oubli des différentes modalités des interactions, ainsi que les moyennes des covariants).

⁴⁷ avec un pourcentage d'oublis bien plus important pour la lettre *t* que pour la lettre *u*.

Variables indépendantes	Modalités des variables indépendantes	Effectif	Moyenne d'oublis	Erreur-type
Nature F(6, 571) = 13,79, p<0,0001 S.	Nom	179	20,77	1,39
	Verbe	187	22,23	1,33
	Adverbe	74	24,64	1,98
	Adjectif	72	24,61	1,98
	Préposition	55	35,16	2,49
	Déterminant	25	49,27	3,50
	Complémenteur	25	46,86	3,79
Position du mot dans la phrase F(3, 571) = 1,3319, p<0,27 N.S.	Premier quart (début)	87	33,15	2,02
	Deuxième quart	174	31,30	1,49
	Troisième quart	141	33,17	1,61
	Quatrième quart (fin)	215	30,12	1,43
Position de la lettre sur la ligne F(3, 571) = 1,2871, p<0,28 N.S.	Premier quart (gauche)	149	31,34	1,65
	Deuxième quart	174	34,16	1,57
	Troisième quart	173	30,49	1,63
	Quatrième quart (droite)	121	31,76	2,23
Position de la lettre dans la hauteur de la page F(3, 571) = 0,5624, p<0,64 N.S.	Premier quart (haut)	188	30,74	1,44
	Deuxième quart	189	31,56	1,43
	Troisième quart	196	32,73	1,43
	Quatrième quart (bas)	44	32,71	2,67
Pos. de la lettre dans le mot F(2, 571) = 0,4147, p<0,67 N.S.	Premier tiers (gauche)	129	30,36	1,64
	Deuxième tiers	191	30,93	1,54
	Troisième tiers (droite)	297	34,52	1,30
Type de lettre F(3, 548) = 17,7532, p<0,001 S.	d	80	28,86	2,20
	r	180	28,29	1,63
	t	236	39,10	1,54
	u	121	31,49	1,98
Longueur de la phrase F(3, 571) = 1,1252, p<0,34 N.S.	Courte	205	30,43	1,58
	Courte-Moyenne	152	32,11	1,70
	Moyenne-Longue	113	30,88	1,92
	Longue	147	34,32	1,78

Tableau 13 — Moyennes des moindres carrés des pourcentages d'oublis de lettres dans les mots (calculées pour les covariants à leurs moyennes). Expérience 2.

La figure 34 montre les pourcentages d'oubli selon les natures : on y voit clairement un continuum avec, à gauche, les mots à rôle sémantique et, à droite, les mots organisant la syntaxe de la phrase. Ces résultats confirment l'opposition entre rôle sémantique et rôle syntaxique. Ils révèlent le pouvoir différent d'organisation de la syntaxe, selon la nature même des mots. Les complémenteurs introduisant des relations entre différents éléments phrastiques (donc très importants dans l'organisation de la syntaxe) et les déterminants (précédant sans ambiguïté un élément à rôle sémantique)

présentent des taux d'oubli élevés. Cependant, les prépositions, bien qu'appartenant sans conteste à l'ensemble des mots fonctionnels, présentent un taux d'oubli moins fort.

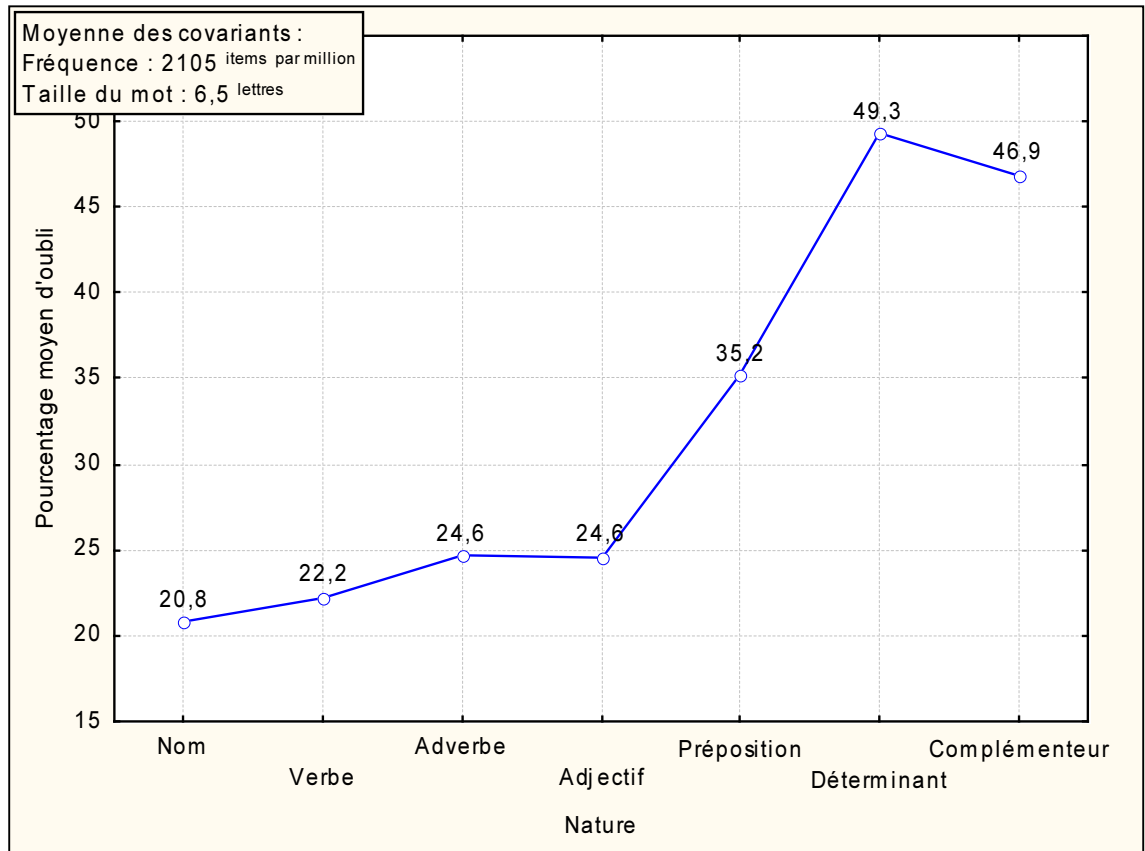


Figure 34 — Pourcentages d'oublis de lettres dans les mots en fonction de leur nature grammaticale. Expérience 2.

L'étude des groupes homogènes à partir des tests Post Hoc présentés dans le tableau 14 montrent trois communautés de natures de mots : la première regroupe toutes les natures dont les occurrences effectives sont chargées de significations, la deuxième ne comprend que les prépositions introduisant un syntagme prépositionnel et la troisième comprend les déterminants introduisant un syntagme nominal et les compléments faisant le lien entre deux syntagmes flexionnels.

	Groupe 1	Groupe 2	Groupe 3
Nature			
Nom	****		
Verbe	****		
Adjectif	****		
Adverbe	****		
Préposition		****	
Déterminant			****
Complémenteur			****

Tableau 14 — Groupes homogènes dans la nature des mots. Tests Post Hoc sur la nature des mots. (HSD de Tuckey pour des effectifs différents ; seuil de significativité <0,05)

8.2.4. Discussion

Le dispositif présenté ici avait pour objectif de préciser l'effet du rôle des mots dans le phénomène d'oublis de lettres, tout en contrôlant d'autres critères susceptibles d'interférer avec cet effet. Nous avons pris comme parti d'observer l'oubli d'une lettre dans tous les mots la contenant, en caractérisant chaque occurrence aussi bien par sa localisation dans la page et dans le mot que par la fréquence en français écrit du mot qui la contient et par son rôle dans la phrase. Cette méthode permet de s'affranchir du petit nombre d'items étudiés et fait porter l'étude sur plus de 39000 mots⁴⁸.

La tentative de différencier l'oubli de lettres en fonction de caractéristiques relatives soit à la géographie de la lettre, soit à la taille du mot et à la position de la lettre dans ce mot n'a pas donné de résultats nouveaux. Il ne semble pas que ces facteurs entrent en jeu dans l'oubli de lettres. Cependant, le fait que les lettres soient plus oubliées en bas à gauche de la feuille qu'en haut à droite n'est pas sans nous étonner. D'autres observations seront nécessaires pour répliquer ce phénomène et mieux comprendre son fonctionnement.

La question seulement posée par Koriat de savoir si certaines natures de mots représentent une plus grande capacité à organiser la syntaxe de la phrase trouve ici des éléments de réponse. Le dispositif différencie en effet les mots en fonction de leur nature au sein même des catégories sémantiques et fonctionnelles. Les deux natures grammaticales qui provoquent le moins

⁴⁸ 617 mots observés dans l'ensemble des textes pour 64 individus = 39488 observations

d'omissions (les noms 20% et les verbes 22%) sont définies dans la littérature linguistique (Martinet, 1979; Bronckart, Bain, Schneuwly, Davaud, & Pasquier, 1985) comme « pivots », c'est à dire qu'elles fonctionnent comme noyaux soit d'un groupe nominal, soit d'un groupe verbal et sont donc indispensables à l'existence même de ces groupes.

Pour ce qui est des mots organisant la syntaxe, deux catégories semblent particulièrement provoquer l'oubli de lettres : d'une part, celle des mots *complémenteurs* qui opèrent soit aux frontières de propositions de niveaux équivalents et/ou de phrases (coordonnants), soit à l'articulation d'un type de proposition dans une autre (subordonnants), d'autre part, les déterminants, tête des syntagmes nominaux⁴⁹, qui introduisent un groupe de taille plus restreinte, facilement identifiable, et qui est toujours suivi par des informations sémantiques et souvent par la reprise de marques portant sur le genre et le nombre. En outre, ces données précisent l'effet de moindre oubli observé sur les mots à rôle sémantique (Koriat & Greenberg, 1996b), puisqu'on remarque un taux d'oubli inférieur sur les pivots des groupes nominaux et verbaux, dont la présence est indispensable à l'établissement du sens.

Il convient de revenir un instant sur l'opposition entre les modèles de Koriat et de Healy. Les données que nous apportons ne semblent contredire aucun des deux. En effet, on observe aussi bien une contribution importante de la fréquence des mots que de leur rôle dans l'oubli des lettres qui les constituent. L'acte lexicale est suffisamment complexe pour qu'il soit possible de concilier un phénomène de construction syntaxique dans le processus de lecture avec des effets de fréquence intervenant dans la reconnaissance de mots, ce qui, comme le suppose Healy, influencerait sur la détection de leurs constituants. Toutefois, les résultats de la régression multiple que nous avons opérée nous permettent de préciser davantage le poids relatif de ces différents facteurs. Le tableau 11 nous rappelle que c'est le rôle du mot, illustré par la dichotomie syntaxique/sémantique, qui possède le poids explicatif le plus important (bêta de 0,44) alors que la fréquence ne possède qu'un bêta de 0,29. En conséquence, même si la fréquence du mot continue de jouer un rôle important

⁴⁹ dans les développements récents de la linguistique (cf. chapitre 2)

dans l'oubli de lettres, le fait pour ce mot de tenir un rôle sémantique ou un rôle syntaxique l'est bien davantage.

On observe un effet très fort de la lettre en elle-même. Si jusque-là plusieurs travaux avaient noté des fréquences d'oubli différentes en fonction de la lettre à barrer (Saint-Aubin & Poirier, 1997), le poids de ce facteur semble avoir été particulièrement sous-estimé. Nous montrons ici que la lettre peut avoir un effet aussi fort que celui de la fréquence du mot. Cette observation est à rapprocher des conclusions de Nazir dans ses travaux sur la visibilité des lettres, où il postule que c'est la forme générale de la lettre, unique information perceptible par la vision parafovéale, qui permet de l'appréhender plus ou moins (Nazir, Jacobs, & O'Regan, 1998). La lettre *t*, beaucoup plus oubliée toutes choses égales par ailleurs que la lettre *r*, est aussi beaucoup mieux perçue par la vision périphérique du fait de l'interaction de deux éléments distinctifs (hauteur et barre).

8.3. Conclusion

Les différents résultats obtenus dans ce chapitre permettent de rejeter les première et deuxième hypothèses nulles.

L'hypothèse n°1 peut être rejetée : des deux phénomènes (fréquence vs. rôle), c'est le rôle du mot dans la phrase qui est le plus important pour expliquer l'oubli de lettres. Les deux expériences confirment, chacune à leur manière, ce résultat. La première parce qu'elle montre que c'est le rôle qui importe : quand on supprime la possibilité (par le mélange des syntagmes) que le mot puisse jouer un rôle syntaxique, l'effet d'oubli sur ces mots disparaît, alors que, bien évidemment, sa fréquence en français reste la même. La seconde, à travers son premier résultat, parce qu'elle permet de calculer le poids respectif des deux phénomènes : celui du rôle est sans conteste plus important que celui de la fréquence du mot. Comme nous le disions auparavant, ces observations ne suggèrent en rien que la fréquence ne soit qu'un artefact dans l'identification du mot. Au contraire, nous avons vu son importance. Ce que nous pouvons

faire, c'est, en hiérarchisant le pouvoir explicatif des variables, montrer que le modèle le plus puissant est celui de lecture structurale, initié par Koriat.

L'hypothèse n°2 peut, elle aussi, être rejetée : le taux d'oubli de lettres dans les mots à rôle syntaxique n'est pas homogène. La nature grammaticale de ces mots influe sur le taux d'oubli. Nos résultats trouvent également une explication dans le modèle de lecture structurale. Comme le proposent les travaux de Koriat, le lecteur extrait le squelette de la phrase qu'il parcourt pour ensuite le compléter par du sémantique. Ce faisant, les mots organisant le contexte sont projetés à l'arrière plan et seuls les mots à rôle sémantique focalisent alors l'attention, au moment où le lecteur les voit nettement. Les données que nous apportons ici laissent apparaître une double organisation : d'une part le lecteur élabore ou tente d'élaborer des cadres assez larges fonctionnant au niveau des propositions et/ou des phrases et, d'autre part, il construit des cadres fonctionnant autour du groupe de mots, le plus souvent relativement réduit quant à leur taille. On peut supposer que le lecteur active simultanément deux schémas syntaxiques, l'un fonctionnant sur des unités longues, organisées principalement par les conjonctions et pronoms relatifs, l'autre sur des unités plus locales, organisées autour du groupe de mots.

Chapitre 9.

Mise en rapport de l'effet d'oubli de lettres et des performances en lecture

9.1. Introduction

Les expériences relatées dans la littérature sur l'effet d'oubli de lettres n'ont jamais recherché d'éventuelles variations du niveau de performances des lecteurs. Seul un article, publié par Greenberg, Koriat et Vellutino, a étudié, dans une perspective développementale, le rapport entre l'effet d'oubli de lettres et l'âge du sujet (Greenberg et al., 1998) : plus les lecteurs progressent en âge (entre 7 et 13 ans), plus l'effet d'oublis de lettres est marqué, ce qui conduit les auteurs à conclure à une prise de conscience progressive du rôle de certains mots dans l'organisation de la structure des phrases. Mais, les populations adultes étudiées jusqu'ici sont considérées comme étant de niveau homogène, et bien peu de travaux ont pris en compte d'autres caractéristiques individuelles, qui mettraient à jour des comportements d'oubli différents en fonction d'habiletés de lecture dissemblables.

Le modèle de Koriat suggère (et nous l'avons confirmé et approfondi dans le chapitre précédent) qu'une forte différence d'oubli de lettres entre les mots fonctionnels et les mots porteurs de sens témoigne d'un recours à ces mots fonctionnels pour construire les cadres syntaxiques. Nous pouvons faire l'hypothèse que la capacité à construire des cadres syntaxiques (observable par un fort différentiel d'oublis) permet une lecture plus efficace en concentrant l'attention sur les éléments sémantiques qu'ils accueillent.

Dans ce chapitre, nous mettrons le différentiel d'oubli entre les mots à rôle syntaxique et les mots à rôle sémantique en relation avec les résultats en lecture tels que nous les avons décrits au chapitre 7. Nous avons choisi d'exprimer la différence entre l'oubli selon les rôles des mots par un rapport plutôt que par une différence : le choix de la différence aurait réintroduit dans

la variable la quantité générale d'oublis⁵⁰, variable que nous garderons par ailleurs.

9.2. Rapports entre performances de lecture et barrage de lettre

9.2.1. L'explication de l'oubli de lettres par les performances du lecteur

Nous observerons la contribution du niveau de lecture à l'effet d'oublis de lettres grâce à une analyse de régression multiple prenant comme variable dépendante le rapport entre les omissions dans les mots à rôle syntaxique et celles dans les mots à rôle sémantique ($m = 3,5$; $\sigma = 2,09$; minimum = 0,75 ; maximum = 10,34 ; Kolmogorov-Smirnov : $d = 0,12$; $p > 0,20$)⁵¹ et comme variables explicatives le sexe, la vitesse de lecture (axe 1 de l'ACP), la compréhension^U (axe 2 de l'ACP) et compréhension^D (axe 3 de l'ACP), ainsi que le pourcentage général d'oublis afin de pouvoir observer le rapport entre les natures d'oublis indépendamment de leur nombre.

Le rapport entre les mots à rôle syntaxique et les mots à rôle sémantique, comme le pourcentage moyen d'oublis, est calculé à partir de l'expérience décrite au chapitre 8, sur les textes 1, 2, 3 et 4, soit sur une observation d'environ 40000⁵² mots. Le regroupement des déterminants, prépositions et complémenteurs (cf. chapitre 8) constituent les mots à rôle syntaxique. Les mots à rôle sémantique sont les noms, les adjectifs, les adverbes et les verbes.

Ce modèle sera utilisé deux fois, une première fois pour repérer les contributions significatives (M3), une seconde fois, après élimination des 6

⁵⁰ En cas de calcul d'une différence, un individu ayant 60% d'oubli pour les mots fonctionnels et 40% pour les mots lexicaux n'aurait pas la même valeur qu'un autre individu ayant eu des oublis de 30% et 20% (20 pour le premier et 10 pour le second). Ce qui nous intéresse, c'est bien que les deux couples sont dans la même rapport (1,5).

⁵¹ Courbe présentée dans l'annexe 3.3.1.

⁵² Nous avons au total 617 mots observés pour 64 individus, soit un total de 39488 mots.

plus forts résidus⁵³, afin de mieux établir les paramètres du modèle M4. Le modèle M4 apporte un gain explicatif de 10 points. Le tableau 15 présente les résultats généraux des deux modèles.

Modèle	N	Corrélation	Pourcentage de variance expliquée	Significativité
M3 : avec tous les individus	64	0,71	50 %	<0,0001
M4 : après élimination des 6 plus forts résidus	58	0,78	60 %	<0,0001

Tableau 15 — Résultats généraux de la régression cherchant à expliquer l’oubli de lettres par les performances des lecteurs. Modèle M3 et M4. Expérience 3.

60% de la variance est expliquée ($R^2 = 0,60$) dans le modèle M4. Le niveau de la corrélation multiple ($R = 0,78$) et la significativité du modèle rendent légitime l’examen des contributions de chacune des variables.

Effet	Paramètre	Erreur Type	t	p	bêta
Pourcentage général d’oublis	-0,10	0,01	-8,52	<0,001	-0,77
Compréhension^U (axe 2 ACP)	0,42	0,14	3,02	<0,004	0,27
Vitesse (Axe 1 ACP)	0,14	0,12	1,15	0,25	0,11
Sexe					
Effet Homme	-0,19	0,18	-1,10	0,27	-0,10
Compréhension ^D (axe 3 ACP)	-0,11	0,14	-0,74	0,46	-0,07

Tableau 16 — Paramètres des variables pour le modèle M4. Expérience 3.

Le tableau 16 présente les coefficients de régression, les coefficients standardisés, et leur significativité, pour les variables continues ou les modalités. On observe que :

- Le sexe se situe en deçà du seuil de significativité ($p = 0,27$).
- La vitesse de lecture ne joue pas de rôle dans le rapport d’oublis ($p = 0,25$).

⁵³ Dont la valeur du résidu est supérieure à 1,5 fois l’écart-type des résidus.

- Le pourcentage général d’oublis est la variable qui contribue le plus fortement à la variation du rapport d’oublis : moins les lecteurs oublient au total de lettres, plus le rapport d’oublis de lettres entre les mots ayant un rôle différent est élevé.
- La compréhension^U joue un rôle important ($p < 0,01$) : elle est corrélée très positivement avec le rapport entre le pourcentage d’oublis de lettres dans les mots à rôle syntaxique et celui dans les mots à rôle sémantique.

9.2.2. L’explication des performances du lecteur par l’oubli de lettres

Cette fois, ce qui nous intéresse, c’est de savoir si la vitesse de lecture ou une des modalités de la compréhension peut s’expliquer par l’effet d’oubli de lettres. Pour répondre à cette question, nous procédons à nouveau à une analyse en régression multiple. Nous introduisons dans le modèle les trois variables issues de l’Analyse en Composantes Principales comme variables dépendantes et les variables sexe, pourcentage général d’oublis et différentiel d’oubli seront nos trois variables indépendantes. Le modèle M5 regroupe donc trois analyses de régression multiple.

Le tableau 17 présente les résultats généraux du modèle⁵⁴.

Modèle	Variables dépendantes	Corrélation	Pourcentage de variance expliquée	p
M5	Vitesse	0,42	18%	<0,01
	Compréhension ^U	0,38	14%	<0,03
	Compréhension ^D	0,09	0,7%	<0,93

Tableau 17 — Résultats généraux de la régression cherchant à expliquer les résultats en lecture par l’oubli de lettres. Modèle M5. Expérience 3.

On remarque que le modèle M5 explique deux variables dépendantes : la vitesse et la compréhension^U. Rappelons que la compréhension^U illustre les

⁵⁴ L’élimination des individus ayant un fort écart-type n’apporte un gain explicatif que de 3 points pour la compréhension. Pour ne pas alourdir la présentation, nous ne l’avons pas retenue.

phénomènes de compréhension quel que soit le type de lecture, à l'inverse de la compréhension^D qui oppose ce qui se passe entre lecture documentaire et lecture littéraire (cf. chapitre 7). La compréhension^U semble illustrer des phénomènes plus généraux, à l'œuvre dans toute compréhension.

			Ord. Origine	Variables explicatives		
				Sexe	Pourcentage général d'oubli	Différentiel d'oubli
Variables à expliquer	Vitesse de lecture	Paramètre	-0,41	-0,61	0,009	-0,03
		Coeff. bêta		-0,4	0,08	-0,05
		Significativité	0,58	<0,001	0,60	0,72
	Compréhension ^U	Paramètre	-1,7	0,1	0,03	0,2
		Coeff. bêta		0,11	0,43	0,45
		Significativité	<0,01	<0,39	<0,01	<0,01
Compréhension ^D	Paramètre	0,18	0,03	-0,0008	-0,04	
	Coeff. bêta		-0,03	-0,01	0,09	
	Significativité	<0,8	<0,85	<0,96	<0,61	

Tableau 18 — Paramètres, coefficients bêta et significativité des variables explicatives du modèle M5.

Le tableau 18 présente les coefficients de régression, les coefficients standardisés, et leur significativité, pour les variables continues ou les modalités. On observe que :

- la variable sexe intervient de manière significative dans la vitesse de lecture avec une vitesse de lecture plus importante pour les hommes⁵⁵.
- Le pourcentage général d'oubli explique significativement la compréhension^U, avec un oubli général plus important pour les meilleurs compreneurs.
- Le différentiel d'oublis explique, lui aussi, la compréhension^U avec un différentiel plus élevé pour les meilleurs compreneurs.

⁵⁵ Le paramètre est certes négatif, mais la vitesse de lecture est issue de l'axe 1 de l'ACP qui situait les vitesses les plus élevées à son côté négatif.

- La compréhension^D n'est expliquée par aucune des variables introduites dans le modèle.

9.3. Discussion

Cette étude fait apparaître une relation significative entre oubli privilégié de lettres dans les mots organisant la syntaxe et compréhension de textes. Conformément à l'hypothèse structurale, le différentiel entre oublis de lettres dans les mots à rôle syntaxiques et oublis dans les mots à rôle sémantique s'explique par le décalage entre le repérage de la syntaxe et le traitement des mots à rôle sémantique qui prennent place dans des structures anticipées. Chacun à leur manière, les deux modèles statistiques que nous avons employés vont dans le sens d'un rapport étroit entre la compréhension et un taux d'oubli de lettre plus important dans les mots ayant un rôle syntaxique.

- Le premier, expliquant la différence dans l'oubli de lettres, montre que le fait d'oublier significativement plus de lettres dans les mots fonctionnels est, en grande partie, lié à une meilleure compréhension. Si ce résultat est important pour mieux comprendre ce qui construit le modèle de Koriat, il l'est moins pour nos préoccupations pédagogiques : ce qui nous préoccupe, c'est d'améliorer la qualité de lecture, ce n'est pas de creuser le différentiel d'oubli entre les mots fonctionnels et les autres.

- Le deuxième, expliquant les performances en lecture, montre qu'une meilleure compréhension^U est liée à un plus fort différentiel d'oubli. Il ne faut pas s'étonner que le pourcentage de variance expliquée pour la compréhension^U ne soit que de 14% : d'autres éléments contribuent bien évidemment à la compréhension, comme la familiarité avec le sujet traité, l'empan de mémoire de travail (Daneman & Carpenter, 1980),... Cependant, les résultats montrent sans conteste que le bon compreneur est celui qui, à vitesse égale, anticipe avec plus de sûreté la construction de la phrase, ce dont témoigne le rapport entre les taux d'oubli de lettres dans les mots à rôle syntaxique et celui dans les mots à rôle non syntaxique.

En outre, il est important de souligner la corrélation négative entre le différentiel d'oublis et le taux général d'oubli de lettres : de manière générale, les lecteurs qui établissent efficacement ou tentent d'établir des cadres structuraux, ce qui se repère à l'aune de leur différentiel, sont en même temps ceux qui sont le plus attentifs au matériau graphique proposé.

La corrélation positive entre la compréhension en lecture et la capacité à avoir construit l'organisation syntaxique avant le traitement des mots semble contredire certaines descriptions classiques de l'acte lexique. En particulier, l'idée, reprise par Lecocq, d'une analyse syntaxique « on line », dépendante d'une identification des mots par un accès au lexique mental, et séquentielle, dans la mesure où *elle s'effectue pas à pas au fur et à mesure que les mots sont identifiés* (Lecocq, Casalis, Leuwers, & Watteau, 1996), devrait ici être questionnée. En effet, certains développements du modèle structural considèrent la formation de la structure syntaxique de la phrase (ou d'un groupe de mots) comme étant édiflée *a priori* à l'aide d'une prise d'information parafovéale (Koriat & Greenberg, 1996a; Saint-Aubin & Klein, 2001). La lecture experte ne relèverait donc pas d'abord d'une activité séquentielle qui implique l'identification des mots les uns après les autres afin de leur attribuer un rôle (syntaxique ou non) mais serait "pilotee" par le repérage d'unités syntaxiques qui organisent le contenu sémantique de la phrase parcourue.

Quoi qu'il en soit, le lien entre la compréhension^U et un différentiel d'oubli important suggère que l'habileté dans la construction des cadres syntaxiques (qui vont être remplis ensuite par un contenu sémantique) est prédictif d'une bonne compréhension. En conséquence, il est nécessaire de vérifier si l'entraînement de cette habileté provoque une amélioration de la compréhension. Ce sera un des objets de notre troisième partie.

Chapitre 10.

L'importance de la vision périphérique pour le modèle structural

10.1. Position du problème

Les expérimentations présentées jusqu'ici permettent de mieux comprendre les relations entre l'oubli de lettres et le rôle organisateur des mots dans la phrase. Néanmoins, d'autres facteurs semblent susceptibles d'influer sur cette relation. Nous rappelions plus haut les hypothèses de Koriat où il suggère une influence possible de la vision périphérique sur l'effet d'oubli de lettres. (Koriat & Greenberg, 1996a).

Ce que nous allons tenter de découvrir, c'est si le lecteur peut, à l'aide d'informations périphériques présentées très rapidement, établir des cadres syntaxiques et jusqu'à quelle distance cette construction est possible. Autrement dit, le lecteur est-il sensible à des variations syntaxiques et/ou lexicales dans la zone parafovéale et jusqu'à quelle distance ?

L'objectif de cette expérimentation est donc de tester notre hypothèse nulle n°3 qui stipule que les informations parafovéales ne permettent pas de prédire la structure syntaxique.

10.2. Méthode

Cette épreuve se passe sur un ordinateur dont la résolution d'écran est fixée en 800x600 pixels. Tous les affichages se font en Times New Roman de taille 13. Un expérimentateur se tient perpendiculairement au sujet. Un temps d'apprentissage est prévu.

10.2.1. Déroulement de l'épreuve

- Un point apparaît sur l'écran, qu'il est demandé au sujet de fixer.
- Après 500 ms, une phrase apparaît et reste affichée pendant 300 ms.
- Le sujet dit à l'expérimentateur le mot qui est apparu à la place du point de fixation. Si ce n'est pas le bon mot, l'expérimentateur passe à l'item suivant.
- Si le sujet a correctement donné le mot, trois phrases s'affichent.
- Le sujet doit sélectionner la phrase qu'il lui semble avoir vue initialement.
- On passe à l'item suivant.

10.2.2. Construction des triplets de phrases

Le plan expérimental comprend trois sources de variations :

- **La fréquence du mot fixé.** Il peut être soit rare, soit fréquent. On peut faire l'hypothèse que l'identification d'un mot rare (comme *botanique*) influe sur la suite de la tâche en consommant davantage de « ressources cognitives » que la reconnaissance d'un mot fréquent (comme *grand*).
- **Le rôle du mot** par où commence la variation entre les trois phrases. Deux catégories sont possibles : soit le rôle du mot qui initie la variation des phrases est sémantique (adjectif, verbe, nom), soit il est syntaxique (pronom relatif, conjonctions,...). Les deux autres phrases varient à partir de ce mot en le remplaçant pour l'une, par un mot à rôle sémantique et pour l'autre par un mot à rôle syntaxique.
- **La distance de la variation.** Les phrases se différencient au bout de 9, 15, 21 ou 27 signes entre le milieu du mot à fixer et le début du mot qui introduit la variation. Pour des questions de faisabilité, une tolérance d'un signe est acceptée dans ces distances. Ce qui est recherché ici, c'est l'éventualité d'une distance maximale à partir de laquelle l'effet ne serait plus perceptible. L'unité de mesure que nous avons choisie est le nombre de signes plutôt que l'angle

qui sépare le point de fixation du mot introduisant la variation. L'emploi d'une police proportionnelle où la largeur des lettres varie nous impose ce choix, dans la mesure où il semble impossible de fabriquer des triplets de phrases possédant un angle de variation identique. En outre, nous ne contrôlons pas la position de la tête du lecteur par rapport à l'écran. Cependant, à titre illustratif, nous avons calculé l'angle moyen pour chacune des distances de variation, en faisant l'hypothèse que les yeux du lecteur se trouvent à 40 cm de l'écran. Le tableau 19 les présente, ainsi que le nombre moyen de mots séparant le point de fixation du mot introduisant la variation.

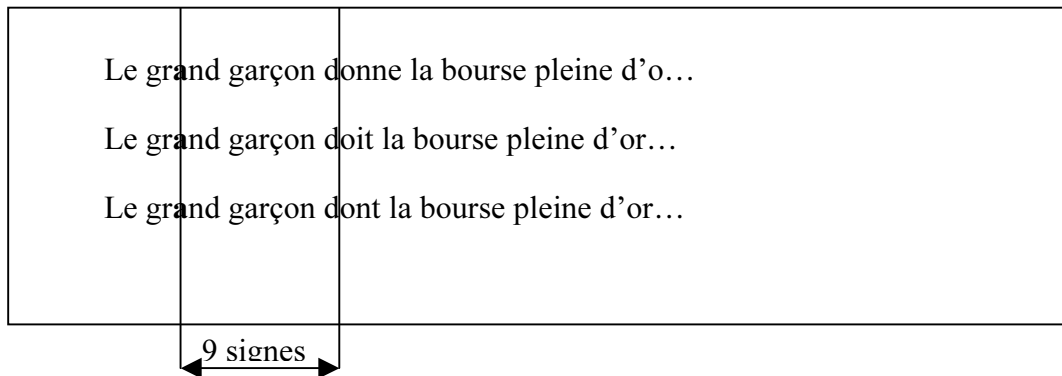
Signes	Angle	Mots
9 signes	3°59'	2,33
15 signes	5°18'	3,42
21 signes	7°1'	4,87
27 signes	8°4'	4,75

Tableau 19 : Equivalences entre le nombre de signes, l'angle moyen d'écartement et le nombre moyen de mots.

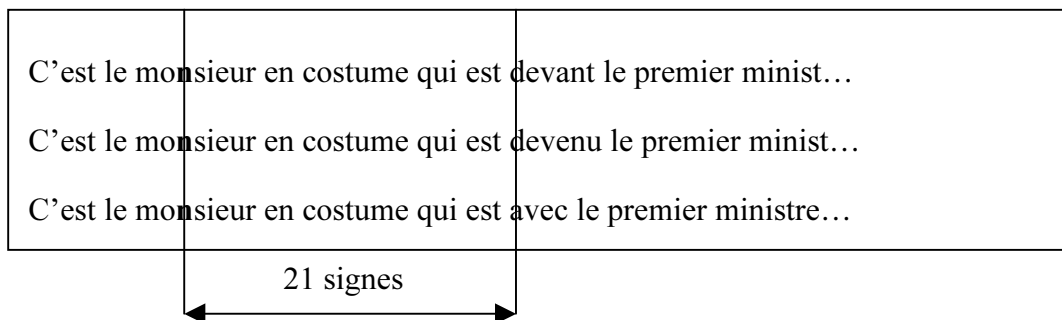
La mise en relation des différents angles rapportés dans ce tableau avec la figure 17 (page 58) montre qu'à partir de la distance de 15 signes la qualité des identifications est inférieure à 50 %.

- En conséquence, ce plan aboutit à seize combinaisons différentes ; nous produirons deux triplets par combinaison soit un total de trente-deux triplets. Ils sont récapitulés dans l'annexe 1.3, mais nous en donnerons ici deux exemples :

Exemple n°1 : il s'agit ici d'un triplet avec un mot d'appel fréquent (*grand*), avec une distance de 9 signes entre le milieu du mot fixé et le début des mots qui différencient les phrases (*donne, doit, dont*). La phrase à retrouver, d'abord affichée à l'écran en flash, a, comme mot débutant la variation, un mot à rôle sémantique (*donne*). La lettre en gras montre la place du point de fixation. La variation dans les deux autres phrases commence pour l'une par un mot à rôle sémantique (*doit*) et pour l'autre par un mot à rôle syntaxique (*dont*).



Exemple n°2 : il s'agit ici d'un triplet avec un mot d'appel fréquent (*monsieur*), avec une distance de 21 signes entre le mot d'appel et le mot qui différencie les phrases (*devant*, *devenu*, *avec*), et dont la phrase affichée à l'écran en flash contient comme mot débutant la variation un mot à rôle syntaxique (*devant*). La lettre en gras montre la place du point d'appel ; *devenu* a un rôle sémantique, *avec* un rôle syntaxique.



Les triplets sont construits pour avoir la plus grande similarité visuelle. On notera que, dans le cas où, dans la phrase présentée en flash, le mot différenciant a un rôle syntaxique, le mot différenciant à rôle sémantique d'une des deux autres phrases a une apparence visuelle très proche (*devant/devenu*), alors que l'autre mot à rôle syntaxique est fort différent (*devant/avec*). En revanche, quand, dans la phrase présentée en flash, le mot différenciant a un rôle sémantique, les trois mots différenciant du triplet sont visuellement proches.

Enfin, rappelons qu'on oblige le sujet à fixer le mot d'appel par la présence d'un point sur l'écran de l'ordinateur. C'est sa reconnaissance qui valide le

triplet. Grâce à cette contrainte, on s'assure que le mot débutant la variation des phrases est bien dans la zone parafovéale.

10.2.3. Population

L'ensemble de notre population a passé cette épreuve, soit un total de 64 individus (cf chapitre 7).

10.3. Résultats

Pour un des triplets, l'enregistrement n'a pas fonctionné, ce qui ramène le nombre de triplets que nous allons observer à 31.

	Modalité	Pourcentage de réussites	Ecart-type
Rôle du mot	Sémantique	31,10	10,12
	Syntaxique	38,14	7,80
Distance	9 signes	36,36	9,09
	15 signes	29,60	8,74
	21 signes	31,20	11,32
	27 signes	40,45	6,66
Fréquence du mot fixé	Fréquent	35,23	7,99
	Rare	33,73	11,33
Total		34,51	9,61

Tableau 20 — Exercice triplet : Pourcentage de réussite en fonction du rôle du mot, de la distance de la variation et de la fréquence du mot fixé.

10.3.1. Première analyse

Comment sont constituées les phrases cibles quand on les retrouve ensuite dans la liste des triplets ? Pour répondre à cette question, nous menons une analyse de variance, avec le pourcentage de réussite pour chacun des triplets comme variable dépendante. Les variables explicatives sont les valeurs qui correspondent aux trois sources de variation : la fréquence du mot fixé, la distance entre le mot fixé et le mot qui initie la variation entre les phrases et enfin le rôle de ce mot, sémantique ou syntaxique. Nous ajoutons dans le modèle l'ensemble des interactions d'ordre 2.

Le tableau 21 montre les résultats univariés issus de l'ANOVA.

Effet	Somme des carrés	Degré de liberté	Moyenne des carrés	F	p<
Distance	530,95	3	176,98	2,259	0,12
Fréquence	9,09	1	9,09	0,116	0,74
Rôle du mot différenciant	390,37	1	390,37	4,983	0,039*
Distance * Fréquence	46,43	3	15,48	0,198	0,9
Distance * Rôle	268,84	3	89,61	1,144	0,36
Fréquence * Rôle	69,61	1	69,61	0,889	0,36
Erreur	1410,05	18	78,34		

Tableau 21 — Tests Univariés de Significativité du Pourcentage de réussite. La ligne en gras montre la contribution significative.

Il montre que le seul facteur qui entretient un rapport significatif avec la réussite est le rôle du mot⁵⁶ initiant la variation.

La figure 35 illustre la différence de réussite à l'épreuve en fonction du rôle du mot introduisant la variation entre les phrases : on voit clairement que quand ce mot a un rôle syntaxique (comme dans l'exemple 2 donné plus haut) la réussite est meilleure que quand son rôle est sémantique (l'exemple 1 donné plus haut).

⁵⁶ Nous avons fait passer cette épreuve à 49 élèves de sixième et 48 élèves de troisième. Pour ne pas perturber la présentation que nous faisons, qui se focalise sur une population homogène d'adultes, nous ne les avons pas intégrés dans le calcul. Cependant, l'analyse donne des résultats proches de ceux des adultes :

- Pour les élèves de sixième, le rôle du mot initiant la variation a un effet quasiment significatif ($F(1, 25) = 4,0424, p < 0,055$) avec 29 % de réussite quand ce mot a un rôle sémantique et 38 % de réussite quand il a un rôle syntaxique.
- Pour les élèves de troisième, le rôle du mot initiant la variation a un effet quasiment significatif ($F(1, 25) = 3,5123, p < 0,07$) avec 30 % de réussite quand ce mot a un rôle sémantique et 37 % de réussite quand il a un rôle syntaxique.

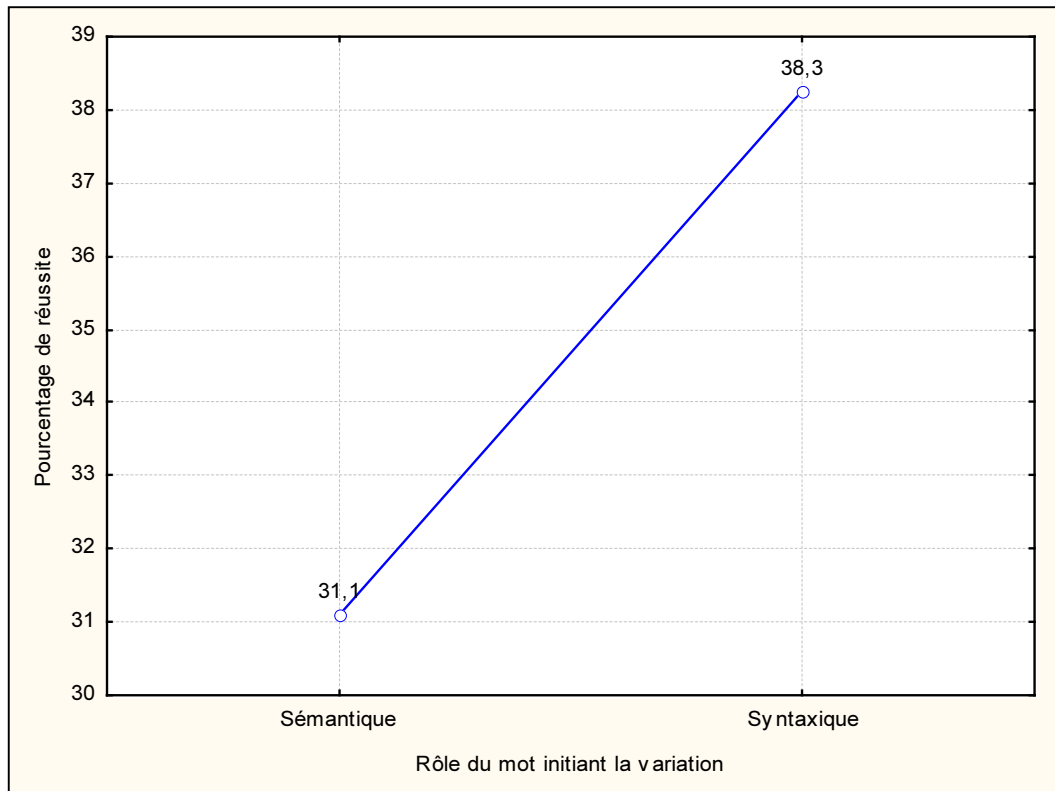


Figure 35 — Pourcentages (des moindres carrés) de réussite en fonction du rôle du mot introduisant la différence à l'intérieur des triplets.

On note également que la distance n'est pas significative ($p < 0,12$), ce qui est corroboré par les tests Post Hoc qui ne montrent pas d'oppositions significatives entre les différentes modalités⁵⁷. On remarquera quand même que les distances importantes ne semblent pas poser plus de problèmes que les distances courtes, puisque c'est à la distance de 27 signes que les sujets obtiennent les meilleurs résultats, la distance de 9 signes venant après.

10.3.2. Les échecs

Quand le sujet n'a pas correctement répondu, c'est-à-dire quand il n'a pas retrouvé la bonne phrase, il a néanmoins dû faire un choix. Si le mot différenciant a un rôle sémantique, les trois mots seront visuellement proches (« *sourd, soumis, sous* », « *salie, saline, selon* »,...). Quand le mot différenciant a un rôle syntaxique, les autres possibilités de réponses possèdent comme mot différenciant soit un mot à rôle sémantique qui est visuellement

⁵⁷ Les deux modalités s'éloignant le plus significativement sont la distance de 15 signes contre celle de 27 signes avec plus d'oubli pour 27. La significativité du test de HSD de Tuckey ne donne qu'un $p < 0,08$.

proche, soit un mot à rôle syntaxique qui est visuellement éloigné (« *par; pur; en* » ; « *mais, mains, et* »,...).

Dans la phrase à retrouver, le mot portant la variation	Dans les deux phrases intruses
a un rôle syntaxique	Le mot à rôle syntaxique ne lui ressemble pas
	Le mot à rôle sémantique lui ressemble
a un rôle sémantique	Le mot à rôle syntaxique lui ressemble
	Le mot à rôle sémantique lui ressemble

D'une certaine manière, on tend un piège au lecteur : une des phrases intruses est visuellement plus proche de la phrase projetée puisque les deux mots différenciants se ressemblent (dans ce cas, la structure n'est pas respectée) alors que l'autre lui ressemble moins puisque les mots différenciants sont visuellement très différents (dans ce cas, la structure est identique). L'alternative est claire : le lecteur va-t-il faire le choix de l'apparence visuelle contre la structure ou celui de la structure contre l'apparence visuelle ?

Par définition, ce schéma n'a pas à s'appliquer aux phrases dont le mot initiant la variation a un rôle sémantique puisqu'on fait l'hypothèse que c'est la perception et le traitement non conscients des mots fonctionnels qui sont possibles dans la vision parafovéale.

Quand un sujet a échoué à un triplet dont le mot différenciant est fonctionnel, nous garderons comme information la nature de sa réponse : a-t-il choisi la phrase visuellement ou structurellement proche ? Pour chacun de ces triplets, nous connaissons donc le pourcentage de réponses privilégiant l'apparence visuelle ou la structure de la phrase.

Un simple calcul du T de Student nous permet de savoir si les différences de moyennes sont significatives entre ces deux types de réponses. Les variances des deux groupes (Apparence proche vs. Rôle identique) étant homogènes (Test de Lévène : $F(1,28) = 0$; $p = 1$), nous pouvons accepter le résultat de

l'analyse du Test t pour des Echantillons Indépendants) présenté dans le tableau 22.

	Moyenne Apparence.	Moyenne Rôle.	valeur t	dl	p	F Levene	DI Levene	P Levene
Apparence Proche vs. Rôle Identique	42,01	57,99	-3,10	28	0,004	0,00	28	1

Tableau 22 — Résultats de la comparaison des catégories de réponse en cas de mauvaise réponse.

La figure 36 montre très clairement que les sujets choisissent plus souvent la phrase qui a une structure identique bien que l'apparence visuelle en soit plus éloignée⁵⁸. Ce résultat montre toute l'importance des informations parafovéales dans la construction du squelette syntaxique de la phrase.

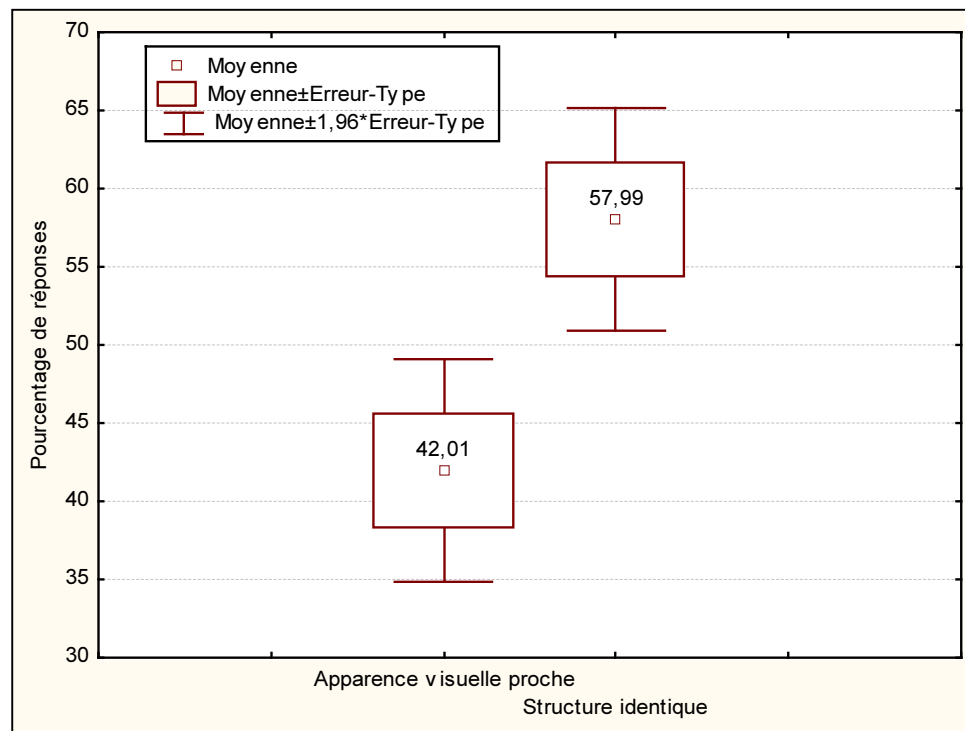


Figure 36 — Pourcentage de réponses aux triplets échoués selon l'apparence visuelle ou la structure identique.

⁵⁸ Pour les 49 élèves de sixième et les 48 élèves de troisième (voir note précédente) les résultats montrent :

- Pour les élèves de sixième, les réponses données pour une apparence visuelle proche sont moins importantes que pour une fonction similaire (41,1% vs. 58,9%). Cette différence est significative ($t = -2,93$; $dl = 28$; $p < 0,007$)
- De même, pour les élèves de troisième, les réponses données pour une apparence visuelle proche sont moins importantes que pour une fonction similaire (41,3% vs. 58,7%). Cette différence est significative ($t = -2,86$; $dl = 28$; $p < 0,008$)

10.4. Discussion

L'ensemble des résultats présentés dans ce chapitre permet de rejeter notre hypothèse nulle n°3 qui prévoyait que les informations parafovéales ne permettent pas de prédire la structure syntaxique. Au contraire, il semble bien que le lecteur est susceptible de prendre appui sur les informations issues de la vision périphérique pour construire le squelette syntaxique de la phrase.

La particularité de cette expérience, c'est de proposer un protocole où le choix se fait entre trois phrases également plausibles, et où il ne peut s'expliquer que par l'utilisation des informations parafovéales. La reconnaissance par le lecteur de la structure ne peut se faire ni à partir des informations perçues fovéalement, ni à l'aide du contexte linguistique dans lequel s'insère la phrase. La seule information dont il a conscience concerne le mot qu'il devait fixer et cette information ne peut lui servir en rien pour prévoir la suite de la phrase, puisqu'elle est présente dans les trois phrases proposées. Pour prendre une décision, il faut donc que le lecteur s'appuie sur les informations visuelles périphériques.

- Le premier résultat que nous obtenons montre que les triplets les mieux réussis sont ceux où le sujet a pu repérer, dans la phrase présentée initialement, une structure syntaxique bien particulière. Quand, de manière tangible, les mots fonctionnels organisent la structure de la phrase, les sujets sont mieux capables de la retrouver, comme s'ils avaient été sensibles à sa présence et que finalement, à ce stade, c'est du squelette syntaxique dont ils se souviennent.
- Le second résultat, qui concerne l'analyse des erreurs, est sans doute le plus intéressant. On aurait pu s'attendre à ce que les sujets soient attirés par la proximité d'une apparence visuelle. Il n'en est rien : ce qui prime dans le choix de la réponse, c'est la reconnaissance d'une structure à laquelle on a dû être sensible dans la phrase initialement projetée. Si on y a été sensible, c'est de manière totalement non consciente puisque l'œil se trouvait fixé dans la phrase bien en amont du lieu critique. Ce résultat est tout à fait congruent avec ceux que nous avons rappelés au chapitre 3 qui montraient que les informations non

conscientes peuvent servir de supports à des manipulations cognitives (Dehaene et al., 1998; Naccache & Dehaene, 2001).

Finalement, le fait marquant de cette expérience, c'est que les sujets organisent leurs réponses à travers la perception qu'ils ont du cadre syntaxique de la phrase. Car nous avons veillé à ce que les mots à rôle sémantique soient les plus proches possibles (si ce n'est identiques) dans les trois phrases du triplet : ce ne sont donc pas les informations sémantiques différentes qui guident le choix de la phrase, c'est uniquement la prégnance d'une organisation syntaxique bien particulière.

La lecture de ces résultats amène une autre remarque concernant cette fois la distance entre le point de fixation et l'endroit de la variation dans les phrases. Nous n'avons pas pu trouver de limites supérieures dans la mesure où, à 27 signes d'écart, on continue à observer un pourcentage de réussite important. Nous pouvons également regretter de n'avoir pas pu observer d'interactions, en raison d'un nombre trop faible d'observations à l'intersection des différents facteurs étudiés⁵⁹. En conséquence, et en s'appuyant sur ces premiers résultats significatifs, d'autres protocoles expérimentaux s'en inspirant seront nécessaires, en prévoyant d'éloigner encore le lieu de la variation et en construisant des ensembles en plus grande quantité.

Nous voudrions rappeler que les résultats présentés ici ne signifient pas que la seule source d'information pour établir un cadre syntaxique soit d'origine parafovéale. Nous avons déjà rapporté les expériences de Saint-Aubin qui fait apparaître les textes mot à mot sur un écran d'ordinateur. Cette présentation empêche toute prise d'information dans la zone périphérique de l'œil. Malgré cela, on observe un effet important de la fonction syntaxique du mot (Saint-Aubin & Klein, 2001). Ce qui se précise, c'est la relation dialectique entre les informations présentes en amont du lieu et du moment de la construction du cadre syntaxique (connaissance de la syntaxe de la langue, relation syntaxe-discours, connaissance du style de l'auteur,...) et la perception subliminale de la majorité des mots qui *crystallisent* l'organisation de la phrase. Ces deux

⁵⁹Sur les 32 triplets de départ, on n'en a plus que 16 dont le mot initiant la différenciation a un rôle syntaxique. Au croisement de ce rôle et des 4 distances, on ne trouve plus que 4 triplets.

types d'informations peuvent suffire, séparément, pour prendre une décision syntaxique (quoique avec moins d'efficacité), mais ce n'est que de leur union qu'on peut attendre les meilleures performances.

Pour finir, si ce résultat a de l'importance dans la compréhension des processus de lecture, il en a également dans une optique pédagogique. Nous faisons l'hypothèse à la fin du chapitre 9 qu'un entraînement spécifique devait permettre une amélioration de la compréhension. Il semble indispensable de réserver, à l'intérieur de cet entraînement, un moment particulier pour travailler la qualité de cette perception subliminale. Cet aspect n'est pas forcément le plus simple, dans la mesure où il semble éloigné des préoccupations des professeurs de français, surtout quand il est question de syntaxe ! Nous verrons pourtant au cours de la troisième partie de ce travail, l'apport d'un tel entraînement.

Bilan de ces études expérimentales

Nous nous sommes proposé dans cette seconde partie de mener des investigations à propos de points restés obscurs en consultant les travaux générés par l'effet d'oubli de lettres et qui semblaient à la portée des dispositifs expérimentaux que nous étions capables de maîtriser. En d'autres termes, cet effet d'oublis de lettres est-il une curiosité dont la connaissance n'a pas d'impact sur la pratique pédagogique ? Ou témoigne-t-il de processus dont la connaissance éclaire notre compréhension de ce qu'est la lecture, compréhension qui permet d'interroger les dispositifs pédagogiques existants autour de son apprentissage ? Incontestablement, sans invalider les modèles qui se réfèrent au rôle de la familiarité des mots dans leur reconnaissance, nos investigations confirment la robustesse des modèles qui se réfèrent à l'importance (sinon la nécessité absolue) pour la compréhension de l'établissement de la structure de la phrase et de ses constituants comme horizon d'attente des unités lexicales à contenu sémantique. Ces investigations montrent aussi que des informations, périphériques non seulement à la vision fovéale mais à l'empan de lecture lui-même jouent un rôle significatif dans l'élaboration de cette organisation syntaxique.

Il convient alors de se demander quelles traces de ces deux aspects se retrouvent déjà dans les prescriptions et les pratiques pédagogiques à l'école et au collège.

Troisième partie :

**Place de la syntaxe
dans la pédagogie de la lecture**

Chapitre 11.

Contexte théorique

11.1. Syntaxe et niveau de lecture

Dans le chapitre 9 de ce travail, nous avons mis en relation les performances en lecture des individus et le rapport d'oubli de lettres entre les mots à rôle syntaxique et à rôle sémantique. Nous considérons que ce rapport illustre la propension du lecteur à élaborer des hypothèses sur la structure de la phrase afin de construire un squelette syntaxique sur lequel viennent ensuite se greffer les informations sémantiques de la phrase. D'après le modèle de Koriat, plus ce rapport est élevé, plus la proportion de mots fonctionnels rejetés à l'arrière-plan cognitif est importante. Les résultats de nos expérimentations chez des lecteurs adultes, familiers du maniement de l'écrit, nous ont montré que cette habileté dans l'établissement des cadres structurels est plus importante chez les bons compreneurs, alors qu'elle ne les différencie aucunement en fonction de leur vitesse de lecture. Une plus grande maîtrise dans l'édification préalable des cadres syntaxiques est associée positivement à la compréhension d'un message écrit. Ce résultat est important. En effet, le rejet à l'arrière-plan des mots fonctionnels témoigne d'une création liminaire de la syntaxe et permet au lecteur de focaliser son attention sur les informations sémantiques contenues par la phrase. On pourrait donc penser que la moins bonne compréhension résulte au moins pour partie d'un traitement séquentiel des mots de la phrase sans anticipation de leur rôle, obligeant à mener de pair l'élaboration de la syntaxe et l'élaboration sémantique.

Le terrain expérimental mettant en relation les connaissances syntaxiques et la lecture n'est pas totalement vierge. Par exemple, les travaux de Cupples portent sur la relation entre les performances en lecture et les connaissances linguistiques de lecteurs adultes. Les expériences montrent que les meilleurs compreneurs sont plus performants que les autres dans des tâches de jugement syntaxique qui demandent de décider si deux mots présentés sous forme de paire peuvent partager la même fonction dans une phrase (Cupples & Holmes,

1992). A l'inverse, la réussite à une épreuve de jugement sémantique où les sujets doivent décider si les deux mots d'une paire ont un sens similaire ne permet pas de discriminer les différents niveaux de compréhension. Les auteurs suggèrent que les différences dans la maîtrise des classes de mots et de leur fonction pourraient être associées à des variations dans l'efficacité des processus syntaxiques et donc au niveau de compréhension en lecture.

L'article publié par Bentin étudie les effets du contexte syntaxique sur l'identification orale des mots et sur la capacité à déterminer et à corriger des erreurs syntaxiques dans des phrases présentées oralement (Bentin et al., 1990). La population étudiée (de 9 à 14 ans) se répartit en trois groupes : des lecteurs particulièrement déficients, des lecteurs médiocres et des bons lecteurs. Les lecteurs médiocres se différencient des bons lecteurs par leur plus faible compréhension en lecture, mais sont équivalents dans leur capacité à appliquer les règles de transformations grapho-phonétiques. Les résultats montrent que les bons lecteurs font significativement moins d'erreurs que les autres dans les tâches de jugement et de correction. Cependant, l'écart se creuse entre les groupes dans la capacité à corriger les erreurs syntaxiques. Indépendamment des lecteurs déficients, ce qui différencie les bons lecteurs des mauvais, c'est plus l'habileté dans l'utilisation de la syntaxe que la simple capacité à juger de la grammaticalité d'une phrase.

Une recherche plus complète a été menée par Gaux et Gombert. Comme pour les études précédentes, l'objectif est de comprendre les relations pouvant exister entre les connaissances syntaxiques et les performances en lecture, cette fois chez des pré-adolescents (Gaux & Gombert, 1999). Sept épreuves syntaxiques et morpho-syntaxiques⁶⁰ ont été conduites sur deux types de phrases agrammaticales : des violations morphémiques ('la petite fille met sa manteau') et des inversions de mots ou de groupe de mots. Ce qui nous

⁶⁰ Les sept épreuves sont : **Répétition** (d'une phrase présentée oralement), **Jugement** (la phrase est-elle correcte ou pas ?), **Localisation** (identifier l'endroit précis où se situe l'erreur), **Correction** (répéter la phrase corrigée), **Explication** (de l'agrammaticalité), **Reproduction** (à partir de phrases correctes, il s'agit de reconstruire des phrases présentant les mêmes incorrections que celle identifiée précédemment), **Identification de la fonction d'un mot ou d'un syntagme** (épreuve écrite où on présente une phrase avec un syntagme souligné et où on demande au sujet de retrouver dans une nouvelle phrase le syntagme ayant une fonction identique : on appelle ici fonction le sujet, le verbe, le déterminant, le complément d'objet direct et l'adjectif).

intéresse ici plus particulièrement, ce sont les résultats des épreuves syntaxiques, c'est-à-dire celles qui manipulent l'ordre de mots ou de groupes de mots. Les résultats montrent un lien élevé entre les différentes tâches portant sur l'ordre des mots et la compréhension en lecture⁶¹. On trouve également un lien important entre l'épreuve demandant une identification de la fonction d'un syntagme et la compréhension. Les auteurs considèrent que, chez les pré-adolescents, l'habileté dans la reproduction des inversions d'ordre des mots, qui est une activité purement syntaxique, influence directement la compréhension en lecture. A l'intérieur de leur paradigme expérimental, ils estiment que la conscience de l'ordre des mots est le facteur d'explication de la compréhension en lecture le plus important ; ces résultats renforcent l'idée de l'importance de la conscience syntaxique dans le traitement d'une phrase.

Une recherche assez similaire a été conduite par Layton : à partir des analyses de Gombert, elle cherche à observer des effets sur la lecture d'exercices de closure orientés sur les mots syntaxiques. Comme Gombert, ces auteurs trouvent une amélioration de la conscience syntaxique, attribuée à l'entraînement, mais aucun effet sur la lecture (Layton, Robinson, & Lawson, 1998).

On le voit, ces quelques trop rares études sur le rapport lecture/syntaxe montrent, par d'autres modalités expérimentales que les nôtres, les liens qui existent entre la compréhension et la perception de la syntaxe. Cependant, les expériences que nous venons de relater attirent deux types de remarques :

- Pour la plupart, elles demandent un jugement sur la grammaticalité d'une phrase. On a vu dans le chapitre consacré à l'imagerie cérébrale que le type de tâche influait grandement sur la géographie des aires cérébrales concernées. Or la différence est grande entre la construction des cadres syntaxiques pendant la lecture et le jugement sur la justesse de la construction d'une phrase. Quant à la correction d'une phrase irrégulière, on est là dans un processus qui, s'il

⁶¹ Encore une fois, on ne trouve aucun lien entre syntaxe et vitesse de lecture.

témoigne sans conteste d'une habileté dans les processus syntaxiques, reste très éloigné de la construction inconsciente de la structure de la phrase⁶².

- Elles font la part belle à une étude de la syntaxe dans ses modalités orales. Pour le moins, les indices qui permettent de traiter la syntaxe différent de l'oral à l'écrit, ne serait-ce que par la présence des informations visuelles parafovéales dans le traitement de phrases écrites et par l'absence d'informations subliminales à l'oral, du moins de nature linguistique, car les mimiques, le contexte, l'intonation, la situation de dialogue peuvent jouer un rôle à l'oral dans l'organisation d'un horizon sémantique.

Malgré ces remarques, il ne fait aucun doute que le rôle de la syntaxe est primordial dans les activités de lecture. Aussi pouvons-nous faire l'hypothèse que des investissements pour construire ces cadres syntaxiques influeraient, en retour, sur les compétences de lecture des apprentis, et, sans doute, particulièrement sur leur compréhension. C'est ce que nous tenterons de vérifier dans la suite de ce travail.

11.2. La place de la syntaxe dans la pédagogie de la lecture

11.2.1. A l'école élémentaire

De manière synchronique, nous nous en tiendrons au dernier texte du ministère de l'éducation nationale qui présente les programmes officiels pour l'école primaire (Ministère de la Jeunesse de l'Éducation nationale et de la Recherche, 2002b). Sur les parties portant spécifiquement sur le français et la maîtrise de

⁶² Fowler suggère que la capacité à corriger une phrase fait appel de manière bien plus importante à la mémoire de travail que la simple tâche de jugement qui est plus immédiate. Ce qui pourrait expliquer en partie les différences de niveau de lecture (Fowler, 1988). Cependant, dans les résultats obtenus par Gaux, l'empan de mémoire fait partie des variables contrôlées (Gaux & Gombert, 1999).

la langue⁶³, nous avons conduit une recherche d'occurrences lexicales de quelques entrées liées à la lecture :

	Cycle 2		Cycle 3	
Nombre total de mots	8766		7948	
Mot	111	(12,7)	29	(3,6)
Lecture	46	(5,2)	54	(6,8)
Ecriture	36	(4,1)	37	(4,7)
Phonétique (phonème+phonémique+phonétique)	21	(2,4)	2	(0,3)
Identification	18	(2,1)	7	(0,9)
Syllabe	18	(2,1)	0	(0,0)
Lettre	15	(1,7)	0	(0,0)
Phrase	9	(1,0)	13	(1,6)
Lexique (+lexical)	8	(0,9)	17	(2,1)
Syntaxe (+ syntaxique)	7	(0,8)	11	(1,4)
Mot-Outils	6	(0,7)	2	(0,3)
(Autres) Groupes de mots	4	(0,5)	3	(0,4)
Groupe (nominal)	2	(0,2)	0	(0,0)
Syntagme	0	(0,0)	0	(0,0)

Tableau 23 — Nombre d'occurrences sur le programme des cycles 2 et 3 de l'école élémentaire. Les nombres entre parenthèses indiquent l'apparition théorique du mot pour 1000 mots.

Pour le cycle 2, on observe très nettement l'extrême propension des programmes à faire travailler les élèves sur des unités inférieures au mot : les items *syllabe*, *lettre* et même *phonétique* sont bien évidemment de cette catégorie. L'accent est mis sur l'*identification* et se résume parfaitement dans ce passage : « Pour identifier des mots, l'apprenti lecteur doit avoir compris le principe qui gouverne le codage de la langue écrite en français : les lettres ou groupes de lettres (graphèmes) représentent le plus souvent des unités distinctives de la langue orale (phonèmes) assemblées en syllabes ». Cette problématique disparaît pratiquement au cycle 3, où on commence à recommander le travail sur des unités plus étendues et sans doute plus abstraites : on observe, en effet, l'introduction de la double catégorisation lexique/syntaxe qui suppose un regard sur le fonctionnement des phrases. Dans ce contexte, il peut sembler surprenant de voir que le terme de mots-outils disparaît quasiment au cycle 3. Dans ce document, les mots-outils sont définis

⁶³ Pour les deux cycles, nous avons extrait dans la partie MAÎTRISE DU LANGAGE ET DE LA LANGUE FRANÇAISE les sous-parties OBJECTIFS et PROGRAMMES. Pour le cycle 3, nous avons extrait en plus la partie LANGUE FRANÇAISE, ÉDUCATION LITTÉRAIRE ET HUMAINE qui comprend les sous-parties LITTÉRATURE (DIRE, LIRE, ÉCRIRE) et OBSERVATION RÉFLÉCHIE DE LA LANGUE FRANÇAISE (grammaire, conjugaison, orthographe, vocabulaire).

par la notion de fréquence (« Il convient que ce soient aussi les plus fréquents de la langue, pour l'essentiel des mots-outils... ») et il y a comme une incohérence à qualifier les mots de cette liste d'outils alors qu'ils ne sont que fréquents. En se privant de la terminologie linguistique qui utilise le terme de *fonctionnel*, voire de syntaxique, on passe à côté du rôle de ces mots dans l'architecture de la phrase. Certes, les listes des mots-outils et des mots fonctionnels se recoupent souvent, puisque les mots fonctionnels sont également très fréquents. D'ailleurs, ce qui préoccupe les auteurs des programmes de cycle 2, c'est bien l'identification rapide de ces mots, puisque la majorité des occurrences de *mot(s)-outils* apparaît à l'intérieur d'une partie plus générale intitulée « *Apprendre à identifier des mots par la voie directe* ».

Pour ce qui nous préoccupe ici plus particulièrement, la syntaxe, l'extraction du contexte dans lequel elle apparaît laisse entendre qu'elle doit principalement être abordée pour deux raisons :

- Les problèmes qu'elle pose dans l'élaboration des phrases à l'écrit. Par exemple, dès le préambule, le texte précise qu'à la fin du cycle 2, les élèves doivent pouvoir rédiger cinq à dix lignes, en maîtrisant les problèmes du vocabulaire, de la syntaxe et de l'orthographe.
- La complexité syntaxique de certaines phrases en perturberait la compréhension. Que ce soit au cycle 2 ou au cycle 3, il est fait référence de nombreuses fois à cet obstacle à la compréhension : certaines structures syntaxiques sont difficiles à comprendre. C'est le cas, par exemple, des phrases passives que beaucoup d'enfants comprennent comme si elles étaient des phrases actives. C'est aussi le cas des phrases complexes comprenant des relatives ou des conjonctives.

Le document d'application des programmes du cycle 3 envisage, lui aussi, des ateliers spécifiques de lecture pour les élèves qui, confrontés à des structures syntaxiques complexes, présentent des problèmes de lecture : l'enseignant « doit être attentif au fait que si, dans certains cas, c'est la langue qui peut faire obstacle (lexique rare, syntaxe complexe, enchaînement problématique des substituts du nom, connecteurs délicats à interpréter, relations temporelles

inhabituelles, etc.), » ... Les ateliers proposés sont donc conçus, avant tout, pour « renforcer [...] le repérage des structures syntaxiques des énoncés ». On préconise donc de faire segmenter la chaîne écrite pour que l'élève ait une « meilleure conscience de la cohésion du groupe syntaxique ». Enfin, notons que le rappel des objectifs des programmes précise explicitement que ces ateliers « visent à faire rencontrer de manière artificielle (dans une situation d'exercice) des mots, des types d'énoncés ou des formes syntaxiques que l'on a peu de chances de rencontrer fréquemment dans les textes qu'on lit par ailleurs » (Ministère de la Jeunesse de l'Éducation nationale et de la Recherche, 2002a).

11.2.2. Au collège

Le même type de recherche d'occurrences a été entrepris sur les documents présentant les programmes et les accompagnements de programmes pour les classes de 6^{ème}, 5^{ème} et 4^{ème}.

	6ème		5 ^{ème} et 4 ^{ème}	
	Programme	Accompagnement	Programme	Accompagnement
Nombre total de mots	3245	19527	4201	37641
Lecture	28 (8,6)	140 (7,2)	33 (7,8)	187 (5,0)
Ecriture	11 (3,4)	59 (3,0)	16 (3,8)	105 (2,8)
Phrase	9 (2,8)	51 (2,6)	8 (1,9)	70 (1,9)
Lexique (+lexical)	5 (1,5)	21 (1,1)	11 (2,6)	56 (1,5)
Mot	4 (1,2)	41 (2,1)	4 (1,0)	49 (1,3)
Syntaxe(+syntaxique)	1 (0,3)	12 (0,6)	1 (0,2)	18 (0,5)
Groupe (de mots)	1 (0,3)	18 (0,9)	0 (0,0)	15 (0,4)
Sémantique	0 (0,0)	6 (0,3)	1 (0,2)	10 (0,3)
Syntagme	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
Mot-Outils	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)

Tableau 24 — Nombre d'occurrences sur le programme de cycles 2 et 3 de l'école élémentaire. Les nombres entre parenthèses indiquent l'apparition théorique du mot pour 1000 mots.

On voit que le travail autour de la phrase prend une plus grande importance au collège. Autour de cette unité, deux grandes options se font jour :

- Montrer comment la phrase est gouvernée par le discours dans lequel elle s'insère (« le type de discours [...] rend fréquent le recours, dans la phrase, à des fonctions grammaticales particulières, comme les diverses formes d'expansion du nom dans le descriptif, les compléments circonstanciels de temps dans le narratif. »).
- Travailler sur la grammaire de la phrase qui s'attache, entre autres, « à identifier et à caractériser les principaux constituants de la phrase ».

Par ailleurs, au collège apparaît un travail portant sur les mots à rôle sémantique et leur organisation à l'intérieur de champs (lexicaux et/ou sémantiques) autour de trois domaines :

- Structuration et relations lexicales, afin d'explorer les principes de constructions des mots (affixes, dérivation,...) et les relations sémantiques entre les mots (synonymie, hyperonymie)
- Construction de champs sémantiques par regroupement de différentes entrées lexicales
- Le rôle du discours et de l'énonciation dans le choix du lexique.

Il faut cependant noter que le travail sur le lexique est supposé être le plus important, puisqu'il est écrit dans l'accompagnement aux programmes de 6^{ème} comme de 5^{ème} et 4^{ème} que « le lexique est le premier des outils de la langue : avant de s'interroger sur les règles d'assemblage grammatical, il faut disposer des éléments à assembler ». S'il semble probable que la syntaxe est un composant de cet « assemblage grammatical », la recherche des occurrences des mots syntaxe ou syntaxique laisse entendre que cette capacité à ordonner les mots est surtout utile dans *la mise en texte* où elle devra être *rectifiée*.

11.3. Discussion

Il est tout à fait clair que les programmes officiels, aussi bien pour l'école élémentaire que pour le collège, considèrent que la syntaxe est le lieu de

difficultés dans la maîtrise langagière : une syntaxe complexe est avant tout une cause d'erreur de compréhension. De ce fait, lorsqu'on travaille sur la syntaxe à l'école et au collège, on cherche à obtenir une production syntaxique correcte. Dès l'école maternelle, cet objectif est énoncé pour la maîtrise de la langue orale, et il perdure au cycle 2 et 3 pour la production écrite. On est ici dans une analyse *a posteriori* des éléments qui posent problème dans l'habileté langagière, ce que Gombert propose de ranger dans des activités d'ordre métalinguistique, qui semblent assez spécifiques des activités de l'école. Il considère que cette activité linguistique particulière qu'est la lecture nécessite un pilotage cognitif conscient. A l'inverse, il propose de qualifier d'épilinguistiques, les comportements qui ne demanderaient pas au sujet un contrôle conscient de ses activités linguistiques (Gombert, 1990). Toutefois, Gombert suppose que ce traitement épilinguistique est plus en rapport avec les processus de traitement de la langue orale, alors que l'écrit demande un accès conscient au fonctionnement langagier. En ce sens, les programmes de l'école sont cohérents avec cette interprétation, puisqu'on travaille la syntaxe pour en renforcer sa conscience explicite : nous sommes donc tout à fait dans une activité d'ordre métasyntactique, spécifiquement en rapport avec la lecture et l'écriture.

Le modèle de Koriat, dans la mesure où la construction du squelette syntaxique est non-conscient, implique la mise en place de processus que Gombert qualifierait d'épilinguistiques. Nous avons décrit en détail les deux principes sur lesquels la création des cadres syntaxiques pouvait se réaliser de cette manière inconsciente, avant même le parcours effectif de la phrase (ou du groupe de mots) :

- la théorie syntaxique nous a permis de mieux comprendre l'importance de l'architecture linguistique de la phrase et le rôle central de la classe fermée des mots fonctionnels (cf. chapitre 2). Nous rappelions alors les propos de Pinker : « Les traces, les cas, les X-barres et autres accessoires de la syntaxe [...] doivent faire partie de notre vie mentale inconsciente. »

- l'habileté perceptive permet d'extraire, à l'aide des informations parafovéales, des informations sur la fonction des mots, en particulier pour ce qui concerne la classe des mots « fonctionnels » (cf. chapitre 3).

D'une certaine manière, c'est le non-conscient, le subliminal, qui réunit la théorie X-Barre et la vision parfovéale. Aussi, il nous semble intéressant de voir s'il est possible de renverser la façon d'aborder le travail sur la syntaxe à l'école. Nous proposons, dans cette partie, une première approche didactique ayant comme objectif de permettre aux apprentis lecteurs d'améliorer leurs habiletés dans la construction épilinguistique de l'architecture des phrases. Si la réalisation effective des processus décrits reste non-consciente, un apprentissage spécifique serait-il susceptible de l'améliorer ? Nous avons évoqué (cf. chapitre 3) les quatre grands mécanismes qui sous-tendent un apprentissage perceptif. Il nous semble important d'en retenir tout particulièrement trois :

1. Une plus grande attention apportée aux formes discriminantes.
2. Une capacité à séparer psychologiquement des stimuli jusqu'alors indistinguables.
3. Une habileté à détecter une construction unique représentant une configuration complexe, en lieu et place d'une multitude de parties séparées.

Ainsi, ce que nous proposons, ce sera de porter explicitement attention aux formes qui discriminent les différentes structures syntagmatiques, à savoir les mots fonctionnels. Au préalable, si cette habileté structurelle est bien de nature épilinguistique, nous devrions voir son apparition dès l'entrée dans l'écrit, sans aucun enseignement explicite. Ce sera le sujet du prochain chapitre.

Chapitre 12.

Le barrage de lettres chez les élèves de cycle 2

Nous rappelions en introduction que peu de travaux ont porté sur la mise en relation de l'oubli de lettres avec les caractéristiques du lecteur. Le rôle de l'âge du lecteur a été souligné : plus les lecteurs vieillissent, plus l'effet d'oublis de lettres est marqué, ce qui conduit les auteurs à conclure à une prise de conscience progressive du rôle de certains mots dans l'organisation de la structure des phrases (Greenberg et al., 1998). Cependant, cette étude ne se situe pas dans une perspective diachronique : elle oppose des élèves d'âges très différents (7 ans vs. 13 ans) et ne répond pas à la question du moment où cette sensibilité aux rôles des mots dans la construction de la phrase s'élabore. Comme ces auteurs supposent la très rapide mise en œuvre de cette habileté, nous nous proposons d'en chercher l'apparition chez des élèves de cycle 2.

Ce que nous allons étudier, dans cette partie, c'est bien le rapport entre l'entrée dans le monde de l'écrit, au cours du cycle 2, et la mise en œuvre des deux modèles explicatifs de l'effet d'oubli de lettres : le modèle structural de Koriat et le modèle d'unités concurrentes de Healy.

12.1. Population et procédure

886 élèves de cycles II provenant de 45 classes différentes ont participé à cette série d'expérimentations. On dénombre parmi eux 116 élèves de Grande Section, 420 de Cours Préparatoire et 350 de CE1. Les épreuves se déroulent en fin d'année scolaire, au mois de juin.

Un texte de 116 mots, *Les baisers du Loup* (Solotareff, 2001), est distribué aux élèves, sans aucune lecture ni présentation par l'enseignant. La seule consigne,

après une épreuve d'essai sur un autre texte et avec une autre lettre, porte sur la tâche à effectuer⁶⁴ (voir l'annexe 1.2.2).

Le texte est présenté dans la fonte Garamond d'une taille de 24 points, avec un interlignage simple. Au total, la lettre *u* est présente 35 fois dans 34 mots différents.

La variable dépendante introduite dans le modèle explicatif est la moyenne d'oublis de la lettre observée. Les variables explicatives sont au nombre de six :

- la lettre *u* étudiée représente (ou non) le phonème [y] dans le mot (comme dans le mot *sur* contrairement à *qui*).
- la taille de ce mot.
- la position de la lettre à l'intérieur du mot. Pour représenter cette valeur, on calcule le rapport entre la position de la lettre et la taille du mot ; plus cette valeur est proche de 1, plus la lettre est en fin de mot.
- la fréquence du mot dans lequel la lettre apparaît ; cette fréquence est issue de la base FRANTEXT, qui donne des informations sur les mots du français contemporain entre 1950 et 2000 (New et al., 2001).
- le rôle du mot dans lequel la lettre est présente (syntaxique vs. sémantique). Cette variable a été construite à partir de la fonction effective des mots dans le texte : les noms, verbes, adjectifs et adverbes sont considérés comme étant de rôle *sémantique*, c'est-à-dire qu'ils transportent de l'information sur un personnage ou une action du texte. Les autres mots ont été considérés comme étant de rôle *syntaxique*, dans la mesure où ils organisent les mots à rôle sémantique entre eux et ne portent pas d'informations sémantiques propres : ce sont les conjonctions, les déterminants, les pronoms relatifs, les prépositions...
- le niveau scolaire auquel ce mot est présenté.

⁶⁴ Vous allez prendre le stylo et en parcourant le texte, en le lisant, chacun comme il peut, à sa vitesse, sans revenir en arrière, vous allez barrer la lettre « **u** » (l'écrire au tableau) à chaque fois que vous la rencontrez..

12.2. Résultats

Les résultats sont à examiner sous deux angles : d'une part, les données analysées permettent-elles de mettre à jour une différence entre les trois niveaux scolaires étudiés et d'autre part à l'intérieur d'un même niveau, une des variables explique-t-elle les différences d'oublis observées ?

Une analyse de covariance a été conduite sur ces résultats en introduisant comme variables indépendantes les variables décrites plus haut ainsi que les interactions entre le niveau scolaire et toutes les autres variables du modèle.

		Niveau					
		GS		CP		CE1	
		M	Err-T	M	Err-T	M	Err-T
Prononciation	Oui	31,42	2,51	16,14	2,51	18,64	2,51
	Non	31,38	1,97	18,53	1,97	20,85	1,97
Rôle du mot	Sémantique	30,96	2,04	11,18	2,04	10,51	2,04
	Syntaxique	31,85	3,03	23,49	3,03	28,99	3,03
Tous les mots		31,4		17,34		19,75	

Tableau 25 — Pourcentages et Erreurs-Type des oublis de lettres en fonction du niveau scolaire, de la prononciation de la lettre et du rôle du mot.

Les résultats présentés dans le tableau 25 nous montrent tout d'abord un taux d'oubli beaucoup plus fort parmi les élèves de GS (31,4%), que pour les deux autres niveaux (17,34% et 19,75%). Une première analyse de variance confirme la significativité de cette différence (Effet courant : $F(2, 87) = 5,42$, $p < 0,007$). Les tests Post Hoc précisent que seule la différence entre la GS et les deux autres niveaux est significative (LSD de Fischer $< 0,0001$ pour les deux autres niveaux).

Les autres résultats montrent qu'il n'existe d'effet significatif ni pour la longueur du mot ($F(1, 87) = 0,58$, $p < 0,44$), ni pour la présence phonétique du son [y] ($F(2, 87) = 0,70$, $p < 0,41$). En revanche, on note :

- *un effet de la position de la lettre dans le mot* ($F(1, 87) = 4,73, p < 0,04$) avec un plus fort pourcentage d'oubli à mesure que la lettre se rapproche de la fin du mot.
- *un effet du rôle du mot* ($F(1, 87) = 20,72, p < 0,0001$) avec, pour l'ensemble de la population, une moyenne d'omission de 28% pour les mots à rôle syntaxique contre 17% pour les mots à rôle sémantique.
- *un effet de la fréquence du mot* ($F(1, 87) = 6,94, p < 0,009$) avec une augmentation de l'oubli plus forte pour les mots plus fréquents.

Les interactions sont ici primordiales, puisqu'elles nous renseignent sur l'activation de ces différents éléments à travers le passage des élèves dans le cycle 2. On constate en premier lieu que l'interaction entre la présence phonétique [y] et le niveau scolaire n'a pas d'effet sur l'oubli de lettres ($F(2, 87) = 0,18, p < 0,83$), de même que l'interaction entre le niveau et la fréquence du mot ($F(2, 87) = 0,22, p < 0,79$). L'interaction entre le niveau et la longueur n'est pas non plus significative ($F(2, 87) = 0,19, p < 0,82$), ni celle entre le niveau et la position de la lettre dans le mot ($F(2, 87) = 0,44, p < 0,72$).

A l'inverse, le rôle du mot évolue significativement au cours du cycle ($F(2, 87) = 4,94, p < 0,01$). La figure 37 illustre l'évolution du rôle du mot au cours de la scolarité. Notons que les tests Post Hoc (HSD de Tuckey pour effectifs différents) précisent que les différences entre les types de mots ne sont pas significatives en GS ($p < 0,99$), mais commencent à l'être au CP ($p < 0,02$) pour se renforcer encore au CE1 ($p < 0,0001$).

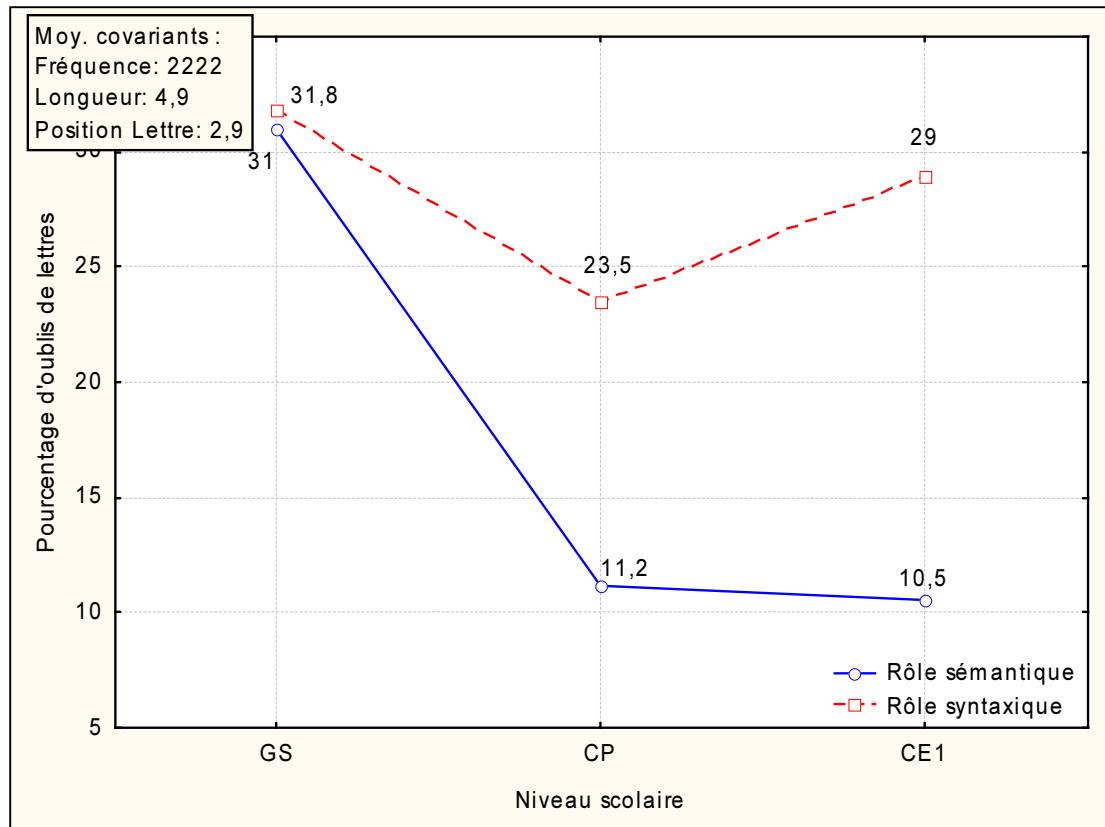


Figure 37 : Pourcentages moyens d'oubli de lettres en fonction du rôle du mot et du niveau scolaire (calculés pour les covariants à leur moyenne).

Si cette première analyse nous permet d'éclairer ce qui se construit au cours du cycle 2, elle ne nous permet pas d'évaluer le poids respectif des différents facteurs introduits dans le modèle explicatif. C'est ce que nous nous proposons de faire à l'aide de trois analyses en régression multiple, chacune d'elles prenant comme variable dépendante le pourcentage d'oubli de lettres dans un des trois niveaux scolaires. Les variables explicatives sont les mêmes que dans l'analyse de variance précédente, à l'exception du niveau scolaire. Si le modèle M6 expliquant l'oubli de lettres en GS n'est pas significatif (cf. tableau 26), la part de variance expliquée par les variables indépendantes (importante puisque près de 50% du phénomène est expliqué par les modèles M7 et M8) et le niveau de la corrélation multiple ($R = 0,7$ & $R = 0,73$), permettent l'examen approfondi des contributions de certaines des variables à l'effet d'oubli de lettres.

	Niveau de corrélation multiple (R)	Pourcentage de variance expliquée (R ²)	Significativité
M6 : L'oubli de lettres en GS	0,4	0,157	<0,38
M7 : L'oubli de lettres au CP	0,7	0,49	0,0007
M8 : L'oubli de lettres en CE1	0,73	0,53	0,00025

Tableau 26 — Résultats généraux des analyses de régression multiple cherchant à expliquer l'oubli de lettre en fonction du niveau scolaire. Modèle M6, M7 et M8.

Les tableaux 27, 28 et 29 présentent les coefficients de régression, les coefficients standardisés et la significativité de ces coefficients, pour les variables ou modalités introduites dans les modèles explicatifs. Il confirme les précédentes analyses en faisant toutefois disparaître le rôle de la position de la lettre dans le mot. Les deux modèles significatifs présentent deux variables très fortement corrélées, toutes choses égales par ailleurs, à l'oubli de lettres dans les mots. La première de ces variables, le rôle du mot dans lequel apparaît la lettre, voit son importance s'accroître au cours du cycle 2, son rôle étant inexistant en fin de GS (bêta de 0,04, $p < 0,87$) pour démarrer de manière spectaculaire en fin de CP (bêta de 0,59, $p < 0,01$) et se renforcer encore au CE1 (bêta de 0,66, $p < 0,001$). La fréquence du mot, quant à elle, ne subit pas la même évolution radicale. En effet, dès la GS, elle a déjà un rôle faible (bêta de 0,44, $p < 0,08$) qui reste approximativement le même au CP et au CE1, mais on ne note pas d'apparition soudaine du phénomène comme on peut le voir avec le rôle du mot.

Pour ce qui est du poids respectif de ces deux variables (dès que les modèles deviennent plus robustes), on note que, dès le CP, le rôle du mot tient une place prépondérante, avec un bêta de 0,59 contre 0,39 pour la fréquence. Cet écart se creuse encore au CE1, où la différence atteint environ 100 % entre l'impact du rôle du mot et celui de la fréquence.

		bêta	paramètre	p<
Rôle du mot (Référence = sémantique (m=31,5 ; σ =1,9))	syntaxique (m=32,1 ; σ =2,7)	0,04	0,29	0,87
Prononciation de la lettre (Référence = Non (m=16,6 ; σ =2,2))	oui (m=18,7 ; σ =1,8)	0,03	0,19	0,89
Fréquence du mot		0,44	0,0009	0,07
Longueur du mot		0,02	0,07	0,93
Position de la lettre dans le mot		0,24	1,18	0,36
Constante			25,75	0,0001

Tableau 27 — Modèle explicatif des scores d'oublis de lettres pour la GS (M6)

		bêta	paramètre	p<
Rôle du mot (Référence = sémantique (m=11,7 ; σ =2,5))	syntaxique (m=23,7 ; σ =2,5)	0,59	6,02	0,01
Prononciation de la lettre (Référence = Non (m=32,1 ; σ =2,5))	oui (m=32,1 ; σ =2,5)	-0,1	-1,04	0,47
Fréquence du mot		0,39	0,0011	0,05
Longueur du mot		0,21	0,81	0,30
Position de la lettre dans le mot		0,20	1,37	0,31
Constante			7,05	0,12

Tableau 28 — Modèle explicatif des scores d'oublis de lettres pour le CP (M7)

		bêta	paramètre	p<
Rôle du mot (Référence = sémantique (m=10,5 ; σ =2,4))	syntaxique (m=29,1 ; σ =3,5)	0,66	9,28	0,001
Prononciation de la lettre (Référence = Non (m=18,7 ; σ =2,9))	oui (m=20,9 ; σ =2,3)	-0,08	-1,14	0,54
Fréquence du mot		0,32	0,001	0,08
Longueur du mot		0,08	0,42	0,68
Position de la lettre dans le mot		0,31	2,86	0,12
Constante			6,46	0,26

Tableau 29 — Modèle explicatif des scores d'oublis de lettres pour le CE1 (M8)

Tous les effets décrits dans ces tableaux sont nets, c'est-à-dire que les autres variables sont contrôlées. On voit par exemple dans la colonne « coefficients » du modèle CP que, pour ce niveau, le fait que la lettre à barrer soit présente dans un mot à rôle syntaxique (variable qualitative) « contribue » à élever le pourcentage d'oubli de 6,02 points et que chaque lettre en plus (variable quantitative), dans le mot qui contient la lettre à barrer, augmente le pourcentage d'oubli de 0,81 point...

12.3. Discussion

Dans cette étude, nous voulions rechercher des éléments pour mieux cerner le moment où les élèves commencent à oublier de barrer des lettres de manière non aléatoire afin d'affiner la fourchette (7↔13 ans) mentionnée précédemment (Greenberg et al., 1998). On retrouve déjà dans les résultats présentés ici un certain nombre de données classiques de l'épreuve de barrage de lettres. Les deux grands modèles en concurrence pour expliquer ce phénomène, s'appuyant l'un sur le rôle de la fréquence du mot et l'autre sur le type de mot trouvent ici des données les confortant.

En GS, les oublis sont presque deux fois plus nombreux qu'en CP et en CE1, reflétant, sans doute, une faible discrimination du matériau écrit. Toutefois, les élèves en fin de GS commencent à connaître un certain nombre de mots très courants, souvent de manière isolée, et ce début de familiarité explique qu'on observe un effet de la fréquence des mots sur l'oubli de lettres. En cela, ces observations apportent crédit au modèle de Healy qui propose que si une reconnaissance globale du mot a lieu rapidement (par une grande familiarité avec ce mot), les autres processus à l'œuvre pour sa reconnaissance, comme l'identification des lettres le composant, s'arrêtent immédiatement. Notons que cet effet de fréquence se renforce très légèrement au CP et au CE1, traduisant sans doute, comme dans la population adulte, un usage plus important des mots très fréquents.

Le deuxième modèle dont nous voulions cerner l'engagement au cours de l'apprentissage de la lecture, est le modèle structural de Koriat où le lecteur extrait de prime abord le squelette de la phrase qu'il parcourt pour ensuite le compléter par du sémantique. Ce faisant, les mots organisant le contexte sont relégués à l'arrière plan et retiennent alors moins l'attention que les mots à rôle sémantique, au moment où le lecteur les voit nettement. Les résultats présentés ici trouvent effectivement à s'expliquer par un modèle structural de la lecture. La figure 37 montre très clairement que, dès que les élèves commencent à maîtriser les techniques de lecture, donc dès la fin du CP, on observe, toutes choses égales par ailleurs (et en particulier l'effet de fréquence du mot), que la distribution des oublis n'est plus aléatoire et que la différence entre les mots à rôle syntaxique et les mots à rôle sémantique est très sensible. Le plus frappant sans doute est la très forte diminution, par rapport à la GS, des oublis dans les mots à rôle sémantique, illustrant par-là la plus forte attention aux informations sémantiques au cours du déplacement de l'œil dans la phrase ; mais les travaux de Koriat avaient déjà montré un renforcement de l'attention sur les mots à rôle sémantique à partir du moment où le lecteur est amené à penser que le mot lu n'a pas de rôle syntaxique, par exemple si le mot est placé juste après un mot à rôle syntaxique non ambigu (Koriat & Greenberg, 1996b). En conséquence cette diminution très forte du pourcentage d'oubli dans les mots à rôle sémantique, dès le début de la lecture, éclaire bien l'habileté très tôt construite par le jeune lecteur à établir des cadres syntaxiques qu'il va garnir d'informations sémantiques.

Toutefois, si les modèles de Healy et de Koriat se trouvent confortés par les données présentées dans ce travail, on peut observer, par l'étude des interactions entre le niveau scolaire et les autres variables explicatives, que seul le rôle du mot dans la phrase interagit avec l'élévation du niveau. Ainsi, que l'on soit en GS ou au CE1 où la transcription grapho-phonologique devrait être mieux maîtrisée, le fait que la lettre U représente ou non le phonème [y] n'est d'aucune influence sur son oubli. De la même façon, la position de la lettre dans le mot qui influe sur l'oubli de lettres (plus la lettre est éloignée du début du mot, plus elle est oubliée) ne semble pas avoir un rôle différent qu'on soit en GS ou au CE1. Ainsi, la fréquentation du matériau écrit et l'expérience

de son fonctionnement au cours du cycle des apprentissages développe prioritairement l'habileté à élaborer des constructions syntaxiques, illustrée par la différence d'oubli de lettres entre les mots à rôle syntaxique et les mots à rôle sémantique.

Une autre information mise à jour par les données présentées dans ce travail chez les jeunes enfants est le poids relatif des deux grands modèles explicatifs de l'oubli de lettres, c'est-à-dire la fréquence ou le rôle du mot. On constate en effet que, dès le CP, à partir du moment où l'oubli de lettres cesse d'être dû à une méconnaissance de l'écrit, le poids du rôle du mot est plus important que celui de la fréquence. Qui plus est, les coefficients standardisés (bêta) présentés dans les tableaux 27, 28 et 29 montrent que plus le niveau des élèves croît, plus le poids du rôle du mot est important, alors que celui de la fréquence reste stable. Ces résultats sont en accord avec les expérimentations précédentes menées avec des adultes qui montraient que si le poids de la fréquence du mot dans ce phénomène se réduit encore, la différenciation du rôle du mot accroît, elle, son importance et se trouve corrélée avec le niveau de compréhension en lecture. Tout se passe comme si cette habileté à élaborer des cadres syntaxiques sur le déroulement futur de la phrase était associée à une meilleure maîtrise de la lecture ; conformément à l'hypothèse structurale, le cadre syntaxique préalablement établi pilote l'extraction des informations sémantiques et permet d'y consacrer attention.

Pour terminer, nous voudrions revenir sur une possible mésentente entre nos résultats, montrant l'apparition d'une habileté syntaxique au cours du cycle 2 et ceux présentés par Asmussen & Bastien-Toniazzo. Dans leur article sur la compréhension de phrases chez des enfants de début de CE1, elles concluent par l'affirmation qu'il n'est pas possible de mettre en évidence un effet du facteur syntaxique dans la compréhension de phrases (Asmussen & Bastien-Toniazzo, 2000). Nous voudrions noter que ce n'est pas parce qu'elles n'en observent pas, qu'il n'y en a pas : à cet égard, elles font remarquer l'existence d'un problème méthodologique quant au type de tâche qu'elles emploient, le jugement *a posteriori* sur les phrases. Nous avons vu précédemment que les activités cérébrales étaient très tributaires de la tâche à exécuter et tout

particulièrement quand il s'agit de juger syntaxiquement une phrase à partir d'une autre présentée préalablement : dans ces conditions (qui sont celles de l'expérimentation de Asmussen), il est indispensable de garder consciemment dans sa mémoire la construction de la phrase précédente, et en cela, nous sommes dans une problématique toute différente de la nôtre où la syntaxe est élaborée en premier lieu et de manière inconsciente. En conséquence, il n'y a pas incompatibilité entre ces deux résultats.

12.4. Observation des différences dans le modèle structural au cours de la scolarité

Si l'expérience précédente semble montrer que, dès l'entrée dans l'écrit, le lecteur met en place des stratégies qui lui permettent de construire des cadres syntaxiques de manière première au cours de sa lecture, nous pouvons également chercher à savoir si le différentiel d'oubli entre les mots à rôle syntaxique et les mots à rôle sémantique s'agrandit au fur et à mesure de l'âge. Pour cela, nous avons fait passer les mêmes épreuves que pour les adultes (voir chapitre 8) à des élèves de deux classes de sixième (soit 47 élèves) et à des élèves de deux classes de troisième (également 47 élèves). Les épreuves se sont déroulées à la fin de l'année scolaire, dans le courant du mois de juin.

Cette analyse ne prendra en compte que deux variables, à savoir la fréquence du mot où se trouve la lettre et le fait que ce mot a un rôle syntaxique ou sémantique. Les trois variables à expliquer sont le pourcentage d'oublis de la lettre dans le mot considéré pour les 6^{ème}, pour les 3^{ème} et pour les adultes.

Les tests multivariés sont significatifs pour le rôle du mot (Lambda Wilk = 0,88, $F(3, 612) = 26,57$, $p < 0,001$) comme pour la fréquence (Lambda Wilk = 0,9, $F(3, 612) = 21,75$, $p < 0,001$) : on observe que la tendance à oublier plus de mots à rôle syntaxique que de mots à rôle sémantique est présente pour les trois classes d'âge.

La figure 38 illustre les résultats obtenus.

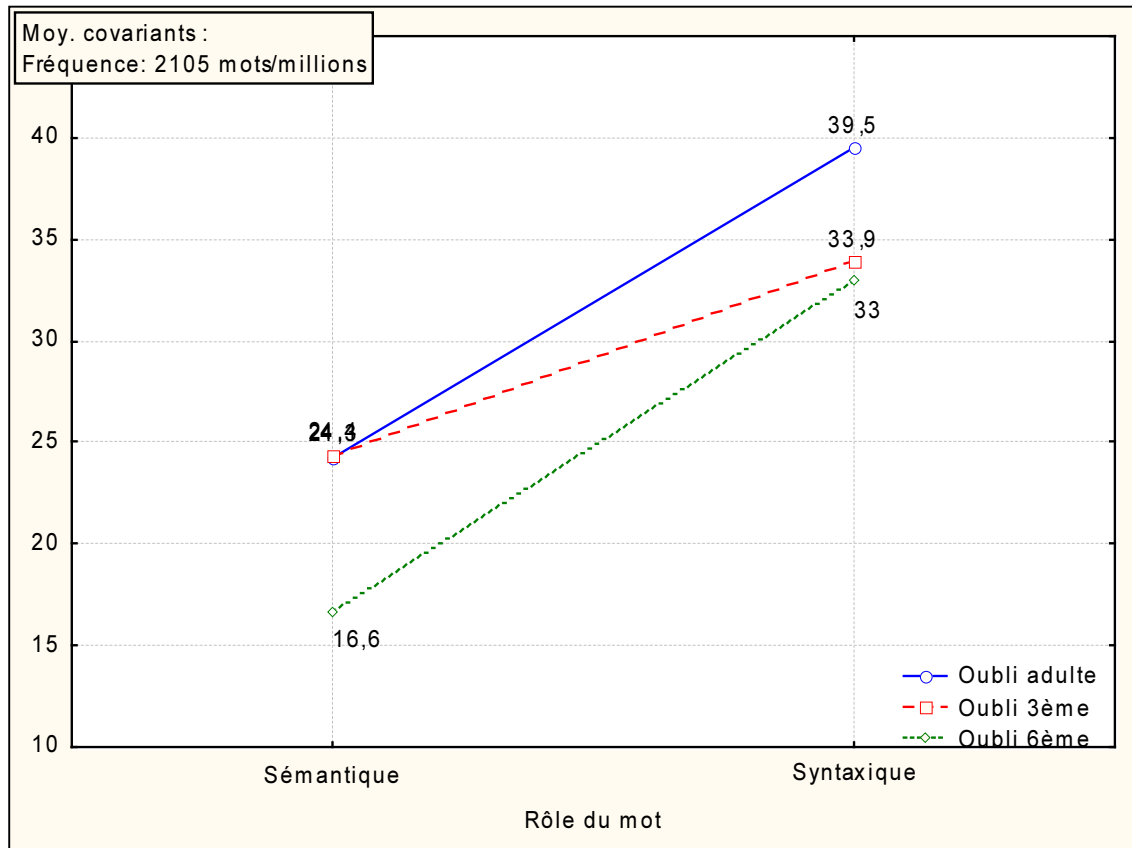


Figure 38 — Pourcentage d’oublis de lettres en fonction du rôle des mots et de l’âge des sujets.

Les résultats univariés donnent, aussi bien pour la fréquence des mots que pour leur rôle, des résultats significatifs à $p < 0,001$ pour les trois classes d’âge. Cependant, la figure 38 ne montre pas de différences significatives entre les trois niveaux, même si on remarque que la tendance est à un accroissement de l’oubli général en fonction de l’âge des sujets, et que les élèves de 6^{ème} oublient globalement moins de lettres que leurs aînés adultes (respectivement 16% et 24% pour les catégories lexicales à rôle sémantique contre 33% et 39% pour les catégories lexicales à rôle syntaxique).

Ce parallélisme, au décalage près, suggère que l’âge n’est pas responsable d’une plus forte proportion d’oublis dans les mots fonctionnels. Les résultats que nous avons présentés au chapitre 9 nous font penser qu’il vaut mieux chercher du côté de la compréhension pour différencier les individus.

Chapitre 13.

Peut-on entraîner le lecteur à construire des cadres syntaxiques ?

13.1. Introduction

Nous avons vu, au cours du chapitre précédent que, dès l'entrée dans l'écrit, se développe une habileté dans la construction de squelettes syntaxiques qui, d'après Koriat, doivent être par la suite peuplés d'informations sémantiques

La question à laquelle va tenter de répondre ce chapitre vise à savoir si les performances en lecture peuvent être améliorées par un entraînement à la construction des cadres syntaxiques. Doit-on considérer que cette habileté est une conséquence de l'acte lexique et se développe donc à notre insu, seulement par nos lectures et des rencontres variées avec l'écrit ? Un entraînement ciblé permettrait-il de développer cette capacité structurante ? En quoi ce développement améliorerait-il la performance de lecture ? La vitesse ? La compréhension ? Car l'intérêt du pédagogue est bien dans l'aide à la maîtrise de lecture, non dans l'augmentation d'un différentiel d'oubli de lettres entre mots à rôle syntaxiques et mots à rôle sémantique.

Afin d'explorer les possibles effets d'un entraînement de cette habileté à construire des squelettes syntaxiques, nous avons préparé un programme didactique reposant sur trois exercices, chacun exerçant respectivement un aspect particulier de ce qui pourrait être à l'œuvre lors de cette construction. Chacun de ces exercices tente d'aider le lecteur à prendre conscience du statut particulier des mots à rôle syntaxique, à mieux les reconnaître en situation, à les percevoir plus rapidement en travaillant avec les informations qui les entourent, sur le contenu sémantique qu'ils organisent.

Les trois exercices sont proposés grâce à un logiciel que nous avons développé spécialement dans le cadre de ce travail.

13.2. Les exercices

Ils sont conçus pour des élèves de collège qui vont les utiliser au cours de séances de travail d'une durée de 20 minutes pendant un trimestre. L'ensemble des exercices est présentée exhaustivement dans l'annexe 2.

13.2.1. Premier exercice : le cliquage sur les mots à rôle syntaxique

Le premier exercice que nous proposons travaille la reconnaissance des mots à rôle syntaxique ainsi que la vitesse à laquelle elle a lieu. Le travail est divisé en trois parties qui arrivent dans un ordre aléatoire. Pour chacune des parties, il s'agit de cliquer le plus rapidement possible, à l'intérieur d'un texte, sur tous les mots appartenant à une même catégorie grammaticale. Les trois catégories successives sont :

- les prépositions
- les déterminants
- les compléments, qui comprennent les différentes conjonctions et les pronom relatifs. Pour des raisons de facilité lexicale, nous l'avons appelée avec les élèves la catégorie des relationnels.

Ces trois catégories ont été choisies en fonction des résultats que nous avons présentés plus haut (cf. chapitre 8) qui montraient leurs rôles prépondérants dans l'oubli de lettres puisqu'elles étaient les natures les plus massivement oubliées, illustrant en cela leur mise à l'arrière plan au cours de la lecture et donc leur rôle important dans la construction du cadre structural de la phrase lue.

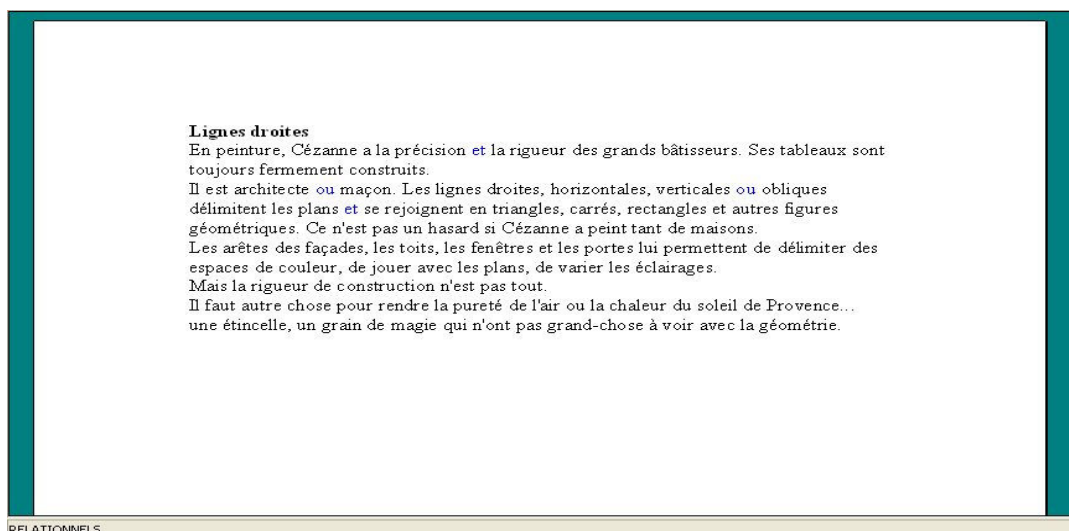
Trente textes issus, pour la plupart d'ouvrages de littérature jeunesse, ont été choisis et ont servi à fabriquer les trente séries de ce premier type d'exercice. Ces textes ont une longueur moyenne de 187 mots, le plus court en comptant 131 et le plus long 290. Le nombre d'éléments à trouver, pour l'ensemble des trois présentations du texte, varie de 41 à 92, avec une moyenne de 63.

Une présentation du texte est toujours limitée dans le temps pour obliger les élèves à travailler rapidement. Le temps maximum par exercice est compris entre 174 et 384 secondes, avec une moyenne de 291 secondes. Pour une présentation, le calcul du temps maximum d’affichage du texte est calculé par la formule $[Temps = (0.4 * N) + (0.4 * Q)]$ où N représente le nombre de mots du texte et Q le nombre de mots à trouver dans la catégorie concernée.

Déroulement

Le logiciel propose un texte à l’élève et la catégorie de mots sur laquelle il devra cliquer. Un clic sur un mot n’appartenant pas à la catégorie demandée ne provoque aucune action ; le mot se colorie en bleu s’il appartient à la bonne catégorie. Si l’élève trouve toutes les unités à chercher, on passe à la deuxième présentation du même texte, avec une nouvelle demande d’une autre catégorie à trouver. S’il n’arrive pas à terminer dans le temps maximum de la présentation, le logiciel présente la correction de la recherche en cours. Les unités oubliées apparaissent alors successivement sur l’écran en clignotant pendant 4 secondes. Si un élève est performant, il va finir les trois présentations avant la fin des cinq minutes : on lui présentera alors le texte suivant de la même catégorie. Au bout de cinq minutes, l’exercice se termine pour pouvoir passer au deuxième type d’exercices. Si l’élève est dans la première présentation du texte, il reviendra sur ce même texte à son prochain passage, sinon le logiciel lui présentera, la fois prochaine, le texte suivant.

Exemple d’écran



13.2.2. Deuxième exercice : l'exercice à trous

Classiquement, l'exercice à trous a essentiellement pour but de stimuler la réflexion métalinguistique de l'apprenant (Besse & Porquier, 1991). Cet exercice propose un texte dans son entier. L'objectif est ici de retrouver un mot du texte en s'aidant d'informations contextuelles, aussi bien d'ordre sémantique que syntaxique ou syntagmatique. La réalisation de cet exercice s'appuie sur le principe de l'explicitation d'un rapport entre l'élément manquant et les éléments du contexte : l'élément est déduit, calculé à partir du contexte. La réalisation de l'exercice suppose donc une réflexion métalinguistique. Toujours pour les mêmes raisons que celles présentées pour l'exercice précédent, la nature grammaticale des mots supprimés est soit un article, soit une préposition, soit un relationnel (notons que pour cet exercice il est fait explicitement appel à la dénomination exacte de la nature des mots : on aura trois possibilités pour cette dernière catégorie, les conjonctions de subordination, les conjonctions de coordination ou les pronoms relatifs).

Ici encore, trente autres textes issus pour la plupart de la littérature jeunesse ont été sélectionnés comme supports de travail. Leur longueur varie de 172 mots à 344 mots, avec une moyenne de 261 mots. Pour chacun des textes, dix mots sont choisis pour être retrouvés dans l'exercice. Pour chaque trou, l'élève a un temps maximum de recherche d'une minute : le temps maximum de l'exercice est donc de dix minutes, ce qui est plus important que le temps maximum prévu pour le passage d'un exercice (5 mn). En conséquence, si un élève voit son temps se terminer alors qu'il a fait moins de cinq mots dans le texte, le même texte lui est re-proposé à son prochain passage. A l'inverse, s'il a terminé son texte en moins de cinq minutes, un nouveau texte lui est proposé, jusqu'à ce qu'il termine son temps de travail.

Déroulement

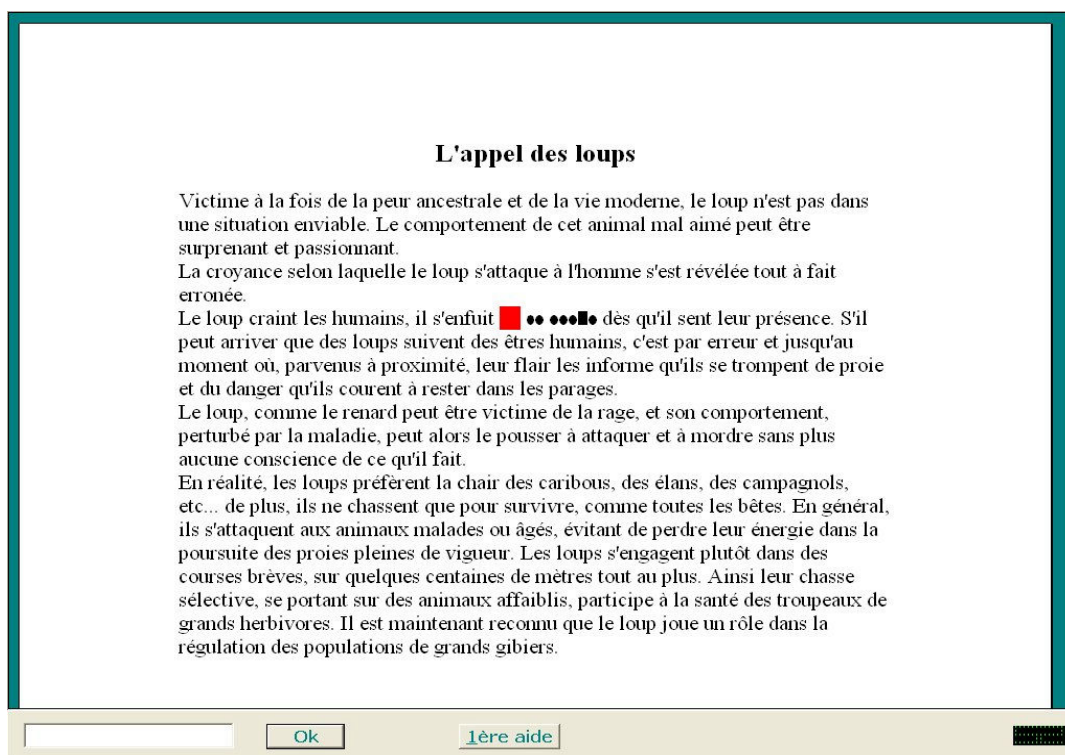
Un texte est proposé à un élève. L'ensemble du texte est visible et le premier trou s'affiche par un rectangle rouge d'une largeur exactement égale au mot qu'il masque. Comme les mots ont été choisis pour être des têtes de syntagme, on attire l'attention sur ce syntagme en l'écrivant d'une manière dégradée. Un

chronomètre égrène le temps restant pour trouver le mot. L'élève peut, à tout moment, demander à voir les aides qui y sont associées et qui sont présentées toujours dans le même ordre :

1. on affiche en bleu et dans sa police d'origine le syntagme introduit par le mot à trouver.
2. on indique à l'élève la nature grammaticale du mot recherché.
3. on présente le mot à trouver en police silhouette.

Si l'élève n'a pas trouvé la bonne réponse au bout d'une minute, le logiciel lui présente le mot qu'il recherchait et passe au mot suivant.

Exemple d'écran



13.2.3. Troisième exercice : les structures syntaxiques en vision parafovéale

Le troisième exercice proposé aux élèves diffère sensiblement des deux premiers. A propos des travaux de Saint-Aubin, nous avons montré le rôle (cf. chapitre 3 et 10) des informations de type parafovéal dans l'établissement des cadres structuraux (Saint-Aubin & Klein, 2001). En conséquence, ce que nous essayons d'exercer ici, c'est un meilleur traitement des informations

perçues par la vision parafovéale et leur meilleure intégration dans un processus de construction syntagmatique. Ce troisième exercice reprend, dans sa construction, un certain nombre des principes qui ont présidé à l'épreuve sur le repérage de structure syntaxique par la vision parafovéale (cf. chapitre 10). Il cherche à travailler sur la qualité de la perception dans un temps le plus court possible.

Cent cinquante-cinq groupes de trois phrases ont été construits d'une manière méthodique. Ces trois phrases démarrent de la même manière et utilisent un matériel lexical le plus proche possible. Elles se différencient au bout d'un nombre variable de mots suivant les triplets. Une des phrases est la phrase cible, celle qui a été présentée à l'écran. Une autre se différencie de la première mais le mot qui initie la différenciation est de même nature grammaticale. Une troisième, enfin, se différencie sur un mot non seulement de nature différente mais plus radicalement de fonction différente. Si le mot qui initie la variation est de rôle syntaxique dans les deux premières phrases, il sera de rôle sémantique dans la troisième. Mais, dans la mesure du possible, le mot initiant la différenciation entre les phrases (qui est de nature différente de celui de la phrase cible) a une apparence visuelle proche ou, du moins, des éléments visuels peuvent se ressembler (figure 39).

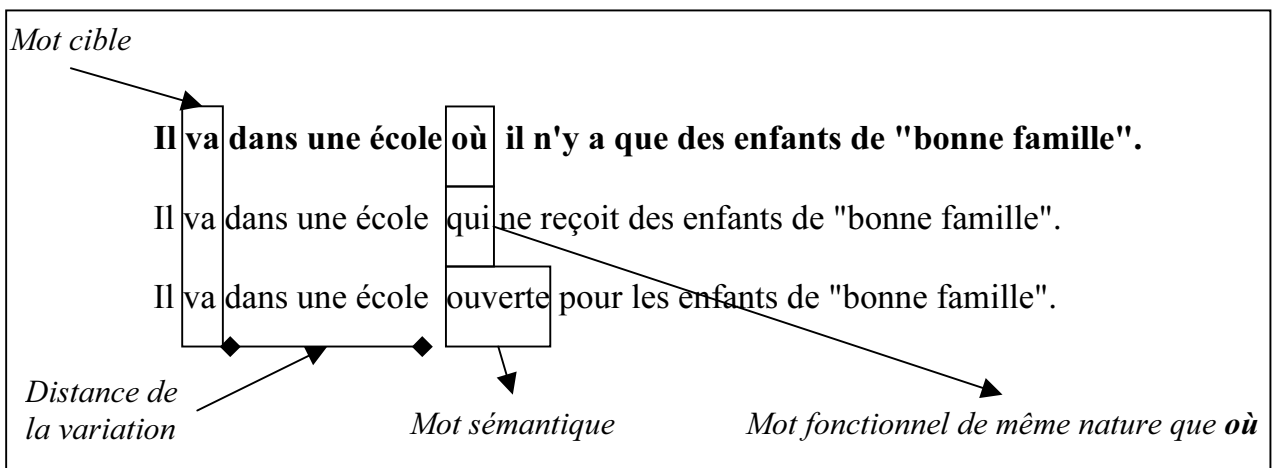


Figure 39 — Exemple de triplet. Le mot initiant la différenciation a un rôle syntaxique

Déroulement

La figure 40 présente l'ensemble des opérations possibles sur un triplet. Un point apparaît sur l'écran pendant 500 millisecondes, point qu'il est demandé au sujet de fixer. La première phrase apparaît et reste affichée pendant 500 ms. Le logiciel demande alors à l'élève de choisir parmi quatre mots lequel était sous le point qu'il devait fixer (étape 1). Si l'élève se trompe, on passe au triplet suivant.

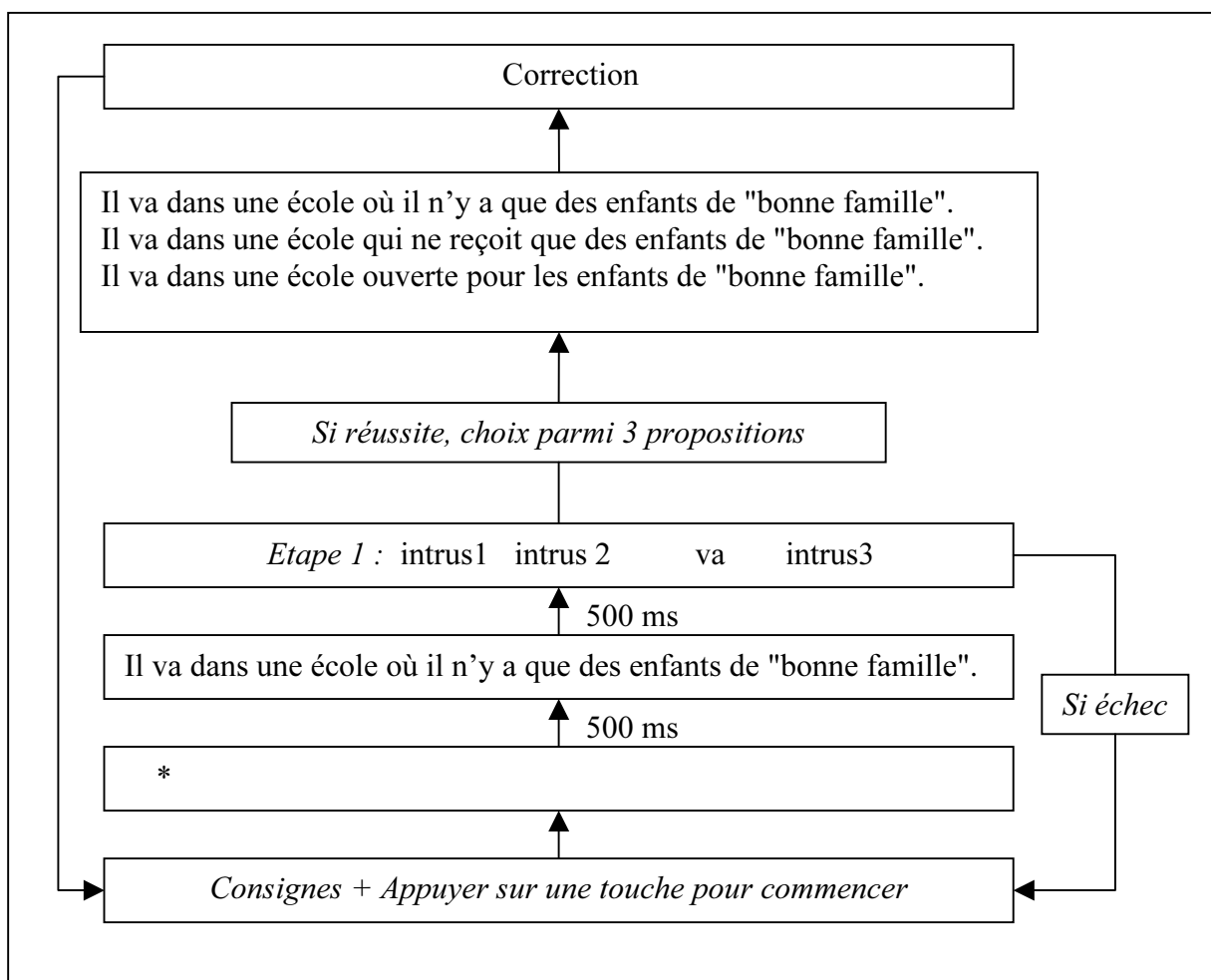
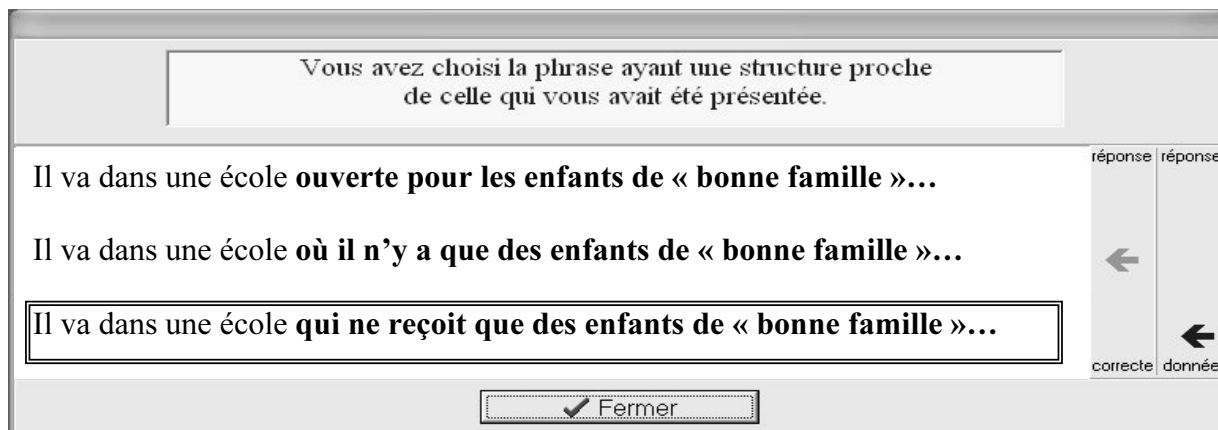


Figure 40 — Déroulement de l'exercice triplet

Si l'élève réussit, on lui présente alors un groupe de trois phrases dans son ensemble et on lui demande de choisir la phrase qu'il avait déjà vue. L'ordre d'affichage des phrases du triplet est aléatoire. Une correction est systématiquement apportée portant mention du type de réponse donnée.

Exemple d'écran de correction :



13.2.4. Récapitulatif

Les deux premiers exercices proposés dans le cadre de cet entraînement reposent sur l'idée qu'une meilleure connaissance des mots organisant la syntaxe de la phrase et qu'une plus grande familiarité avec leur fonctionnement favorisent, par la suite, la mise en place des cadres structuraux grâce à une plus facile identification des têtes de syntagmes. Le propos est bien de faire fonctionner de manière explicite des unités qui, en situation normale de lecture, le font implicitement.

Par exemple, le deuxième exercice proposé (exercices à trous) isole visuellement la structure que commande le mot fonctionnel à trouver. Non seulement la taille de la structure est mise en valeur dès le départ mais la première aide consiste précisément à l'explicitier, à la rendre visible par son changement de couleur. L'élève est ainsi amené à voir fonctionner chacun des mots fonctionnels qu'il a à trouver dans la structure où il s'intègre.

Le premier exercice (les cliques), lui, nécessite un traitement plus automatique. Il suffit simplement de chercher, de manière rapide, la succession des mots introduisant une structure identique. Il s'agit de conduire l'élève à focaliser son attention sur ces mots peu nombreux par rapport à la totalité des mots du français, mais très nombreux et indispensables dans chacun des textes écrits.

Par ce premier exercice qui nommait *a priori* la catégorie recherchée dans les textes, nous voulions travailler sur la catégorie en elle-même, non de manière abstraite mais par le repérage rapide des unités la constituant dans leur environnement naturel. Force fut bien de constater que, pour les élèves de collège engagés sur ces exercices, l'identification de tous ces mots à rôle syntaxique n'est pas aussi automatique qu'on pouvait l'espérer. Le premier exercice nécessite également une prise en compte des informations parafovéales. En effet, par sa dynamique qui oblige à se déplacer rapidement dans un texte tout en gardant une discrimination fine, il est probable que les informations périphériques sont utiles pour percevoir l'ensemble des mots demandés.

Dans leur ensemble, ces deux premiers exercices travaillent sur la conscientisation de l'existence des unités syntagmatiques, par un repérage explicite et le plus exhaustif possible des mots les initiant : il s'agit de bien maîtriser le fonctionnement de ces unités linguistiques par une analyse du contexte les gouvernant.

En revanche, le troisième exercice a une ambition différente. Cette fois, il ne s'agit plus de travailler sur l'explicite mais de devenir plus performant dans des activités dont les modalités restent le plus souvent hors de portée de la conscience. Grâce aux travaux déjà cités de Dehaene, nous savons maintenant que des éléments dont nous n'avons pas conscience influent non seulement sur nos réponses ou nos sensations, mais également aussi sur les processus cognitifs (Dehaene et al., 1998). C'est un des objectifs de cet exercice : concevoir que la présence d'un certain nombre d'éléments dans la vision parafovéale influe sur le processus cognitif de construction des cadres syntaxiques. Le chapitre 9 a déjà montré qu'un lecteur expert adulte est capable de réagir à de telles informations et qu'il s'en sert justement pour échafauder la syntaxe de la phrase. Ici, il s'agit de vérifier l'hypothèse que l'exercice systématique de ce processus participe à l'amélioration de la compréhension en lecture. Les travaux sur l'apprentissage perceptif (cf Chapitre 3) nous ont montré qu'il est tout à fait possible de conduire un entraînement pour améliorer l'habileté perceptuelle, à condition de positionner

les éléments à percevoir dans un système qui les intègre et où on les cherche. D'où l'importance de la correction donnée juste après l'essai. La fenêtre qui apparaît présente le lieu de la séparation des trois phrases proposées et l'ensemble des structures qui en découlent. L'élève peut ainsi mettre en relation le mot qui initiait la structure avec la structure construite. Il nous semble que cet exercice va dans le sens de ce que Goldstone propose : il s'agit bien là de porter attention aux formes discriminantes, de séparer psychologiquement des stimuli jusqu'alors indistinguables, et de détecter une construction unique sous des configurations multiples (Goldstone, 1998).

Ce qui ressort de l'observation des experts perceptifs et de l'apprentissage de cette expertise (cf. chapitre 3), c'est bien d'une part qu'ils organisent les stimuli par la présence d'un cadre conceptuel et d'autre part qu'on peut accélérer la maîtrise de ce cadre en rendant conscientes les informations décisives qui organisent la tâche (Biederman & Shiffrar, 1987). Pour Léontiev, « la prise de conscience d'un contenu est déterminée par la place qu'il occupe dans la structure de l'activité », ce qui implique que ce qui est accommodé dans la conscience d'un élève faisant une activité, c'est « ce qui entre dans l'activité en tant qu'objet d'une action qu'il réalise, en tant que but immédiat de cette action » (Léontiev, 1984). Pour cet auteur, un contenu ou un processus qui n'a jamais été amené à un niveau conscient, qui est toujours resté inconscient, reste « figé, rigide », donc peu manipulable et surtout incapable à la fois de réagir de manière créatrice face à un stimulus surprenant et d'améliorer ses propres procédures de traitement. C'est seulement lorsque le processus ou le contenu est susceptible d'être actualisé de manière consciente (le processus devient « éventuellement conscient »), qu'il acquiert la capacité de descendre à un niveau inconscient quand il est requis dans une autre tâche. C'est modestement ce que nous voulons tenter : afin d'améliorer l'habileté des élèves dans l'établissement de l'architecture d'une phrase (processus inconscient dans une situation de lecture normale), exposer à la conscience les éléments qui autorisent ce fonctionnement. Lors de l'exercice, nous hissons le processus du stade inconscient à celui d'« éventuellement conscient » pour pouvoir travailler sur ces éléments (mots fonctionnels repérés en vision

parafovéale), de façon à permettre à l'apprenti, en situation réelle, de tirer le meilleur parti de ce qui sera redevenu inconscient mais non ignoré.

En outre, la disposition des salles impliquant le partage d'un ordinateur entre deux élèves, l'un commandant la machine, l'autre le regardant travailler et inversement, nous a encouragé à vouloir ces exercices comme des supports de discussions métalinguistiques entre élèves. En accord avec les professeurs, nous avons sollicité les élèves pour qu'ils échangent entre eux pendant le déroulement des exercices. Nous avons transformé cette contrainte matérielle en une chance didactique, dans la mesure où cette coopération métalinguistique pouvait favoriser la prise de conscience par les élèves de l'importance de la syntaxe dans leur habileté de lecture. Cependant, cet aspect nous invite à la prudence quant à l'interprétation des résultats.

13.3. Les effets de l'entraînement sur la lecture.

13.3.1. Population

Deux groupes d'élèves de cinquième ont été constitués : un groupe expérimental qui a travaillé avec le logiciel que nous avons décrit et un groupe témoin. Deux classes de cinquième participent à l'expérience. Le groupe témoin a été constitué par tirage au sort parmi les élèves d'une troisième classe. A partir de cette souche témoin, l'appariement s'est fait pour le groupe expérimental en respectant les variables suivantes :

- la note globale obtenue à l'épreuve de français de l'évaluation 5^{ème} ($\pm 5\%$) ;
- l'âge de l'élève ;
- la note globale obtenue à l'épreuve de mathématiques de l'évaluation 5^{ème} n'a servi qu'à départager les élèves si plusieurs respectaient les deux premiers critères.

Pour chaque élève du groupe témoin, on retrouve, dans le groupe expérimental, trois élèves qui lui correspondent au regard de l'âge, de la performance en français et éventuellement de la performance en mathématiques lors des épreuves de l'évaluation cinquième.

Au total, nous avons constitué une population de 71 élèves qui sera ramenée à 67⁶⁵. 16 appartiennent au groupe témoin et 51 au groupe expérimental. Cette population se partage en 41 garçons et 26 filles avec une différence non significative entre les groupes (Chi² de Pearson : 0,015 ; dl = 1 ; p<0,91). L'âge, calculé par le nombre de jours écoulés entre la première évaluation et la date de naissance ne présente pas de différence significative entre les deux groupes (ANOVA âge x groupe : F(1, 64) = 0,27, p<0,61). La note en mathématique ne différencie pas, elle non plus, les deux groupes (ANOVA Mathématiques x Groupe) F(1, 65) = 1,19 ; p<0,29).

		Groupe Expérimental	Groupe Témoin	Total
Sexe	M	N = 31	N = 10	N = 41
	F	N = 20	N = 6	N = 26
Age moyen (en jours)		4696	4658	4687
Note moyenne en mathématiques		47,81	42	46,42

Tableau 30 — Répartition du sexe, de l'âge et de la note en mathématique suivant les deux groupes.

Enfin, une analyse de variance ANOVA entre les groupes constitués sur la note de français à l'évaluation cinquième corrobore leur homogénéité (F(1, 65) = 0,022, p<0, 89). (figure 41)

⁶⁵ Une erreur sur les listes d'inscription nous a fait perdre un élève dans le groupe expérimental et 3 élèves n'ont pas pu participer soit au pré-test, soit au post-test.

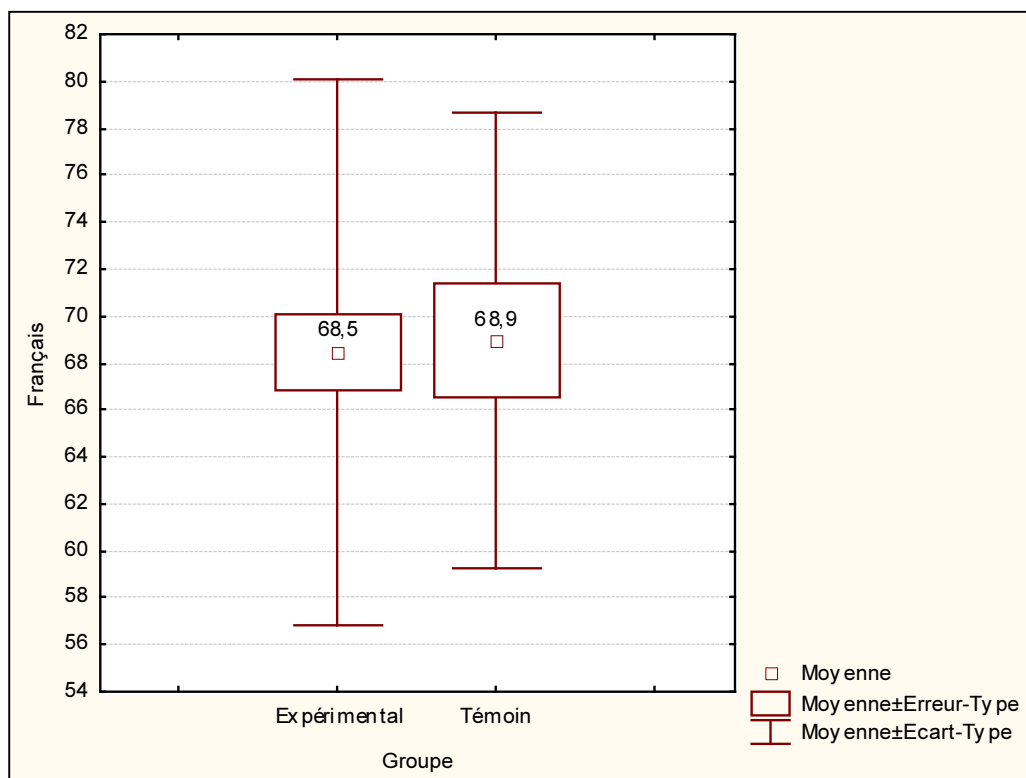


Figure 41 — Notes en français à l'évaluation nationale de 5^{ème}. Moyennes, erreurs-type et écarts-type pour le groupe expérimental et le groupe témoin.

13.3.2. Descriptif des séances de travail

Les élèves du groupe expérimental ont consacré une séance par semaine prise à l'intérieur de l'horaire normal de français à l'entraînement décrit plus haut. Pour les élèves du groupe témoin, ce temps est consacré aux activités habituelles de la classe de français en cinquième. Les équipements informatiques standard des collèges pour les classes de français obligent les élèves à n'utiliser un ordinateur que pendant la moitié de l'horaire⁶⁶ total. En conséquence, deux élèves se trouvent devant une machine, alors qu'un seul en tient les commandes. Comme nous l'écrivions un peu plus haut, nous avons profité de cette situation pour favoriser les échanges métalinguistiques autour des exercices, au prix d'un bruit parfois un peu plus élevé qu'à l'accoutumée.

L'entraînement s'est déroulé pendant tout le deuxième trimestre, soit un maximum de 10 séances pour chaque élève, 200 minutes d'entraînement

⁶⁶ On compte approximativement moitié moins d'ordinateurs que d'élèves par classe.

personnel et 200 minutes d'accompagnement d'un camarade (toujours le même) sur les mêmes exercices.

Le volume horaire de l'enseignement de français est donc inchangé et reste identique entre groupe témoin et groupe expérimental. Ce qui est introduit dans le groupe expérimental, c'est une substitution d'activité sur 25% du temps sur un trimestre. Pendant le temps consacré à l'entraînement, les élèves du groupe témoin continuent leurs activités habituelles de la classe de français : lecture, grammaire, travail sur ordinateur, travail en groupe, etc...

13.3.3. Les épreuves d'évaluation de la lecture

Les épreuves d'évaluation du niveau de lecture des élèves ont été passées, pour le pré-test, durant la dernière semaine du premier trimestre de l'année scolaire et pendant la dernière semaine du deuxième trimestre pour le post-test, donc à trois mois d'intervalle.

Pour évaluer la compréhension et la vitesse, nous nous sommes servi des deux épreuves déjà décrites au chapitre 7. Rappelons ici que la première épreuve porte sur la lecture de plusieurs textes courts, suivis d'une question sur un point explicitement présent dans le texte alors que la deuxième se déroule sur un texte long, suivi de plusieurs questions qui évaluent non seulement la compréhension de l'explicite du texte, mais aussi de son implicite.

Pour éviter la répétition de mêmes épreuves aux pré-test et post-test, nous avons opté pour un passage contrebalancé des épreuves : la moitié de l'effectif de chacun des groupes passera, au pré-test, l'épreuve sur les textes courts pendant que l'autre moitié passera l'autre épreuve. Pour les épreuves de post-tests, les élèves passent l'épreuve qu'ils n'ont pas faite au pré-test.

	Groupe expérimental (51)		Groupe témoin (16)	
Pré-test	Epreuve 1 (27)	Epreuve 2 (24)	Epreuve 1 (9)	Epreuve 2 (7)
Post-Test	Epreuve 2 (27)	Epreuve 1 (24)	Epreuve 2 (9)	Epreuve 1 (7)

Tableau 31 — Répartition des individus en fonction des épreuves. Les effectifs figurent entre parenthèses

Nous avons veillé à ce que la répartition des épreuves soit indépendante des résultats en français à l'évaluation 5^{ème}. Une analyse de variance cherchant à expliquer la note de français par le groupe et le type d'épreuve donne une interaction entre ces deux variables dépendantes non significative (Effet courant : $F(1, 63) = ,12, p < 0,73$), montrant par là que le groupe témoin et le groupe expérimental possèdent une répartition homogène suivant le niveau de départ des élèves entre les deux épreuves.

13.3.4. Les effets de l'entraînement sur la compréhension en lecture

Pour décider si l'entraînement a eu des effets sur la compréhension en lecture des élèves, il faut que nous ayons une notation homogène à l'intérieur des épreuves de pré-test et de post-test entre les deux épreuves. Aussi, afin de neutraliser la notation et la différence des épreuves, nous avons standardisé séparément les résultats des deux épreuves pour pouvoir raisonner toutes choses égales quant à la dispersion de chaque distribution.

Cette transformation permet de positionner chaque individu par rapport à une grille de performance commune aux résultats des deux épreuves. La valeur ainsi obtenue représente, pour chaque élève, la qualité de la compréhension au pré-test. La même opération est répétée pour le post-test, ce qui nous permet d'obtenir deux variables comparables représentant la compréhension, dans un jeu à somme nulle.

Une analyse de variance (ANCOVA) avec mesures répétées va servir à comparer les deux groupes de pratiques au regard de leurs progrès en compréhension⁶⁷. Pour contrôler la variation intragroupe, nous introduisons dans le modèle en covariants la note globale obtenue à l'évaluation 5^{ème} en français et les variables sexe et type d'épreuve au pré-test.

⁶⁷ Les moyennes, écarts-types et erreurs-types des différentes variables sont présentés dans l'annexe 3.6.1.

			dl	F	p
Effets Intergruppes		Sexe	1	1,56	<0,22
		Groupe	1	2,84	<0,1
		Type d'épreuves	1	0,19	<0,67
		Français	1	17,91	<0,001
Effets Intragruppes		Temps	1	0,58	<0,45
		Temps * Sexe	1	3,34	<0,08
		Temps *Groupe	1	8,88	<0,004
		Temps *Type d'épreuve	1	0,06	<0,82
		Temps *Français	1	0,33	<0,058

Tableau 32 — Résultats de l'ANCOVA expliquant les compréhensions. Significativité des effets intergruppes et intragruppes.

Les résultats présentés dans le tableau 32 montrent que :

- Pour les effets intergruppes, seul le niveau de départ illustré par la note en français est significatif. Que ce soit au pré-test ou au post-test, meilleurs sont les élèves à l'évaluation 5^{ème}, plus forte sera leur compréhension.
- Pour la variation intragroupe, et c'est ce qui nous intéresse en priorité dans cette analyse, le groupe (témoin vs. expérimental) permet d'expliquer significativement cette progression de la qualité de la compréhension pour les élèves du groupe expérimental par rapport au groupe témoin (figure 42).

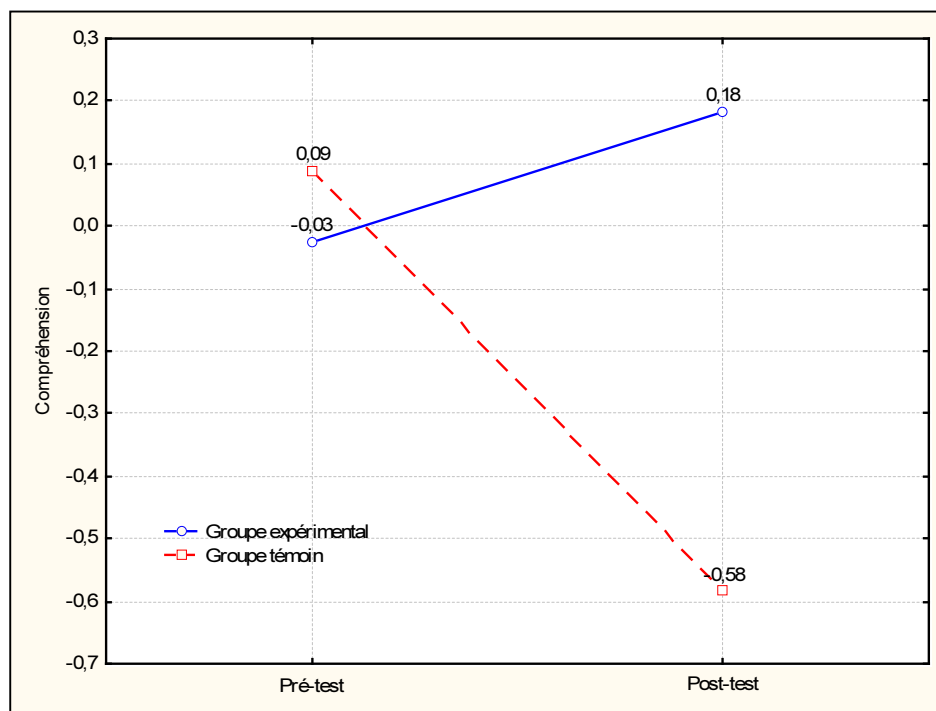


Figure 42 — Différence de compréhension entre le groupe témoin et le groupe expérimental après un trimestre d'entraînement

- L'interaction entre le groupe et le sexe a un effet quasiment significatif ($p < 0,08$) avec une plus forte progression pour les filles que pour les garçons (tableau 33).

	Garçons (N=41)		Filles (N=26)	
	Pré-test	Post-test	Pré-test	Post-test
Score	0,029	-0,395	0,015	0,09

Tableau 33 — Compréhension en fonction du sexe et de l'épreuve

13.3.5. Les effets de l'entraînement sur la vitesse de lecture

La même analyse de variance avec mesures répétées est reconduite en prenant cette fois comme variables dépendantes les résultats au pré-test et au post-test en vitesse de lecture⁶⁸. Le tableau 34 présente la significativité des différents facteurs explicatifs intra et intergroupes.

		dl	F	p<
Effets Intergroupes	Sexe	1	6,083	0,016*
	G	1	2,419	0,125
	Type d'épreuve	1	0,002	0,966
	Français	1	0,869	0,355
Effets Intrajgroupes	Temps	1	0,236	0,629
	Temps * Sexe	1	1,924	0,170
	Temps * G	1	0,020	0,889
	Temps * Type d'épreuve	1	0,003	0,960
	Temps * Français	1	0,191	0,663

Tableau 34 — Résultats de l'ANCOVA expliquant les vitesses de lecture. Significativité des effets intergroupes et intrajgroupes.

Les résultats montrent que seul le sexe est significatif, mais uniquement comme source de variation intergroupe et non dans son interaction avec la mesure répétée. Le tableau 35 montre que les élèves masculins ont une vitesse supérieure à celle des filles que ce soit au pré-test ou trois mois plus tard.

⁶⁸ Les moyennes, écarts-types et erreurs-types des différentes variables sont présentés dans l'annexe 3.6.2.

Sexe	N	Vitesse pré-test	Vitesse post-test
M	41	0,08	0,25
F	26	-0,13	-0,39

Tableau 35 — Vitesse de lecture en fonction du sexe et de l'épreuve

En revanche, on n'observe aucune variation de la vitesse chez les élèves s'étant entraînés pendant un trimestre.

13.4. Discussion

Les résultats confirment nos attentes. Pour commencer, il est intéressant de constater que la vitesse de lecture n'est absolument pas sensible à l'entraînement proposé, comme elle ne l'était déjà pas, chez les adultes, au différentiel d'oublis de lettres entre les mots à rôle syntaxique et les mots à rôle sémantique (cf. chapitre 7). Il se confirme que la vitesse est indépendante de cette habileté à la création première de structures syntaxiques.

En revanche, on observe une amélioration manifeste dans la compréhension des élèves ayant participé aux séquences expérimentales. Chez les lecteurs experts, nous avons constaté que les meilleurs compreneurs sont ceux qui construisaient le plus de structures syntaxiques, rejetant ensuite les mots fonctionnels à l'arrière plan cognitif. D'une manière parallèle, nous enregistrons dans ces résultats qu'un entraînement explicitement conçu pour améliorer la fréquentation des structures syntaxiques conduit les élèves à progresser de manière tout à fait significative dans la compréhension d'un texte. Cette constatation valide plusieurs de nos hypothèses :

- Une meilleure habileté dans la création des cadres syntaxiques de la phrase est associée à une meilleure compréhension en lecture
- Cette habileté est susceptible de s'entraîner au travers de deux (au moins) facteurs :

1. Un travail explicite sur la classe des mots fonctionnels qui *crystallisent la structure de la phrase* (Pinker, 1999). Une structure des phrases mieux maîtrisée, c'est un système où les informations lexicales vont mieux s'intégrer, pour le plus grand profit sémantique.
2. Un travail explicite sur une meilleure perception parafovéale de ces mots fonctionnels, avec une plus grande attention apportée aux formes discriminantes, amenant à une habileté dans la détection d'une construction unique représentant une configuration complexe, en lieu et place d'une multitude de parties séparées (Goldstone, 1998).

Nous devons cependant rester prudents devant les résultats obtenus. En effet, nous semblons tenir pour acquis que l'entraînement que nous avons conçu pour aider l'élève à bien anticiper l'organisation syntaxique dans laquelle va être traitée l'information sémantique, exerce effectivement ce pour quoi il a été conçu. On peut le penser, mais les progrès observés en lecture n'en établissent pas la preuve indiscutable. Il s'est évidemment passé quelque chose qui a un effet positif significatif sur la qualité de lecture des élèves. Ce qui s'est passé, comparé à ce qui ne s'est pas passé pour le groupe témoin, n'est pas lié à un temps supplémentaire consacré à l'horaire de français. Les autres différences entre groupe témoin et expérimental que nous n'avons pas contrôlées strictement sont liées au recours à l'informatique et à un accroissement probable des échanges métalinguistiques entre les élèves. Il est commun d'observer des effets de cette coopération sur les améliorations de résultats des élèves, et pas simplement dans le domaine de la lecture (Greenwood et al., 1984; Pigott, Fantuzzo, & Clement, 1986; Greenwood et al., 1987). Aussi, on pourrait imputer le gain de compréhension que nous observons dans ces expérimentations à un seul effet de la coopération avec des pairs. Il aurait fallu que le groupe témoin pratique, lui aussi, pendant une heure par semaine dans le cadre de l'horaire de français, une activité demandant une coopération entre les élèves. Le protocole expérimental en aurait été alourdi et aurait sans doute dépassé le cadre de ce travail, où nous voulions étudier l'impact d'un entraînement spécifiquement conçu pour améliorer, d'une part, la conscience

(au sens de Léontiev, voir plus haut) que peuvent avoir les élèves des unités qui permettent au lecteur d'établir le cadre structural de la phrase qu'il parcourt et, d'autre part, l'habileté perceptive (par la vision parafovéale) de ces mêmes unités. Cette analyse se voulait avant tout exploratoire dans le sens où il ne semble pas qu'un tel entraînement, centré explicitement sur ces aspects ait jamais été entrepris. Les discussions métalinguistiques entre les élèves, que nous avons encouragées, portaient soit sur le rapport entre les unités observées (les mots fonctionnels) et les structures construites, soit sur les structures perçues de manière parafovéale⁶⁹. Elles participent à la prise de conscience des mécanismes de construction de l'architecture syntaxique d'une phrase puisque, en extrapolant à partir de Léontiev, cette discussion est partie prenante de l'action des élèves à l'intérieur de l'activité d'exercices qui leur est proposée. Elles sont donc partie prenante de l'apprentissage et ne sauraient en être retirées, même pour fabriquer un protocole expérimental plus convenable. Dans ce cas, il reste à espérer que le groupe témoin a lui aussi consacré du temps à des échanges métalinguistiques entre les élèves à propos d'expériences linguistiques, ce qui est, qu'on le veuille ou non, nécessaire sinon inévitable, dans une classe de collège ayant à sa tête un enseignant de français compétent.

⁶⁹ Bien entendu, dans la réalité d'une classe, il est possible que toutes les discussions n'aient pas eu ce seul sujet... Cependant, nous en avons observé de nombreuses.

Chapitre 14.

Différence de résultats en fonction du type d'entraînement dans le seul groupe expérimental

Pour étudier les effets des types d'entraînement, nous construirons une variable représentant la différence de vitesse entre post-test et pré-test. Pour les pré-tests, les vitesses de lecture des deux épreuves sont séparément standardisées, cette valeur représentant la vitesse de lecture de chaque individu au pré-test. La même opération est réalisée pour les vitesses de lecture issues du post-test. On soustrait alors, pour chaque individu, la valeur de la vitesse obtenue au post-test par celle obtenue au pré-test. Cette nouvelle valeur constitue la différence de position de chacun des individus, à l'intérieur de la population, pour ce qui concerne la vitesse de lecture. Une valeur positive signifie que l'élève a progressé dans la population, une valeur négative qu'il a régressé dans un jeu, au niveau global à somme nulle. Y a-t-il des types d'entraînement pour lesquels les bénéficiaires se sont hissés dans la distribution ?

La même opération est effectuée en ce qui concerne la compréhension.

Le groupe expérimental comporte trois sous-groupes qui se différencient par une répartition du temps de travail entre les trois types d'exercices différents. Rappelons que les séances de français pendant lesquelles se déroule l'entraînement sont divisées en deux parties où se succèdent deux élèves sur une même machine ; un élève travaille donc pendant vingt minutes avec le logiciel d'entraînement. Cette durée est elle-même divisée en quatre moments : les trois premiers voient se succéder les trois exercices décrits plus haut, chacun pendant une durée de cinq minutes, puis un des trois exercices est poursuivi pendant le quatrième temps, soit cinq minutes.

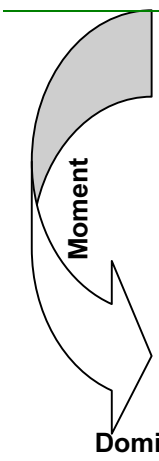
Sous-groupe				Durée	
	1	2	3		
	1	Ex n°2 : exercice à trous	Ex n°1 : clique sur natures	Ex n°3 : triplets	5mn
	2	Ex n°1 : clique sur natures	Ex n°2 : exercice à trous	Ex n°2 : exercice à trous	5mn
	3	Ex n°3 : triplets	Ex n°3 : triplets	Ex n°1 : clique sur natures	5mn
	4	Ex n°2 : exercice à trous	Ex n°1 : clique sur natures	Ex n°3 : triplets	5mn
	Dominante	Exercice à trous	Clique sur natures	Triplets	

Tableau 36 — Répartition des exercices dans une séance d'entraînement de 20 minutes.

Le tableau 36 montre la répartition des exercices pendant une séance d'entraînement. Il permet de voir également que l'exercice sur lequel l'élève doit retravailler encadre la séance, de façon à ne pas être exécuté de manière simultanée. L'ordre des exercices est le même dans les trois groupes.

En conséquence, tous les élèves s'entraînent à partir d'un socle commun auquel s'ajoute, pour chacun d'eux, un passage supplémentaire sur un des trois exercices qui constituent l'entraînement.

L'objectif de cette construction est de rechercher l'influence spécifique éventuelle d'un des exercices sur l'amélioration de la compréhension en lecture. Cependant, une rapide observation des séances d'entraînement laisse entrevoir des différences de comportement importantes à l'intérieur de chacun des groupes : élèves plus ou moins assidus pendant les séances, investissements différents dans l'entraînement....

Le logiciel conserve la trace, pour chacun des exercices, de la quantité de matériau travaillé et la réussite à chacun des items présentés. L'objectif que nous nous fixons maintenant étant de chercher si la quantité d'exposition et/ou le degré de réussite aux exercices expliquent une amélioration de la

compréhension, nous garderons comme variables décrivant l'entraînement les valeurs véritables du volume de l'activité et de la réussite dans chacun des trois types d'exercice.

14.1. Population

Les analyses porteront seulement sur 48 élèves parmi les 51 du groupe expérimental. Trois élèves ont été supprimés parce que les données sur leur entraînement ont été endommagées et se sont révélées inexploitable.

Le gain de compréhension est homogène entre les trois groupes ($F(2, 45) = 0,28645, p < 0,76$), comme l'est la différence de vitesse de lecture entre le début et la fin de la période d'entraînement ($F(2, 45) = 1,3758, p < 0,27$)⁷⁰. Ce résultat illustre bien le fait que, plus que l'affectation à tel ou tel groupe, c'est la quantité et la réussite qui sont susceptibles d'apporter des réponses quant à l'efficacité respective des exercices proposés.

14.2. Procédures d'analyse

Nous analyserons les résultats à l'aide de régressions multiples avec, comme objectif de déterminer le poids respectif du volume de l'activité et de la réussite dans chacun des exercices. Nous introduisons dans les modèles outre les variables indépendantes décrivant l'exposition et la réussite, les covariants Sexe, Type d'épreuve au pré-test et Résultat au pré-test, pour contrôler leurs effets respectifs.

Dans la mesure où ces analyses sont avant tout exploratoires et ont pour objectifs de mieux comprendre le rôle des différentes activités, nous affinerons les modèles de départ en éliminant les individus dont les résidus sont supérieurs à une fois et demie l'écart-type de l'ensemble des résidus de la population de départ. Toutefois, nous n'utiliserons cette procédure que si les premiers modèles se sont avérés significatifs.

⁷⁰ Voir l'annexe 3.7.1 pour une description de la moyenne des groupes

Pour chacun des exercices, une variable sert à décrire le volume de l'activité de l'élève et une autre représente la qualité du travail fourni par l'élève. Ce seront :

	Variables utilisées pour décrire	
	le volume d'activité	le degré de réussite
exercice de cliques	nombre de mots sur lesquels l'élève était susceptible de cliquer.	nombre de mots sur lesquels l'élève a cliqué divisé par le nombre de mots sur lesquels l'élève était susceptible de cliquer.
exercice à trous	nombre de mots que l'élève a dû chercher au cours de son entraînement	nombre de mots trouvés par l'élève divisé par le nombre de mots que l'élève a dû chercher au cours de son entraînement.
exercice triplet	nombre de triplets présenté à l'élève, c'est-à-dire ceux pour lesquels il a correctement identifié le mot cible.	deux résultats sont à prendre en considération : tout d'abord l'élève peut retrouver la phrase projetée, c'est la meilleure réponse. Mais il peut aussi choisir la phrase qui possède une structure identique à celle projetée ; dans ce cas, on considèrera la réponse comme acceptable dans la mesure où l'élève a été sensible à la construction de la phrase. On accorde trois points dans le premier cas et un seul dans le deuxième. Le total des points obtenu est divisé par le nombre de triplets présenté à l'élève.

Tableau 37 — Variables décrivant le volume de l'activité et la qualité du travail d'un élève

14.3. Résultats

Le modèle M9 (tableau 38) expliquant le gain de compréhension en lecture est tout de suite significatif, et explique environ la moitié de la variance. Nous l'affinerons en éliminant les individus dont la différence entre la note attendue et la performance réelle excède un écart-type et demi des résidus. L'élimination de 7 individus dans le modèle M10 provoque un gain explicatif de près de 13 points.

	N	Corrélation	Pourcentage de variance expliquée	p
M9 : modèle avec tous les élèves	48	0,68	46 %	<0,001
M10 : modèle avec élimination des élèves à fort résidu	41	0,77	59 %	<0,001

Tableau 38 — Résultats généraux de la régression cherchant à expliquer le gain en compréhension par le type d'entraînement. Modèle M9 et M10.

Le tableau 39 présente les contributions et significativités des différentes variables introduites dans les modèles.

Le premier résultat à observer, c'est que moins on a été performant au pré-test, plus on a de gain en compréhension. Quel qu'ait été le type d'entraînement suivi, les résultats montrent qu'il provoque une amélioration plus grande de la compréhension chez les élèves qui, au départ, étaient les plus faibles⁷¹. On voit également que le sexe n'est pas sans influence sur le gain de compréhension : les filles présentent un gain plus important.

⁷¹ Ce résultat était déjà présent dans l'analyse précédente, où nous comparions le groupe expérimental dans sa totalité au groupe témoin.

		Modèle M10		
		Coefficients bêta	Paramètres	Significativité
Sexe	femme	0,34	0,24	<0,012
	homme	-0,34	-0,24	<0,012
Type d'épreuve au pré-test				
	Textes courts	0,11	0,08	<0,38
	Texte long	-0,11	-0,08	<0,38
Résultats au pré-test		-0,65	-0,46	<0,001
Exposition Exercice Clique		0,095	0,0004	<0,46
Exposition Exercice Trou		-0,17	-0,001	<0,21
Exposition Exercice Triplets		0,29	0,009	<0,03
Réussite Exercice Clique		-0,02	-0,11	<0,88
Réussite Exercice Trou		0,03	0,31	<0,86
Réussite Exercice Triplets		0,18	0,4	<0,19
Constante			-2,06	<0,22

Tableau 39 — Modèle M10 visant à expliquer la compréhension en lecture.

Indépendamment de ces deux premiers résultats, la quantité d'exposition à l'exercice présentant des triplets de phrase a des effets significatifs sur le gain en compréhension des élèves. Plus les élèves ont travaillé sur cet exercice, plus importante a été leur amélioration dans la compréhension de la lecture des textes. Cet effet est avéré quel qu'ait été leur niveau de départ. Par contre, on ne peut pas dire qu'une exposition supplémentaire aux deux autres exercices ait entraîné un gain de compréhension.

Du côté des réussites dans chacun des exercices, on ne note aucun rapport significatif avec un gain de compréhension.

14.4. Discussion

La tentative de repérer un exercice ayant une contribution significative des différents types d'entraînement fait ressortir le point suivant :

On observe un gain significatif quand on travaille davantage sur l'exercice des triplets. Pour autant, il ne faudrait pas en conclure que ce seul exercice suffirait. Nous avons construit notre plan d'entraînement en essayant de garder actif à la fois la notion de cadre explicatif et d'éléments à percevoir. Il nous semble que l'ensemble des exercices proposés correspond à cet objectif que nous avons emprunté aux travaux sur l'apprentissage perceptif (cf. chapitre 3). L'exercice de triplets n'a de sens qu'à l'intérieur d'un plan qui entraîne les différents aspects qui président à la construction rapide de la syntaxe pendant la lecture. Cependant, cet exercice est sans doute le plus complet, puisqu'il travaille aussi bien sur la qualité de la discrimination parafovéale des éléments, que sur la décision qui doit être faite de la structure qui construisait la phrase observée très rapidement. Dans un sens, cet exercice permet de mettre en place les mécanismes décrits par Berthoz sur la perception, quand il rappelle qu'elle est bien projection d'une pré-perception, que percevoir n'est que vérifier une catégorie d'hypothèses que l'expérience du monde permet de construire (Berthoz, 1997; Berthoz, 2003). En conséquence, les deux premiers exercices permettent un développement de l'expérience explicite des mots fonctionnels et des structures qui y sont associées, tandis que le troisième fait fonctionner la création d'hypothèses sur la structure de la langue écrite à partir desquelles, comme le dit Berthoz, *le cerveau choisit dans le monde ce qui lui convient*. Un peu plus de temps à choisir qu'à connaître donne visiblement de meilleurs résultats.

Conclusion

Nous annonçons en préambule que le modèle structural de lecture énoncé par Koriat et Greenberg allait servir de fil conducteur à ce travail. Les différents résultats que nous apportons dans cette étude confortent incontestablement la validité de ce modèle explicatif et permettent d'envisager de nouvelles approches quant à la place de la syntaxe dans l'enseignement de la lecture.

Rappelons ici une dernière fois que Koriat et Greenberg envisagent la lecture comme une activité où la construction de la structure de la phrase guide l'intégration des unités sémantiques. Le lecteur, à travers le repérage des morphèmes fonctionnels, élabore un cadre structurel dans lequel les informations porteuses de sens viennent se « nicher ». Une fois que les mots fonctionnels ont joué ce rôle, ils sont rejetés à l'arrière plan, ce qui explique les résultats des différentes expérimentations observant l'oubli sélectif de lettres pendant la lecture.

En premier lieu, nous nous sommes attaché au modèle en lui-même pour essayer de préciser son ampleur et son fonctionnement.

Son ampleur d'abord, puisque la question de la taille des structures construites par le lecteur est centrale dans ce modèle. La majorité des travaux avait montré jusqu'ici que le lecteur est capable de construire des structures correspondant aux syntagmes nominaux ou prépositionnels. Toutefois, les théories linguistiques de la syntaxe considèrent que les syntagmes, qu'ils soient nominaux ou flexionnels, fonctionnent selon des principes similaires. La question se posait alors de savoir si cette unicité de principes se retrouvait, au niveau du lecteur, dans une unicité de traitements ; en d'autres termes, le lecteur construit-il également des cadres sur des structures aussi longues que la phrase ? Les expériences que nous rapportons dans ce travail confirment, à l'aide d'un protocole complètement différent, les premières réponses apportées par Musseler (Musseler et al., 2000). Nous montrons que les lecteurs expérimentés ont un taux d'oubli similaire dans les mots fonctionnels en position de complémentateur (les conjonctions de coordination et de subordination et les pronoms relatifs), c'est-à-dire de tête de projections flexionnelles (Pollock, 1997) et dans ceux en position de tête de syntagmes

nominaux (les déterminants). Pour le modèle structural, ce résultat indique que les complémenteurs (dont le rôle est d'introduire une projection maximale d'une tête flexionnelle, c'est-à-dire une proposition) sont également rejetés à l'arrière-plan après avoir servi à construire le cadre de la structure syntaxique qu'ils commandent.

Son fonctionnement ensuite, puisque la question n'est pas close de savoir comment s'établissent ces cadres structuraux dans des situations normales de lecture. Certains travaux antérieurs ont montré que l'oubli de lettres dans les mots qui ont pour fonction de construire la syntaxe s'observait même lorsque la vision parafovéale était empêchée (par exemple par des techniques d'auto-présentation segmentée mot à mot). Cependant, la littérature fait état de l'importance des zones périphériques de la rétine dans la vision en général et dans celle de l'écrit en particulier ; plus précisément, de son utilité pour l'identification des mots ayant un rôle fonctionnel (et qui ont également une fréquence élevée) dans la phrase. Aussi, avons-nous mis au point une expérimentation qui se focalise sur le rôle de la vision périphérique dans la reconnaissance des structures syntaxiques. Les résultats que nous présentons montrent que le lecteur est en mesure d'extraire le squelette de la phrase à partir des informations perçues par les zones périphériques de la rétine, hors de toute conscience. Ce résultat suggère que les cadres syntaxiques créés par le lecteur résultent de l'interaction entre la perception parafovéale de la phrase qu'il parcourt et sa connaissance implicite des règles syntaxiques.

En second lieu, nous nous sommes attaché à mettre en relation l'habileté dans la construction de squelettes syntaxiques et la célérité ou la compréhension en lecture.

Les résultats que nous rapportons montrent que meilleure est la compréhension en lecture, plus grand est le rapport d'oubli de lettres entre les mots ayant un rôle syntaxique et ceux ayant un rôle sémantique. Parallèlement, nous montrons qu'un pourcentage d'oubli plus important dans les mots ayant un rôle fonctionnel est très significativement lié à une forte compréhension en lecture. A partir de ces deux résultats, nous pouvons affirmer que le meilleur compreneur est celui qui, à vitesse égale, anticipe avec le plus de créativité la

construction de la phrase. Bien entendu, la compréhension n'est pas uniquement tributaire d'une bonne anticipation des cadres syntaxiques : nos modèles explicatifs lui accordent un pouvoir explicatif d'environ 20%. Mais ce résultat est tout à fait cohérent avec l'extrême complexité des processus de compréhension qui intègrent des éléments nombreux : « les signifiants disponibles (lexiques) ; les signifiés (base de connaissances) ; les procédures ; les capacités de contrôle, etc. » (Fayol, 1992).

Ainsi, le modèle de Koriat semble être suffisamment robuste pour, d'une part, s'adapter aux structures syntaxiques de tailles variables allant du syntagme nominal au syntagme flexionnel et, d'autre part, être étroitement associé à la compréhension du lecteur. La combinaison de ces deux résultats nous a encouragé à étudier son développement chez l'apprenti lecteur :

Dans un premier temps, nous avons observé l'évolution de l'oubli de lettres chez des élèves des trois niveaux constituant le cycle 2. Il apparaît que, dès que les élèves commencent à maîtriser les techniques de lecture, donc dès la fin du CP, on observe, toutes choses égales par ailleurs (et en particulier l'effet de fréquence du mot), que la distribution des oublis cesse d'être aléatoire, et que la différence entre les mots à rôle syntaxique et les mots à rôle sémantique devient significative. Cette diminution très forte du pourcentage d'oubli dans les mots à rôle sémantique entre les élèves de GS et ceux de CP témoigne de l'habileté que construit le jeune lecteur à établir des cadres syntaxiques qu'il va garnir d'informations sémantiques.

Dans un deuxième temps, nous avons élaboré, pour des élèves de 5^{ème}, un programme didactique en nous aidant des écrits de Léontiev sur l'apprentissage. Nous avons cherché à « exposer à la conscience » des élèves (Léontiev, 1984), deux éléments qui permettent la construction de l'architecture de la phrase, et qui restent, dans une pratique normale de lecture, inaccessibles à la conscience : la vision parafovéale et l'utilisation des mots fonctionnels qui cristallisent la syntaxe de la phrase (Pinker, 1999). Ce que nous voulions entraîner chez les élèves (le *but*, au sens de Léontiev, de l'entraînement), ce n'est pas directement la compréhension, c'est l'habileté dans la construction des squelettes syntaxiques. Nous avons tenté d'amener à

la conscience des élèves les éléments sur lesquels s'appuie cette habileté, car selon Léontiev, c'est bien « la place qu'il [le but] occupe dans la structure de l'activité » qui importe pour sa conscientisation et de là pour son possible entraînement.

Les résultats que nous avons obtenus sont encourageants. En effet, après un trimestre d'entraînement à raison d'une heure par semaine soustraite de l'horaire ordinaire de la classe de français, les élèves du groupe expérimental ont vu leur compréhension s'améliorer significativement par rapport aux élèves du groupe témoin. Il est intéressant d'observer que, seule, la compréhension (et non la vitesse) s'est élevée : ces résultats sont cohérents avec ce que nous observions chez les lecteurs adultes où la compréhension (et non la vitesse de lecture) était associée à une meilleure habileté dans l'établissement des cadres syntaxiques. Il nous semble ainsi avoir mis à jour une relation forte entre compréhension et habileté structurale de lecture. Tout se passe comme si cette habileté à mettre en place le squelette syntaxique de la phrase libérait l'attention sur les seuls éléments riches de sens qui s'intègrent à l'intérieur du cadre. A l'inverse, les lecteurs ayant une habileté moindre dans cette construction doivent mener de front, dans leurs rencontres avec les différents mots de la phrase, la construction de la syntaxe et celle du sens. L'amélioration de la compréhension n'est qu'une conséquence de cette meilleure habileté.

Si les objectifs que nous nous étions fixés au début de ce travail semblent remplis, il n'en reste pas moins que nous avons rencontré, au cours de ce travail, de nouvelles interrogations.

La principale se situe assurément au niveau de l'approfondissement du modèle structural grâce au phénomène d'oubli de lettres. La théorie syntaxique que nous avons rappelée, si elle accorde une place toute particulière aux mots fonctionnels comme têtes de syntagmes (les déterminants pour les syntagmes nominaux, les prépositions pour les syntagmes prépositionnels, les complémenteurs pour les syntagmes flexionnels,...), indique également qu'un syntagme flexionnel (une proposition) n'est, somme toute, qu'un syntagme comme un autre. Sa tête est la *flexion*, et elle prend pour complément le

syntagme verbal ; la flexion (qui marque le temps, les accords et le mode) peut être soit un verbe auxiliaire (*Le maçon a réparé le mur*), soit une simple marque de temps et de personne sur le verbe (*Le maçon réparera le mur*). Le modèle structural de Koriat invite à faire l'hypothèse que les lettres présentes dans les morphèmes qui organisent le syntagme flexionnel devraient être plus souvent oubliées que dans des noms, des adjectifs ou d'autres parties des verbes. Koriat et Greenberg ont commencé leurs travaux sur le modèle structural en observant l'oubli de lettres dans les morphèmes ayant un rôle dans l'organisation de la syntaxe et qui sont intégrés à l'intérieur des noms⁷² (Koriat et al., 1991) : il ne serait donc pas totalement nouveau d'observer des oublis de lettres dans des morphèmes intégrés à l'intérieur des mots. Ce qui serait inédit, c'est l'observation d'un oubli important dans les morphèmes flexionnels, qu'ils aient une réalisation autonome (sous la forme d'auxiliaire) ou qu'ils soient intégrés à l'intérieur du verbe conjugué. Si cette observation se confirmait, le modèle de lecture structural en serait renforcé dans la mesure où le lecteur ne serait plus seulement sensible au rôle des mots fonctionnels mais aussi à d'autres indices linguistiques servant à organiser la structure de la phrase.

Une autre question découle de la distance à laquelle semblent encore perceptibles les variations syntaxiques. Dans notre expérimentation, à 27 signes à droite (à peu près cinq mots), l'effet s'observe toujours. Plus que la recherche d'une 'limite', ce qui est important, c'est que, à cette distance, les informations sont totalement floues. Que voit-on sous un tel angle ? Sûrement pas des lettres. Et comment apprend-on à travailler avec ce type d'informations si les démarches pédagogiques ne font travailler que sur des mots ou des syllabes, bref à ce qui peut être perçu en vision fovéale ? Ce rôle de la vision parafovéale fait penser aux travaux sur la lecture des lecteurs experts en braille. Leurs deux mains sont rapidement dissociées : la gauche explore la ligne suivante de celle où se déplace la main droite. Ces lecteurs sont capables de lire simultanément deux plages de textes différents, sans en avoir explicitement conscience (Portalier, 2001). On retrouve là les descriptions d'Alain Berthoz sur le caractère pro-actif du cerveau dans la prise

⁷² voir chapitre 1.

d'information et dans les mouvements oculaires (Berthoz & Petit, 1996) (Berthoz, 1997). Dès lors, il est vraisemblablement nécessaire de questionner les expérimentations majoritaires en lecture (essentiellement pour des raisons techniques) qui travaillent sur l'identification séquentielle de mots isolés, la lecture étant le résultat de l'itération de cette opération. N'est-ce pas comme si on étudiait la lecture des aveugles simplement en regardant ce que fait leur main droite ? D'une manière parallèle, ne faut-il pas intégrer, dans les enseignements, cette habileté dans le traitement simultané d'informations provenant de sources différentes ? Dans notre protocole didactique, l'exercice qui semble apporter le plus de bénéfices aux élèves est celui qui les conduit à utiliser la totalité de l'information visuelle disponible, en synergie avec leur connaissance du fonctionnement syntaxique de la langue.

Pour terminer, nous voudrions revenir sur le sort habituellement fait à la syntaxe dans les pratiques scolaires.

Sans doute parce qu'elle est considérée comme une sous-partie de la grammaire, spécialisée dans l'ordonnement des mots en syntagmes et en phrases, la syntaxe est, le plus souvent, abordée lors d'activités grammaticales prescriptives ou stylistiques qui sont, avant tout, des moments de description de l'écrit ou de correction des phrases produites par les élèves. Dans ces conditions, le but immédiat de l'action didactique est de travailler la production de l'écrit plutôt que sa réception. Pourtant, les derniers programmes de français, pour l'école primaire comme pour le collège, commencent à accorder une place à l'étude de la syntaxe dans une optique de réception, mais l'apparente complexité de la théorie syntaxique sert parfois de repoussoir aux professeurs de français. Les résultats issus du présent travail confirment l'utilité de compléter l'approche traditionnelle de la syntaxe, afin de participer à l'élévation du niveau de lecture des élèves, préoccupation dont l'urgence sociale se fait chaque jour plus présente.

Bibliographie

- Abernethy, B. (1991). Visual search strategies and decision-making in sport. Special issue : Information processing and decision making in sport. *International Journal of Sport Psychology*, 22, 189-210.
- Abernethy, B., Neil, R. J., & Koning, P. (1994). Visual-perceptual and cognitive differences between expert, intermediate and novice snooker players. *Applied Cognitive Psychology*, 8, 185-211.
- Abrahamsen, E. P. & Shelton, K. C. (1989). Reading comprehension in adolescents with learning disabilities : semantic and syntactic effects. *J.Learn.Disabil.*, 22, 569-572.
- Adams, M. (1979). Models of word recognition. *Cognitive Psychology*, 11, 133-179.
- Altman, G. & Steedman, M. (1988). Interaction with context during human sentences processign. *Cognition*, 30, 191-238.
- Andreewsky, E. (1980). Pathologie alexique. Les troubles de la lecture survenant chez les lecteurs confirmés. In *Cinq contributions pour comprendre la lecture* (pp. 73-83). Paris : AFL.
- Andreewsky, E. & Seron, X. (1975). Implicit processing of grammatical rules in a classical case of agrammatism. *Cortex*, 9, 379-390.
- Asmussen, C. & Bastien-Toniazzo, M. (2000). Comprendre des phrases complexes en lecture quand on est débutant. *Psychologie Française*, 45, 245-252.
- Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interation of contextual constraint and parafoveal visual information during reading. *Cognitive Psychology*, 17, 364-390.
- Bavalier, D., Corina, D., Padmanabhan, S., Clark, V. P., Karni, A., Prinster, A. et al. (1997). Sentence reading : a functional MRI study at 4 Tesla. *Journal of Cognitive Neuroscience*, 9, 664-686.

- Bentin, S. (1987). Visual word perception and semantic processing : an electrophysiological perspective. *Isr.J.Med.Sci.*, 23, 138-144.
- Bentin, S., Deutsch, A., & Liberman, I. Y. (1990). Syntactic competence and reading ability in children. *J.Exp.Child Psychol.*, 49, 147-172.
- Berthoz, A. (1997). *Le sens du mouvement*. Paris : Odile Jacob.
- Berthoz, A. (2003). *La décision*. Paris : Odile Jacob.
- Berthoz, A. & Petit, L. (1996). Les mouvements du regard : une affaire de saccades. *La Recherche*, 289, 58-65.
- Besse, H. & Porquier, R. (1991). *Grammaire et didactique des langues*. Paris : Didier.
- Biederman, I. & Shiffrar, M. S. (1987). Sexing day-old chicks : A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 13, 640-645.
- Bowers, J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological, and articulatory contributions to masked letter and word priming. *J.Exp.Psychol.Hum.Percept.Perform.*, 24, 1705-1719.
- Bronckart, JP. (1977). *Théories du langage. Une introduction critique*. Bruxelles : Mardaga.
- Bronckart, JP., Bain, D., Schneuwly, B., Davaud, C., & Pasquier, A. (1985). *Le fonctionnement des discours*. Lausanne : Delachaux, Niestlé.
- Calef, T., Pieper, M., & Coffey, B. (1999). Comparisons of eye movements before and after a speed-reading course. *J.Am.Optom.Assoc.*, 70, 171-181.
- Caplan, D. & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behav.Brain Sci.*, 22, 77-94.
- Caroll, L. (1971). *De l'autre côté du miroir, et ce que Alice y trouva*. Paris : Aubier-Flammarion.

- Carpenter, P. A. & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading : perceptual and language processes*. (pp. 275-307). New-York.
- Castiello, U. & Umilta, C. (1992a). Orienting attention in volleyball players. *International Journal of Sport Psychology*, 23, 301-310.
- Castiello, U. & Umilta, C. (1992b). Splitting focal attention. *J.Exp.Psychol.Hum.Percept.Perform.*, 18, 837-848.
- Changeux, JP. (2002). *L'homme de vérité*. Paris : Odile Jacob.
- Chase, W. G. & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 5, 55-81.
- Chiss, J. L., Filliolet, J., & Mainguenaud, D. (1998). *Linguistique Française : Communication - Syntaxe - Poétique*. Paris : Hachette.
- Chomsky, N. (1969). *Structures Syntaxiques*. Paris : Seuil.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A. et al. (2000). The visual word form area : spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123 (Pt 2), 291-307.
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125, 1054-1069.
- Colmant, M. & Mulliez, A. (2003). *Les élèves de CMI. Premiers résultats d'une évaluation internationale en lecture (PIRLS)* (Rep. No. Note d'information 03.22). Direction de la Programmation et du Développement. Ministère de la Jeunesse de l'Éducation nationale et de la Recherche.
- Corcoran, D. W. (1966). An acoustic factor in letter cancellation. *Nature*, 210, 658.
- Crain, S. & Steedman, M. (1985). On not being led up the garden path : The use of context by the psychological syntax processor. In D.R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing : Psychological*,

computational and theoretical perspective. (Cambridge : Cambridge University Press.

- Cupples, L. & Holmes, V. M. (1992). Evidence for a difference in syntactic knowledge between skilled and less skilled adult readers. *J.Psycholinguist.Res.*, 21, 249-274.
- Daneman, M. & Carpenter, P. A. (1980). Individual difference in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Dapretto, M. & Bookheimer, S. Y. (1999). Form and content : dissociating syntax and semantics in sentence comprehension. *Neuron*, 24, 427-432.
- Dauphin, L., Rebmeister, B., & Zelty, C. (2003a). *L'évaluation des compétences des élèves de cinquième en septembre 2002.* (Rep. No. Note d'information 03.21). Direction de la Programmation et du Développement. Ministère de la Jeunesse de l'Éducation nationale et de la Recherche.
- Dauphin, L., Rebmeister, B., & Zelty, C. (2003b). *L'évaluation des compétences des élèves de sixième en septembre 2002.* (Rep. No. Note d'information 03.20). Direction de la Programmation et du Développement. Ministère de la Jeunesse de l'Éducation nationale et de la Recherche.
- Dehaene, S., Le Clec'H, G., Poline, J. B., Le Bihan, D., & Cohen, L. (2002). The visual word form area : a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13, 321-325.
- Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G. et al. (1998). Imaging unconscious semantic priming. *Nature*, 395, 597-600.
- Dopkins, S., Morris, R. K., & Rayner, K. (1992). Lexical ambiguity and eye fixations in reading : a test of competing models of lexical ambiguity resolution. *Journal of Memory and Language*, 31, 461-476.
- Draine, S. C. & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology : General*, 127, 286-303.

- Drewnowsky, A. (1978). Detection errors on the word *the* : Evidence for the acquisition of reading levels. *Memory & Cognition*, 6, 403-409.
- Drewnowsky, A. & Healy, A. F. (1977). Detection errors on *the* and *and* : Evidence for reading units larger than word. *Memory & Cognition*, 5, 636-647.
- Drewnowsky, A. & Healy, A. F. (1982). Phonetic factors in letter detection : a reevaluation. *Memory & Cognition*, 10, 145-154.
- Ehrlich, S. F. & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20, 641-655.
- Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Res.*, 35, 3003-3013.
- Fayol, M. (1992). Comprendre ce qu'on lit : de l'automatisme au contrôle. In M.Fayol, J. E. Gombert, P. Lecocq, L. Sprenger-Charolles, & D. Zagar (Eds.), *Psychologie cognitive de la lecture*. (pp. 73-106). Paris : PUF.
- Ferrand, L., Grainger, J., & Segui, J. (1994). A study of masked form priming in picture and word naming. *Memory & Cognition*, 22, 431-441.
- Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- Fodor, J. (1986). *La modularité de l'esprit. (Modularity of mind)*. Paris : Minuit.
- Foucambert, D. (1997). *Conscience graphique et performance en lecture : Étude statistique sur un échantillon d'enfants de onze ans*. Mémoire de DEA Université de Caen.
- Foucambert, D. (2000). Les effets d'une année d'entraînement à la lecture avec un logiciel éducatif : résultats en classe de sixième de collège. *Revue Française de Pédagogie*, 133, 63-73.
- Fowler, A. E. (1988). Grammaticality, judgement and reading skill in grad 2. *Annals of Dyslexia*, 38, 73-94.

- Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension : Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing : early and late event-related brain potential effects. *J.Exp.Psychol.Learn.Mem.Cogn*, 22, 1219-1248.
- Friederici, A. D., Opitz, B., & von Cramon, D. Y. (2000). Segregating semantic and syntactic aspects of processing in the human brain : an fMRI investigation of different word types. *Cereb.Cortex*, 10, 698-705.
- Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension : localizing syntactic and semantic processes. *Cereb.Cortex*, 13, 170-177.
- Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration : sequential effects of syntactic and semantic information. *Mem.Cognit.*, 27, 438-453.
- Frith, U. (1986). A developmental framework for developmental dyslexia, , 69-81, 1986. *Annals of Dyslexia*, 69-81.
- Gaux, C. & Gombert, J. E. (1999). Implicit and explicit syntactic knowledge and reading in pre-adolescent. *British Journal of Developmental Psychology*, 17, 169-188.
- Gobet, F. & Simon, H. A. (1998). Expert chess memory : revisiting the chunking hypothesis. *Memory*, 6, 225-255.
- Goldstone, R. L. (1998). Perceptual learning. *Annu.Rev.Psychol.*, 49, 585-612.
- Gombert, J. E. (1990). *Le développement métalinguistique*. Paris : PUF.
- Gombert, J. E. (1996). Les activités cognitives en œuvre dans la lecture et son acquisition. In *Illettrisme : Quels chemins vers l'écrit*. (Magnard ed., pp. 129-144).

- Gorrell, P. (1989). Establishing the loci of serial and parallel effects in syntactic processing. *Journal of Psycholinguistic Research*, 18, 61-73.
- Greenberg, S. N. & Koriat, A. (1991). The missing-letter effect for common function word depends on their linguistic function in the phrase. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 17, 1051-1061.
- Greenberg, S. N., Koriat, A., & Shapiro, A. (1992). The effects of syntactic structure on letter detection in adjacent function words. *Mem. Cognit.*, 20, 663-670.
- Greenberg, S. N., Koriat, A., & Vellutino, F. R. (1998). Age changes in the missing-letter effect reflect the reader's growing ability to extract the structure from text. *J.Exp.Child Psychol.*, 69, 175-198.
- Greenfield, P. M., deWinstanley, P., Kilpatrick, H., & Kaye, D (1994). Action video games and informal education : effects on strategies for dividing visual attention. *Journal of Applied Developmental Psychology*, 15, 105-123.
- Greenwood, C. R., Dinwiddie, G., Bailey, V., Carta, J. J., Dorsey, D., Kohler, F. W. et al. (1987). Field replication of classwide peer tutoring. *J.Appl.Behav.Anal.*, 20, 151-160.
- Greenwood, C. R., Dinwiddie, G., Terry, B., Wade, L., Stanley, S. O., Thibadeau, S. et al. (1984). Teacher- versus peer-mediated instruction : an ecobehavioral analysis of achievement outcomes. *J.Appl.Behav.Anal.*, 17, 521-538.
- Grevisse, M. & Goosse, A. (1980). *Nouvelle grammaire française*. Paris-Gembloux : Duculot.
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660-676.
- Haber, R. N. (1966). Nature of the effect of set on perception. *Psychol.Rev.*, 73, 335-351.
- Hadley, J. A. & Healy, A. F. (1991). When are reading units larger than letter? Refinement of unitization reading model. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 17, 1062-1073.

- Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Brain Res. Cogn Brain Res.*, *16*, 38-50.
- Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D. et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage.*, *17*, 1101-1116.
- Healy, A. F. (1976). Detection errors on the word the : Evidence for reading units larger than letters. *Journal of Experimental Psychology : Human Perception and Performance*, *2*, 235-242.
- Healy, A. F. (1980). Proofreading errors on the word *the* :New evidence on reading units. *Journal of Experimental Psychology : Human Perception and Performance*, *6*, 45-57.
- Healy, A. F. (1994). Letter detection : a window to unitization and other cognitive processes in reading texts. *Psychonomic Bulletin and Review*, *1*, 333-344.
- Healy, A. F., Conboy, G. L., & Drewnowsky, A. (1987). Characterizing the processing units in reading :effects of intra- and inter-word space in a letter detection task. In B.K.Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 279-297). Hillsdale, NJ : Erlbaum.
- Healy, A. F. & Drewnowsky, A. (1983). Investigating the boundaries of reading units : Letter detection in misspelled words. *Journal of Experimental Psychology : Human Perception and Performance*, *9*, 413-426.
- Healy, A. F., Oliver, W. L., & McNamara, T. P. (1987). Detecting letters in continuous text : Effect of display size. *Journal of Experimental Psychology : Human Perception and Performance*, *13*, 279-290.
- Hoen, M. & Dominey, PF. (2000). ERP analysis of cognitive process sequencing : a left anterior negativity related to structural transformation processing. *Neuroreport*, *11*.

- Hoen, M., Golembiowski, M., Guyot, E., Deprez, V., Caplan, D., & Dominey, P. F. (2003). Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *Neuroreport*, *14*, 495-499.
- Hyönä, J. (1995). Do irregular letter combinations attract readers' attention ? Evidence from fixation location in words. *Journal of Experimental Psychology : Human Perception and Performance*, *21*, 68-81.
- Hyönä, J. & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers : effects of word length and word frequency. *J.Exp.Psychol.Learn.Mem.Cogn*, *21*, 1430-1440.
- Inhoff, A. W. (1989). Lexical access during eye fixations in reading : are word access codes used to integrate lexical information accross interword fixations? **28** : 444-461, 1989. *Journal of Memory and Language*, *28*, 444-461.
- Inhoff, A. W. (1982). Parafoveal word perception : a further case against semantic preprocessing. *Journal of Experimental Psychology : Human Perception and Performance*, *8*, 137-145.
- Inhoff, A. W. & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading : effects of word frequency. *Perception and Psychophysics*, *40*, 431-439.
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension : individual differences in working memory. *Psychol.Rev.*, *99*, 122-149.
- Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension : new frontiers of evidence and arguments. *Psychol.Rev.*, *103*, 773-780.
- Kang, A. M., Constable, R. T., Gore, J. C., & Avrutin, S. (1999). An event-related fMRI study of implicit phrase-level syntactic and semantic processing. *Neuroimage.*, *10*, 555-561.
- Kennedy, A. (2000). Parafoveal processing in word recognition. *Q.J.Exp.Psychol.A*, *53*, 429-455.

- King, J. W. & Just, M. A. (1991). Individual differences in syntactic processing : The role of working memory. *Psychol.Rev.*, 99, 122-149.
- Koriat, A. & Greenberg, S. N. (1991). Syntactic control of letter detection : Evidence from English and Hebrew nonwords. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 17, 1033-1048.
- Koriat, A. & Greenberg, S. N. (1993). Prominence of leading functors in function morpheme sequences as evidenced by letter detection. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 19, 34-59.
- Koriat, A. & Greenberg, S. N. (1994). The extraction of phrase structure during reading : Evidence from letter detection errors. *Psychonomic Bulletin and Review*, 1, 345-356.
- Koriat, A. & Greenberg, S. N. (1996a). Syntactic Control of Letter Detection : Evidence From English and Hebrew NonWords. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 17, 1035-1050.
- Koriat, A. & Greenberg, S. N. (1996b). The Enhancement Effect in Letter Detection : Further Evidence for the Structural Model of Reading. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 22, 1184-1195.
- Koriat, A., Greenberg, S. N., & Goldshmid, Y. (1991). The missing-letter effect in Hebrew : Word frequency or word function ? *Journal of Experimental Psychology : Learning, Memory and Cognition*, 17, 66-80.
- Koriat, A., Greenberg, S. N., & Kreiner, H. (2002). The extraction of structure during reading : evidence from reading prosody. *Mem.Cognit.*, 30, 270-280.
- Layton, A., Robinson, J., & Lawson, M. (1998). The relationship between syntactic awareness and reading performance. *Journal of Research in Reading*, 21, 5-23.
- Le Ny, JF. (1989). *Science cognitive et compréhension du langage*. Paris : PUF.

- Lecocq, P., Casalis, S., Leuwers, C., & Watteau, N. (1996). *Apprentissage de la lecture et compréhension d'énoncés*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.
- Léontiev, A. (1984). *Activité, Conscience, Personnalité*. Moscou : Editions du Progrès.
- Lipka, S. (2002). Reading sentences with a late closure ambiguity : Does semantic information help? *Language and cognitive processes*, 17, 271-298.
- Luck, S. J., Vohel, E. K., & Shapiro, K. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383, 616-618.
- MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- Marmoz, L. (1988). *Les Sciences de l'Education en France, histoire et réalité*. Paris : E.A.P.
- Martinet, A. (1979). *Grammaire fonctionnelle du français*. Paris : Didier.
- Mazoyer, B. M., Dehaene, S., Tzourio, N., Frak, V., Murayama, N., Cohen, L. et al. (1997). The cortical representation of speech. *J.Cognitive Neurosci.*, 4, 467-479.
- McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception : I. An account of basic findings. *Psychological Review*, 88, 375-407.
- McConkie, G. W., Kerr, P. W., Reddix, M. D., & Zola, D. (1988). Eye movement control during reading : I. The location of initial eye fixations in words. *Vision Research*, 28, 1107-1118.
- McCormick, P. A., Klein, R. M., & Johnston, S. (1998). Splitting versus sharing focal attention : comment on Castiello and Umiltà (1992). *J.Exp.Psychol.Hum.Percept.Perform.*, 24, 350-357.

- Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions : an analysis with event-related potentials. *Mem.Cognit.*, 23, 477-494.
- Ministère de la Jeunesse de l'Éducation nationale et de la Recherche (2002a). *Document d'application des programmes*. (Direction de l'enseignement scolaire ed.) Paris : CNDP.
- Ministère de la Jeunesse de l'Éducation nationale et de la Recherche. (2002b). Programmes d'enseignement de l'école primaire. A. du 25-01-2002. Journal officiel du 10-02-2002.
Ref Type : Bill/Resolution
- Miyake, A., Carpenter, P. A., & Just, M. A. (1995). Working memory constraints on the resolution of lexical ambiguity :maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, 33, 175-202.
- Moravcsik, J. E. & Healy, A. F. (1995). The effect on meaning on letter detection. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 21, 82-95.
- Moravcsik, J. E. & Healy, A. F. (1998). Effects of syntactic role and syntactic prominence on letter detection. *Psychon.Bull.Rev.*, 5, 96-100.
- Morris, R. K., Rayner, K., & Pollatsek, A. (1990). Eye movement guidance in reading : The role of parafovea letter and space information. *Journal of Experimental Psychology : Human Perception and Performance*, 16, 501-509.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165-178.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *IEA's Study of Reading Literacy Achievement in Primary Schools* (Rep. No. PIRLS 2001 International Report). Chestnut Hill, MA : Boston College.

- Musseler, J., Koriat, A., & Nisslein, M. (2000). Letter-detection patterns in German : a window to the early extraction of sentential structure during reading. *Mem.Cognit.*, 28, 993-1003.
- Naccache, L. & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, 80, 215-229.
- Nazir, T. A., Jacobs, A. M., & O'Regan, J. K. (1998). Letter legibility and visual word recognition. *Mem.Cognit.*, 26, 810-821.
- Neville, H., Mills, D. L., & Lawson, D. S. (1992). Fractionating Language : Different Neural Subsystems With Different Sensitive Periods. *Cerebral Cortex*, 2, 244-258.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE(tm). *L'année psychologique*.
- Nougier, V., Ripoll, H., & Stein, J.-F. (1989). Orienting of attention with highly skilled athletes. *International Journal of Sport Psychology*, 20, 205-223.
- O'Regan, J. K. (1979). Saccade size control in reading : Evidence for the linguistic control hypothesis. *Perception and Psychophysics*, 25, 501-509.
- O'Regan, J. K. (1992). Facteurs sensoriels et moteurs dans la lecture : la position optimal du regard. In P.Lecocq (Ed.), *La lecture. Processus, apprentissage, troubles*. (Lille : Presses universitaires de Lille.
- O'Regan, J. K. & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition :a challenge to current theory. *Journal of Experimental Psychology : Human Perception and Performance*, 18, 185-197.
- Passeron, J. C. (1987). La notion de pacte. *Les actes de lecture*, 17, 55-59.
- Perea, M. & Gotor, A. (1996). Effects of masked repetition priming and orthographic neighborhood in visual recognition of words. *Percept.Mot.Skills*, 83, 179-186.

- Perea, M. & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition*, 62, 223-240.
- Philippeau, G. (1986). *Comment interpréter les résultats d'une analyse en composantes principales?* Paris : ITCF.
- Piatelli-Palmarini, M. (1982). *Théories du langage, Théories de l'apprentissage. Le débat entre Jean Piaget et Noam Chomsky.* Paris : Seuil.
- Pigott, H. E., Fantuzzo, J. W., & Clement, P. W. (1986). The effects of reciprocal peer tutoring and group contingencies on the academic performance of elementary school children. *J.Appl.Behav.Anal.*, 19, 93-98.
- Pinker, S. (1999). *L'instinct du langage.* Paris : Editions Odile Jacob.
- Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, 256, 1018-1021.
- Pollatsek, A. & Rayner, K. (1982). Eye movement control in reading : the role of word boundaries. *Journal of Experimental Psychology : Human Perception and Performance*, 8, 817-833.
- Pollock, J. Y. (1997). *Langage et Cognition.* Paris : PUF.
- Portalier, S. (2001). Réorganisation neuro-sensorielle : vivariance et/ou transfert intermodal. Etude sur le sujet pathologique dans le cas de l'héminégligence et dans celui de la déficience visuelle. *Arobase*, 5, 3-16.
- Posner, M. I., Abdullaev, Y. G., McCandliss, B. D., & Sereno, S. C. (1999). Anatomy, circuitry and plasticity of word reading. In J.Everatt (Ed.), *Reading and Dyslexia : Visual and Attentional Processes* (pp. 137-163). London : Routledge.
- Pugh, A. K., Bassou, L., Granié, M., & Morucci, J.-P. (1991). Binocular Vision and Reading Performance. In F.Satow & B. Gatherer (Eds.), *Literacy without Frontiers* (pp. 182-190). Edinburgh : United Kingdom Reading Association.

- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav. Brain Sci.*, 22, 341-365.
- Rayner, K. (1979). Eye guidance in reading : Fixation locations in words. *Perceptions*, 8, 21-30.
- Rayner, K. (1984). Visual selection in reading picture perception and visual search : a tutorial review. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance*, X (pp. 67-96). Hillsdale, NJ : Erlbaum.
- Rayner, K. (1998). Eye movements in reading and information processing : 20 years of research. *Psychological Bulletin*, 124, 372-422.
- Rayner, K. & Duffy, S. A. (1986). Lexical complexity and fixation times in reading : Effect of word frequency, verb complexity and lexical ambiguity. *Mem. Cognit.*, 14, 191-201.
- Rayner, K. & Fischer, M. H. (1996). Mindless reading revisited : eye movements during reading and scanning are different. *Percept. Psychophys.*, 58, 734-747.
- Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, 38, 829-837.
- Rayner, K. & Pollatsek, A. (1987). Eye movements in reading : a tutorial review. In M. Coltheart (Ed.), *Attention and performance XII : The psychology of reading*. (pp. 327-362). Hillsdale, NJ : Erlbaum.
- Rayner, K. & Pollatsek, A. (1989). *The psychology of Reading*. Englewood Cliffs, NJ : Prentice-Hall.
- Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading : A comparison of two types of models. *Journal of Experimental Psychology : Human Perception and Performance*, 22, 1188-1200.
- Rayner, K. & Well, A. D. (1996). Effect of contextual constraint on eye movements in reading : a further examination. *Psychon. Bull. Rev.*, 3, 504-509.

- Reichter, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 274-280.
- Robin, I. & Rocher, T. (2003). *Les évaluations en lecture dans le cadre de la journée d'appel de préparation à la défense. Année 2001-2002.* (Rep. No. Note d'information 03.12). Direction de la Programmation et du Développement. Ministère de la Jeunesse de l'Éducation nationale et de la Recherche.
- Rubin, S. S., Newhoff, M., Peach, R. K., & Shapiro, L. P. (1996). Electrophysiological indices of lexical processing : the effect of verb complexity and age. *J.Speech Lang Hear.Res.*, 38, 850-863.
- Saint-Aubin, J. & Klein, R. M. (2001). Influence of parafoveal processing on the missing-letter effect. *J.Exp.Psychol.Hum.Percept.Perform.*, 27, 318-334.
- Saint-Aubin, J. & Poirier, M. (1997). The influence of word function in the missing-letter effect : Further evidence from French. *Memory & Cognition*, 25, 666-676.
- Schuberth, R. E. & Eimas, P. D. (1977). Effects of context on the classification of words and nonwords. *Journal of Experimental Psychology : Human Perception and Performance*, 3, 27-36.
- Schustack, M. W., Ehrlich, S. F., & Rayner, K. (1987). The complexity of contextual facilitation in reading :Local and global factors. *Journal of Memory and Language*, 26, 322-340.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meaning of ambiguous words in context : some limitation of knowledge-based processing. *Cognitive Psychology*, 14, 489-537.
- Sereno, S. C. & Rayner, K. (1992). Fast priming during eye fixation in reading. *Journal of Experimental Psychology : Human Perception and Performance*, 18, 173-184.

- Shapiro, K. & Caramazza, A. (2003). The representation of grammatical categories in the brain. *Trends Cogn Sci.*, 7, 201-206.
- Shapiro, L. P. (1997). Tutorial : an introduction to syntax. *J.Speech Lang Hear.Res.*, 40, 254-272.
- Shapiro, L. P., Brookins, B., Gordon, B., & Nagel, H. N. (1991). Verb effects during sentence processing. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 17, 983-996.
- Shapiro, L. P., Zurif, E., & Grimshaw, J. (1987). Sentences processing and the mental représentation of verbs. *Cognition*, 27, 219-246.
- Shapiro, L. P., Zurif, E., & Grimshaw, J. (1989). Verb representation and sentence processing : contextual impenetrability. *Journal of Psycholinguistic Research*, 18, 223-243.
- Solotareff, G. (2001). *Contes d'été*. Paris : Ecole des Loisirs.
- Sprenger-Charolles, L. (1992). L'évolution des mécanismes d'identification des mots. In M.Fayol, J. E. Gombert, P. Lecocq, L. Sprenger-Charolles, & D. Zagar (Eds.), *Psychologie cognitive de la lecture*. (pp. 141-194). Paris : PUF.
- Starkes, J., Allard, F., Lindley, S., & O'Reilly, K. (1994). Abilities and skill in basketball. Special issue : expert-novice differences in sport. *International Journal of Sport Psychology*, 25, 265.
- Stuart, M. & Coltheart, M. (1988). Does reading develop in a sequence of stage ? *Cognition*, 30, 139-181.
- Taft, M. & Forster, KI. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15, 607-620.
- Tardieu, J. (1951). *La comédie du langage*. Paris : Gallimard.
- Townsend, J. T., Taylor, S. G., & Brown, D. R. (1971). Lateral masking for letters with unlimited viewing time. *Perception and Psychophysics*, 10, 375-378.

- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing : Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285-318.
- Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension : the relationships between reading skill, reading speed, and visual inspection. *Lang Speech*, 33 (Pt 1), 69-81.
- van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York : Academic Press.
- Vitu, F. (1991). The influence of parafoveal processing in linguistic context on the optimal landing position effect. *Percept.Psychophys.*, 50, 58-75.
- Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading : eye-movement characteristics are similar in scanning letter strings and reading texts. *Percept.Psychophys.*, 57, 352-364.
- Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception and Psychophysics*, 47, 583-600.
- Vugalic, S. (1996). *Les compétences en lecture, en calcul et en géométrie des élèves à l'entrée au CE2 et en sixième*. (Rep. No. Note d'information, 96.22.). Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche.
- Waters, G., Caplan, D., & Yampolsky, S. (2003). On-line syntactic processing under concurrent memory load. *Psychon.Bull.Rev.*, 10, 88-95.
- Zagar, D. (1992). L'approche cognitive de la lecture : de l'accès au lexique au calcul syntaxique. In M.Fayol, J. E. Gombert, P. Lecocq, L. Sprenger-Charolles, & D. Zagar (Eds.), *psychologie cognitive de la lecture* (pp. 15-72). Paris : PUF.

Index des figures

Figure 1 : Ligne du temps montrant différents processus pendant la lecture de mots.....	6
Figure 2 — Les différents domaines d'analyse selon l'hypothèse d'unités concurrentes et parallèles de Healy.....	12
Figure 3 — Arbre représentant la phrase (d)	36
Figure 4 — Arbre représentant la phrase (e)	37
Figure 5 — Représentation syntaxique de la phrase (f) avec l'interprétation (g)	38
Figure 6 — Représentation syntaxique de la phrase (f) avec l'interprétation (h)	38
Figure 7 — Schéma d'un N-Barre	41
Figure 8 — Schéma d'un N-Barre avec actant et modifieur	42
Figure 9 — V-Barre avec actant et modifieur	42
Figure 10 — Spécifieur dans un syntagme verbal	43
Figure 11 — Spécifieur dans un syntagme nominal	43
Figure 12 — Exemple de syntagme flexionnel.....	45
Figure 13 — Syntagme nominal représenté à l'aide de la notation N-Barre.....	46
Figure 14 — Représentation syntaxique de la phrase (q).....	47
Figure 15 — Schéma X-Barre pour tout type de syntagme.....	48
Figure 16 — Schéma de l'œil	57
Figure 17 — Densité relative des cônes (ligne continue) et des bâtonnets (ligne en tirets), et la fidélité d'identification d'un mot-cible à une distance variable du centre de la rétine (ligne pointillée).	57
Figure 18 — Etude du champ visuel. Les croix représentent le masquage des lettres	61
Figure 19 — Saccades oculaires sans perturbation du champ visuel	62
Figure 20 — Saccades oculaires après retrait du champ visuel gauche	62
Figure 21 — Saccades oculaires après retrait du champ visuel droit	62
Figure 22 — Positionnement des 3 modèles qui tentent de décrire le processeur syntaxique.	82
Figure 23 — Attachement non-minimal (7 nœuds) et minimal (6 nœuds) d'un syntagme prépositionnel ambigu	84
Figure 24 — Analyses d'une phrase selon le principe de clôture tardive	85
Figure 25 — Aires associatives principales et polygone du langage.....	97
Figure 26 — Régions activées lors de la lecture de phrases en anglais et de la lecture de pseudo-mots. Technique d'IRMF.	101
Figure 27 — Potentiels évoqués enregistrés durant la lecture de mots de différentes catégories.	103
Figure 28 — Aires cérébrales activées lors de traitements sémantiques et syntaxiques.....	105
Figure 29 — Facteur 1, 2 et 3 de l'ACP sur les épreuves de lecture.....	127
Figure 30 — Distribution de la vitesse de lecture.....	128
Figure 31 — Distribution de la compréhension ^U	129
Figure 32 — Distribution de la compréhension ^D	129
Figure 33 — Pourcentages d'oubli de lettres dans les mots en fonction du rôle du mot et du type de texte. Expérience 1.	132
Figure 34 — Pourcentages d'oublis de lettres dans les mots en fonction de leur nature grammaticale. Expérience 2.....	142
Figure 35 — Pourcentages (des moindres carrés) de réussite en fonction du rôle du mot introduisant la différence à l'intérieur des triplets.	160
Figure 36 — Pourcentage de réponses aux triplets échoués selon l'apparence visuelle ou la structure identique.	162

Figure 37 : Pourcentages moyens d'oubli de lettres en fonction du rôle du mot et du niveau scolaire	182
Figure 38 — Pourcentage d'oublis de lettres en fonction du rôle des mots et de l'âge des sujets.	189
Figure 39 — Exemple de triplet. Le mot initiant la différenciation a un rôle syntaxique	195
Figure 40 — Déroulement de l'exercice triplet	196
Figure 41 — Notes en français à l'évaluation nationale de 5 ^{ème} . Moyennes, erreurs-type et écarts-type pour le groupe expérimental et le groupe témoin.	202
Figure 42 — Différence de compréhension entre le groupe témoin et le groupe expérimental après un trimestre d'entraînement.....	205

Index des tableaux

Tableau 1 — Pourcentage d'omission de lettres en fonction du type de mot et du type de fixation.....	24
Tableau 2 — Pourcentages de mots sautés en fonction de leur taille.....	66
Tableau 3 — Catégories de phrases présentées dans l'expérience de Taraban et MacClelland (1988, 1990).....	90
Tableau 4 — Statistiques descriptives de la compréhension. Epreuves de lecture n°1.....	122
Tableau 5 — Statistiques descriptives de la vitesse de lecture. Epreuves de lecture n°1.....	122
Tableau 6 — Statistiques descriptives de la vitesse de lecture. Epreuves de lecture n°2.....	123
Tableau 7 — Statistiques descriptives de la compréhension. Epreuves de lecture n°2.....	124
Tableau 8 — Récapitulatif des variables issues des mêmes épreuves selon les âges.....	124
Tableau 9 — Résultats de l'ACP : Variance expliquée et valeurs propres de chacun des axes.....	126
Tableau 10 — Résultats généraux de la régression cherchant à expliquer le pourcentage d'oubli de lettres. Modèles M1 et M2. Expérience 2.....	138
Tableau 11 — Analyse de régression multiple : modèle M2. Paramètres des variables ou modalités présentant un bêta significatif à 0,001. Expérience 2.....	138
Tableau 12 — Analyse de covariance. Résultats univariés de chaque variable explicative. Les contributions significatives sont en gras.....	139
Tableau 13 — Moyennes des moindres carrés des pourcentages d'oublis de lettres dans les mots. Expérience 2.....	141
Tableau 14 — Groupes homogènes dans la nature des mots. Tests Post Hoc sur la nature des mots.....	143
Tableau 15 — Résultats généraux de la régression cherchant à expliquer l'oubli de lettres par les performances des lecteurs. Modèle M3 et M4. Expérience 3.....	149
Tableau 16 — Paramètres des variables pour le modèle M4. Expérience 3.....	149
Tableau 17 — Résultats généraux de la régression cherchant à expliquer les résultats en lecture par l'oubli de lettres. Modèle M5. Expérience 3.....	150
Tableau 18 — Paramètres, coefficients bêta et significativité des variables explicatives du modèle M5.....	151
Tableau 19 : Equivalences entre le nombre de signes, l'angle moyen d'écartement et le nombre moyen de mots.....	156
Tableau 20 — Exercice triplet : Pourcentage de réussite en fonction du rôle du mot, de la distance de la variation et de la fréquence du mot fixé.....	158
Tableau 21 — Tests Univariés de Significativité du Pourcentage de réussite.....	159
Tableau 22 — Résultats de la comparaison des catégories de réponse en cas de mauvaise réponse.....	162
Tableau 23 — Nombre d'occurrences sur le programme des cycles 2 et 3 de l'école élémentaire.....	172
Tableau 24 — Nombre d'occurrences sur le programme de cycles 2 et 3 de l'école élémentaire.....	174
Tableau 25 — Pourcentages et Erreurs-Type des oublis de lettres en fonction du niveau scolaire, de la prononciation de la lettre et du rôle du mot.....	180
Tableau 26 — Résultats généraux des analyses de régression multiple cherchant à expliquer l'oubli de lettre en fonction du niveau scolaire. Modèle M6, M7 et M8.....	183
Tableau 27 — Modèle explicatif des scores d'oublis de lettres pour la GS (M6).....	184
Tableau 28 — Modèle explicatif des scores d'oublis de lettres pour le CP (M7).....	184
Tableau 29 — Modèle explicatif des scores d'oublis de lettres pour le CE1 (M8).....	184

Tableau 30 — Répartition du sexe, de l'âge et de la note en mathématique suivant les deux groupes.....	201
Tableau 31 — Répartition des individus en fonction des épreuves.....	203
Tableau 32 — Résultats de l'ANCOVA expliquant les compréhensions..	205
Tableau 33 — Compréhension en fonction du sexe et de l'épreuve.....	206
Tableau 34 — Résultats de l'ANCOVA expliquant les vitesses de lecture..	206
Tableau 35 — Vitesse de lecture en fonction du sexe et de l'épreuve.....	207
Tableau 36 — Répartition des exercices dans une séance d'entraînement de 20 minutes.....	211
Tableau 37 — Variables décrivant le volume de l'activité et la qualité du travail d'un élève.....	213
Tableau 38 — Résultats généraux de la régression cherchant à expliquer le gain en compréhension par le type d'entraînement. Modèle M9 et M10.....	214
Tableau 39 — Modèle M10 visant à expliquer la compréhension en lecture.	215

Index des auteurs

- Abdullaev, 6
Abernethy, 71
Abrahamsen, 117
Adams, 60
Allard, 71
Altman, 88
Andreewsky, 74, 99
Asmussen, 188
Avrutin, 109
Bain, 144
Balota, 64
Bassou, 68
Bastien-Toniazzo,
188
Bavalier, 101
Bentin, 106, 117,
169
Berthoz, 7, 59, 63,
64, 77, 110, 114,
217, 223
Besse, 194
Biederman, 72, 200
Bienkowski, 16
Bookheimer, 104
Bowers, 74
Bronckart, 30, 32,
52, 95, 144
Brookins, 89
Brown, 26, 108
Calef, 125
Caplan, 93, 94
Caramazza, 76
Caroll, 19
Carpenter, 13, 93,
94, 152
Casalis, 153
Castiello, 71
Changeux, 32, 99,
105, 110
Chase, 70
Chiss, 35, 36, 37
Chomsky, 7, 29, 30,
31, 32, 33, 34, 40,
52
Clement, 209
Clifton, 86
Cohen, 68
Colmant, 4
Coltheart, 60
Conboy, 18
Constable, 109
Corcoran, 11
Crain, 87
Cramon, 109
Cupples, 117, 168
Daneman, 93, 152
Dapretto, 104
Dauphin, 4
Davaud, 144
Dehaene, 68, 75,
116, 164, 199
Deutsch, 117
deWinstanley, 71
Dominey, 110
Dopkins, 16
Draine, 74
Drewnowsky, 11,
12, 18, 22, 130
Duffy, 64
Edelman, 73
Ehrlich, 64
Eimas, 75
Fahle, 73
Fantuzzo, 209
Fayol, 127, 221
Ferrand, 74, 137
Ferreira, 86
Fiebach, 108
Filliolet, 35
Fischer, 64, 66, 181
Fodor, 79
Forster, 59
Foucambert, 124
Fowler, 171
Frazier, 82, 83, 86
Friederici, 82, 107,
109
Frisch, 107
Frith, 60
Garnsey, 90
Gaux, 117, 169, 171
Gobet, 70
Goldshmid, 15
Goldstone, 72, 200,
209

Gombert, 60, 66,
 117, 169, 170,
 171, 176
 Goosse, 18
 Gordon, 89
 Gore, 109
 Gorrell, 92
 Gotor, 74
 Grainger, 74
 Granié, 68
 Greenberg, 7, 15,
 16, 17, 18, 19, 20,
 23, 24, 25, 27, 52,
 115, 116, 140,
 144, 147, 153,
 154, 179, 186,
 187, 219, 223
 Greenfield, 71
 Greenwald, 74
 Greenwood, 209
 Grevisse, 18
 Grimshaw, 89
 Gunter, 107
 Haan, 74
 Haber, 70
 Hadley, 13, 22
 Hagoort, 108
 Hahne, 108
 Halgren, 106
 Healy, 11, 12, 13,
 14, 15, 16, 17, 18,
 22, 25, 113, 130,
 144, 179, 186,
 187
 Hoen, 110
 Holmes, 118, 168
 Hyönä, 26, 64
 Inhoff, 58, 59, 60,
 64, 65
 Jacobs, 26, 65, 145
 Johnston, 71
 Just, 13, 93, 94
 Kang, 109
 Kaye, 71
 Keller, 93
 Kennedy, 4, 59, 60
 Kerr, 65
 Kilpatruck, 71
 King, 94
 Kintsch, 21
 Klein, 22, 24, 25,
 27, 71, 153, 164,
 195
 Koning, 71
 Koriat, 7, 15, 16,
 17, 18, 20, 23, 25,
 26, 27, 52, 53, 54,
 77, 78, 95, 96,
 111, 113, 114,
 115, 116, 118,
 130, 140, 141,
 143, 144, 146,
 147, 152, 153,
 154, 168, 176,
 179, 187, 191,
 219, 221, 223
 Kreiner, 19
 Lawson, 103, 170
 Layton, 170
 Le Bihan, 68
 Le Clec'H, 68
 Lecocq, 93, 153
 Leiman, 16
 Léontiev, 200, 210,
 221
 Leuwers, 153
 Liberman, 117
 Lindley, 71
 Lipka, 92
 Luck, 74
 MacDonald, 82, 90,
 95
 Mainguenaud, 35
 Marmoz, 7
 Martinet, 30, 31,
 144
 Matos, 137
 Mazoyer, 105
 McClelland, 60, 82
 McConkie, 65
 McCormick, 71
 McNamara, 22
 Mecklinger, 82, 108
 Mills, 103
 Mittau, 26
 Miyake, 93
 Moravcsik, 14
 Morris, 16, 65
 Morton, 81
 Morucci, 68
 Mulder, 107
 Mulliez, 4

Musseler, 20, 26,
 53, 78, 95, 114,
 219
 Naccache, 75, 116,
 164
 Nagel, 89
 Nazir, 145
 Neil, 71
 Neville, 103
 New, 131, 135, 137,
 154, 180
 Newhoff, 89
 Nisslein, 20
 Nougier, 71
 Ny, 18
 Oliver, 22
 Olson, 64
 Opitz, 109
 Pallier, 137
 Pasquier, 144
 Passeron, 5
 Peach, 89
 Pearlmutter, 82
 Perea, 74
 Petit, 59, 63, 64, 77,
 224
 Philippeau, 126
 Piatelli-Palmarini,
 29, 52
 Pigott, 209
 Pinker, 29, 32, 33,
 36, 40, 44, 46, 52,
 79, 98, 115, 177,
 209, 221
 Poggio, 73
 Poirier, 21, 140, 145
 Poline, 68
 Pollatsek, 13, 19,
 56, 57, 64, 65,
 109
 Pollock, 30, 36, 49,
 51, 219
 Porquier, 194
 Portalier, 223
 Posner, 6
 Pugh, 68
 Pylyshyn, 70
 Raney, 64
 Rayner, 13, 16, 19,
 26, 56, 57, 58, 64,
 65, 66, 75, 82, 86,
 109
 Rebmeister, 4
 Reddix, 65
 Reichter, 59
 Ripoll, 71
 Rubin, 89
 Rumelhart, 60, 82
 Ruschemeyer, 108
 Saint-Aubin, 21, 22,
 24, 25, 27, 140,
 145, 153, 164,
 195
 Schneuwly, 144
 Schriefers, 82
 Schuberth, 75
 Schustack, 64
 Segui, 74
 Seidenberg, 16, 82
 Serenio, 6, 64, 74
 Seron, 99
 Shapiro, 15, 29, 48,
 74, 76, 88, 89
 Shelton, 117
 Shiffrar, 72, 200
 Simon, 70
 Solotareff, 179
 Sprenger-Charolles,
 60
 Starkes, 71
 Steedman, 87
 Stein, 71
 Steinhauer, 82, 107
 Stowe, 107
 Stuart, 60
 Taft, 59
 Tanenhaus, 16, 90
 Tardieu, 19
 Taylor, 26
 Topolski, 65
 Townsend, 26
 Trueswell, 90, 95
 Umilta, 71
 Underwood, 125
 van Dijk, 21
 Vellutino, 27, 147
 Vigliocco, 74
 Vitu, 26, 64, 65, 66,
 67
 Vohel, 74
 Vugalic, 121
 Wassenaar, 108
 Waters, 93, 94
 Watteau, 153

Well, 64
Yampolsky, 94
Zagar, 58, 92

Zola, 65
Zurif, 89

Table des matières

Introduction	3
Première partie : Approches théoriques	10
Chapitre 1. L'effet d'oubli de lettres	11
1.1. Le modèle de Healy et ses évolutions	11
1.1.1. Les unités concurrentes	11
1.1.2. Le traitement parafovéal.....	13
1.1.3. Le temps de travail	14
1.2. Le modèle de Koriat et ses évolutions	15
1.2.1. Le rôle de la syntaxe.....	15
1.2.2. La lecture structurale.....	17
1.2.3. La question de la taille des structures.....	20
1.3. Modèle structural et vision parafovéale	22
1.4. Commentaires et hypothèses.....	25
Chapitre 2. Syntaxe et linguistique	29
2.1. Introduction	29
2.2. Un peu d'histoire.....	30
2.3. L'analyse syntaxique.....	32
2.4. De la succession des mots à l'analyse en constituants	34
2.5. De l'analyse en constituants à la théorie X-barre.....	40
2.6. Relations structurelles	49
2.7. Syntaxe et lexique	50
2.7.1. La sous-catégorisation.....	50
2.7.2. La grille thématique d'un item lexical	51
2.8. Conclusion.....	52
Chapitre 3. La vision et la lecture	56
3.1. Introduction	56
3.2. Vision nette et vision dégradée	56
3.3. La vision parafovéale et la lecture.....	58
3.4. Quelques travaux en ophtalmologie.....	60
3.5. La saccade oculaire	63
3.5.1. Généralités.....	63
3.5.2. Où fixer et quand partir ?	64
3.6. Une fixation par mot ?.....	66
3.7. La zone de traitement de la forme des mots.....	68
3.8. Les experts en perception visuelle	70
3.9. Apprentissage perceptif.....	72
3.10. Le traitement inconscient des informations visuelles	73
3.11. Conclusion.....	75
Chapitre 4. Traitement syntaxique et compréhension de phrase.	78
4.1. Les grands modèles explicatifs	78

4.2.	Les modèles sériels prévoyant une syntaxe autonome.....	83
4.3.	Les modèles parallèles prévoyant une syntaxe autonome.....	87
4.3.1.	Le modèle référentiel	87
4.3.2.	Le modèle de représentations lexicales.....	88
4.4.	Les modèles interactifs et parallèles.....	90
4.5.	La question de la mémoire de travail	92
4.6.	Conclusion.....	95
Chapitre 5. Les observations neurophysiologiques.....		97
5.1.	Les lésions cérébrales.....	97
5.2.	Apports des nouvelles techniques d'imagerie cérébrale	100
5.2.1.	L'IRMf. : IRM fonctionnelle	100
5.2.2.	La Tomographie par Emission de Positons.....	101
5.2.3.	Les potentiels évoqués	102
5.3.	Quelques résultats	104
5.4.	Conclusion.....	108
Deuxième partie : Approches expérimentales		112
Chapitre 6. Hypothèses.....		113
6.1.	Première hypothèse	113
6.2.	Deuxième hypothèse	114
6.3.	Troisième hypothèse	116
6.4.	Quatrième hypothèse.....	117
6.4.1.	Récapitulatif	119
Chapitre 7. Les performances en lecture de la population expérimentale.....		120
7.1.	Deux épreuves de lecture	120
7.1.1.	Epreuve n°1 : Lecture de textes courts.....	121
7.1.1.1.	Déroulement	121
7.1.1.2.	Variables primaires	121
7.1.2.	Epreuve n°2 : Lecture d'un texte long et plurisémiq.....	122
7.1.2.1.	Déroulement	123
7.1.2.2.	Variables primaires	123
7.2.	Commentaires sur les variables primaires.....	124
7.3.	Construction des indices du niveau de lecture	125
7.3.1.	Résultats de l'ACP	126
7.3.2.	Description des trois indices retenus.....	128
Chapitre 8. L'oubli de lettres : évaluations des fréquences, rôles et natures des mots.....		130
8.1.	Expérience 1	130
8.1.1.	Population.....	130
8.1.2.	Matériel	130
8.1.3.	Procédure.....	131
8.1.4.	Résultats	131
8.1.5.	Discussion	133
8.2.	Expérience 2.....	134
8.2.1.	Population.....	135
8.2.2.	Matériel et procédure	135

8.2.3.	Résultats	135
8.2.4.	Discussion	143
8.3.	Conclusion.....	145
Chapitre 9. Mise en rapport de l'effet d'oubli de lettres et des performances en lecture		147
9.1.	Introduction	147
9.2.	Rapports entre performances de lecture et barrage de lettre	148
9.2.1.	L'explication de l'oubli de lettres par les performances du lecteur	148
9.2.2.	L'explication des performances du lecteur par l'oubli de lettres	150
9.3.	Discussion	152
Chapitre 10. L'importance de la vision périphérique pour le modèle structural.....		154
10.1.	Position du problème.....	154
10.2.	Méthode.....	154
10.2.1.	Déroulement de l'épreuve	155
10.2.2.	Construction des triplets de phrases	155
10.2.3.	Population.....	158
10.3.	Résultats	158
10.3.1.	Première analyse	158
10.3.2.	Les échecs	160
10.4.	Discussion	163
Bilan de ces études expérimentales		166
Troisième partie : Place de la syntaxe dans la pédagogie de la lecture.....		167
Chapitre 11. Contexte théorique.....		168
11.1.	Syntaxe et niveau de lecture.....	168
11.2.	La place de la syntaxe dans la pédagogie de la lecture	171
11.2.1.	A l'école élémentaire	171
11.2.2.	Au collège	174
11.3.	Discussion	175
Chapitre 12. Le barrage de lettres chez les élèves de cycle 2.....		178
12.1.	Population et procédure.....	178
12.2.	Résultats	180
12.3.	Discussion	185
12.4.	Observation des différences dans le modèle structural au cours de la scolarité	188
Chapitre 13. Peut-on entraîner le lecteur à construire des cadres syntaxiques ?.....		190
13.1.	Introduction	190
13.2.	Les exercices	191
13.2.1.	Premier exercice : le cliquage sur les mots à rôle syntaxique.....	191
13.2.2.	Deuxième exercice : l'exercice à trous	193
13.2.3.	Troisième exercice : les structures syntaxiques en vision parafovéale	194
13.2.4.	Récapitulatif	197
13.3.	Les effets de l'entraînement sur la lecture.	200
13.3.1.	Population.....	200

13.3.2.	Descriptif des séances de travail	202
13.3.3.	Les épreuves d'évaluation de la lecture	203
13.3.4.	Les effets de l'entraînement sur la compréhension en lecture	204
13.3.5.	Les effets de l'entraînement sur la vitesse de lecture.....	206
13.4.	Discussion	207
Chapitre 14.	Différence de résultats en fonction du type d'entraînement dans le seul	
	groupe expérimental.....	210
14.1.	Population.....	212
14.2.	Procédures d'analyse.....	212
14.3.	Résultats	214
14.4.	Discussion	216
Conclusion.....		217
Bibliographie.....		224
Index des figures.....		242
Index des tableaux.....		244
Index des auteurs.....		246
Index des auteurs.....		246
Table des matières.....		250

RÉSUMÉ en français :

Cette thèse s'articule autour du modèle structural de lecture dans lequel la structure syntaxique de la phrase, élaborée immédiatement, guide l'intégration des unités sémantiques. Au cours de la première partie, nous examinons cette prégnance de la syntaxe par l'apport des études de linguistique, de la perception visuelle, des modèles de compréhension et des observations neurophysiologiques du cerveau.

Dans notre seconde partie, nous présentons plusieurs résultats expérimentaux obtenus chez des lecteurs adultes et lettrés. En premier lieu, nous montrons que les structures syntaxiques construites par le lecteur peuvent s'étendre jusqu'aux syntagmes flexionnels ; en second lieu, que la création du squelette syntaxique de la phrase résulte à la fois du traitement de la vision parafovéale et de la connaissance implicite des règles syntaxiques. Enfin, nous montrons que la qualité de compréhension en lecture est significativement associée à l'habileté syntaxique.

Notre dernière partie met en évidence l'apparition, au cours du cycle 2 de l'école primaire, de cette habileté syntaxique. Nous y montrons également qu'un entraînement d'élèves de 5^{ème} de collège, spécifiquement centré sur la reconnaissance des mots organisant la syntaxe et sur le repérage des structures syntaxiques par la vision parafovéale, améliore leur compréhension en lecture.

TITRE et RÉSUMÉ en anglais :

SYNTAX, PARAFOVEAL VISION AND READING PROCESS.

Contribution of structural model to pedagogy.

This thesis is based on the structural model of reading in which the syntactical sentence structure, immediately apparent, guides the assimilation of the semantic units.

During the first part, we examine this assimilation of the syntax by what is known of the study of linguistics, visual perception, models of comprehension, and neurophysiological observations of the brain.

In the second part, we present several experimental results obtained with adult and educated readers. First, we show that the syntactical structures built by the reader can extend to inflexional phrase. Second, we prove that the creation of the sentence's syntactical skeleton results from the treatment of the parafoveal vision and also of the implicit knowledge of syntactical rules. Lastly, we demonstrate that the quality of comprehension in reading is significantly associated to the syntactical skill.

Finally we highlight the appearance of this syntactical skill, during the second cycle of primary school. We also show that when the training of secondary school students (5^e) is focused specifically on the recognition of the words that organize the syntax, and on the location of the syntactical structures by the parafoveal vision, reading comprehension is notably improved.

DISCIPLINE

Sciences de l'Éducation

MOTS-CLES :

lecture, syntaxe, compréhension, vision parafovéale, entraînement.

Université Grenoble II – Pierre Mendès-France

U.F.R. Sciences de l'Homme et de la Société

Département des Sciences de l'Éducation