

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

Introduction

Les expériences relatées dans la littérature sur l'effet d'oubli de lettres n'ont jamais pris en compte 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, Koriat, & Vellutino, 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 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 cet article, 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. 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'oublis¹, variable que nous garderons par ailleurs.

1) Les performances en lecture de la population expérimentale

Pour 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

¹ 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).

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% ($F(1, 59) = 3,0878$, $p = ,08407$).

1.1) Deux épreuves 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.

1.1.1) 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).

♦ **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.

♦ 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)². Le tableau 1 présente les statistiques descriptives de cette compréhension.

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

Tableau 1 - Statistiques descriptives de la compréhension. Épreuves 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 2 présente les résultats principaux pour cette variable.

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

Tableau 2 - Statistiques descriptives de la vitesse de lecture. Épreuves de lecture n°1.

Plus de 18 000 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.

1.1.2) 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

² La transformation linéaire se fait au moyen de la formule Points = (-100/7) x (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, 100 points).

atteindre l'intention de l'auteur et apprécier les moyens qu'il emploie.

♦ **Déroulement** : un texte de fiction de Gianni Rodari, long de 1 526 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-écrans 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. Ce score de compréhension peut s'échelonner de 15 à 94 points.

♦ Variables primaires

La vitesse de lecture en mots/heure

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

Tableau 3 - Statistiques descriptives de la vitesse de lecture. Épreuves 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 quoi qu'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.

Moyenne	Médiane	Ecart-type	Minimum	Maximum
76,68	79,50	9,77992	40,00000	91,000000

Tableau 4 - Statistiques descriptives de la compréhension. Épreuves de lecture n°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 5 présente les différents résultats issus de ces travaux :

Références	Effectif	Niveau	Épreuve n°1		Épreuve n°2	
			Compr.	Vitesse	Compr.	Vitesse
Foucambert, 1997	59	CM2		14 940	54,22	11 835
Foucambert, 2000	85	6 ^{ème}		14 574	57,41	14 331
Etude présente	64	Adulte	78	18 202	76,68	20 393

Tableau 5 - 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 significative³ 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 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).

1.2) 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 varia-

³ 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.

bles 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 principale⁴. 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.

1.2.1) Résultats de l'ACP

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

Les trois premiers axes de l'analyse rendent compte, à eux seuls, de plus de 95% de la variance totale (voir tableau 6 et figure 1 page suivante)

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

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

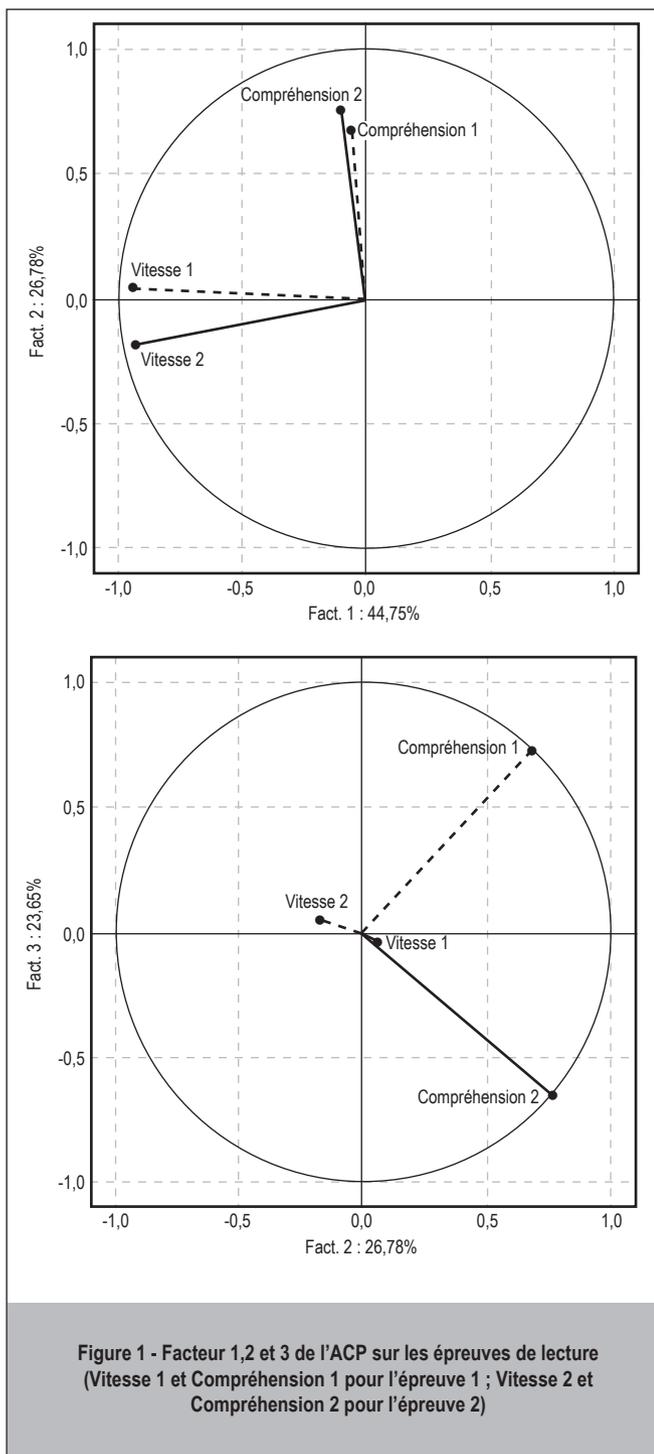
Tableau 6 - Résultats de l'ACP : Variance expliquée et valeurs propres de chacun des axes.

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%.

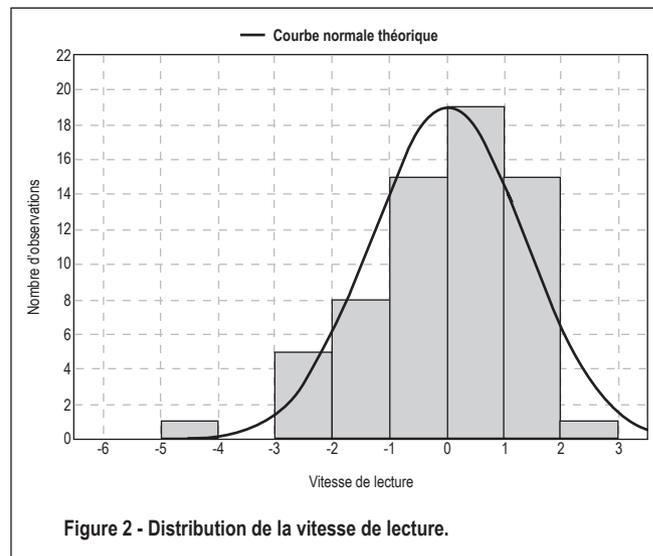
L'axe 1 représentant massivement la vitesse de lecture, la coordonnée des individus sur cet axe sera réutilisée pour définir leur vitesse de lecture dans les analyses ultérieures. De la même façon et pour des raisons similaires, les coordonnées des individus sur l'axe 2 et sur l'axe 3 définiront respectivement leur niveau et leur style de compréhension.

⁴ Cette explication du fonctionnement de l'analyse en composantes principales doit beaucoup à (Philippeau, 1986)

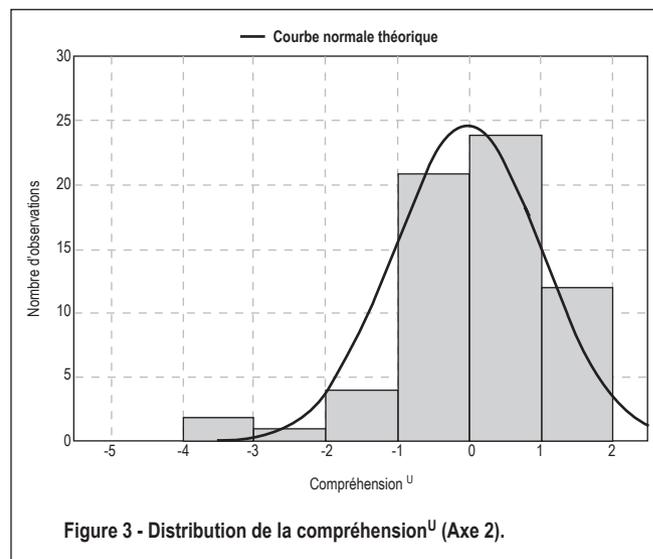


Dorénavant, nous appellerons la compréhension construite par l'axe 2 de l'ACP la compréhension^U (pour Union) et celle construite par l'axe 3 la compréhension^D (pour Différence).

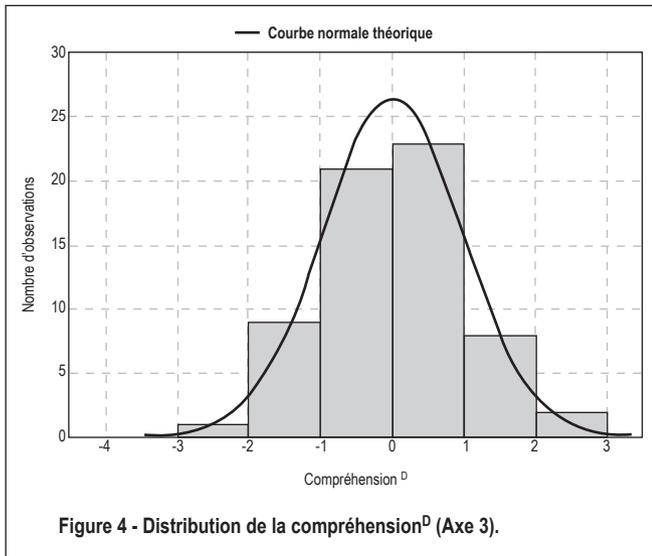
1.2.2) Description des trois indices retenus



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



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



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

2) Rapports entre performances de lecture et barrage de lettres

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 ($m = 3,5$; $s = 2,09$; minimum = 0,75 ; maximum = 10,34 ; Kolmogorov-Smirnov : $d = 0,12$; $p > 0,20$) 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'observation de l'oubli de lettres sur quatre textes différents, soit sur environ 40000 mots⁵. Le regroupement des déterminants, prépositions et compléments constitue 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 plus forts résidus⁶, afin de mieux établir les paramètres du modèle M4. Le modèle M4 apporte un gain explicatif de 10 points. Le tableau 7 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 7 - 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 8 - Paramètres des variables pour le modèle M4. Expérience 3.

⁵ Nous avons au total 617 mots observés pour 64 individus, soit un total de 39 488 mots.

⁶ Dont la valeur du résidu est supérieure à 1,5 fois l'écart-type des résidus.

Le tableau 8 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 deça du seuil de significativité ($p = 0,27$).
- La vitesse de lecture ne joue pas de rôle dans le rapport d'oublis ($p = 0,25$).
- 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.

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 9 présente les résultats généraux du modèle⁷.

Modèle	Variabes 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 9 - 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 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. La compréhension^U

semble illustrer des phénomènes plus généraux, à l'œuvre dans toute compréhension.

			Ord. Origine	Variables explicatives		
				Sexe	% 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	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	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 10 - Paramètres, coefficients bêta et significativité des variables explicatives du modèle M5.

Le tableau 10 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 hommes⁸.
- 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.
- La compréhension^D n'est expliquée par aucune des variables introduites dans le modèle.

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 structurale⁹, 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 com-

⁷ 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.

⁸ 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.

⁹ Voir Deux modèles pour un effet : explications de l'oubli de lettres, Denis FOUCAMBERT, p.66

préhension 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, Leuwens, & 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 édifée *a priori* à l'aide d'une prise d'information parafovéale (Koriat & Greenberg, 1996; 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 « pilotée » 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 l'objet d'un prochain article.

Denis FOCAMBERT ■■■

Bibliographie

- 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.
- DANEMAN, M. & CARPENTER, P. A. (1980). *Individual difference in working memory and reading*. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.
- 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.
- FOCAMBERT, D. (1997). *Conscience graphique et performance en lecture : Étude statistique sur un échantillon d'enfants de onze ans*. Mémoire de DEA Université de Caen.
- FOCAMBERT, 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.
- 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.
- KORiat, A. & GREENBERG, S. N. (1996). *Syntactic Control of Letter Detection : Evidence From English and Hebrew NonWords*. Journal of Experimental Psychology : Learning, Memory and Cognition, 17, 1035-1050.
- LECOCQ, P., CASALIS, S., LEUWERS, C., & WATTEAU, N. (1996). *Apprentissage de la lecture et compréhension d'énoncés*. Villeneuve d'Ascq: Presses Universitaires du Septentrion.
- PHILIPPEAU, G. (1986). *Comment interpréter les résultats d'une analyse en composantes principales ?* Paris : ITCF.
- 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.
- 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.
- 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.