L’apprentissage de la lecture et ses mécanismes :

de la vue sur les sons, et du son dans l’image !

AVANT-PROPOS

Les travaux présentés ici sont le résultat d'une quête sans fin : savoir "comment ça marche". S'il y a des domaines où la réponse est admise de façon quasi définitive, si l'on sait comment les cellules d'un corps se divisent, si l'on connaît les causes et les effets de nombreuses lois physiques, à l'inverse il y a peu de chances d'affirmer de façon certaine comment la lecture est acquise, quels mécanismes d'apprentissage interviennent chez l'enfant quand il s'investit dans la lecture.

Peu de domaines sont aussi délicats d'approche que les processus cognitifs engagés dans l'apprentissage de la lecture, et l'aphorisme de Wittgeinstein, "Que le soleil se lèvera demain est une hypothèse" (...) s'applique parfaitement à ces recherches.

Qu'on ne s'y méprenne pas : les démarches qui m'ont guidé ne sont pas hasardeuses, et les hypothèses ou théories qui en ont résulté possèdent un fondement tangible. Mais il serait toutefois immodeste d'oublier que ce domaine de la linguistique, comme d'autres, demande beaucoup d'humilité et de prudence, tant l'objet étudié est difficile à capter, à disséquer, à analyser.

INTRODUCTION

Mes travaux ne sont pas un aboutissement, mais une continuité dont le but ultime est, à travers la recherche fondamentale et la recherche pratique, l'utilité sociale.

La lutte contre l'illettrisme est une préoccupation de plus en plus fondée, car si autrefois cultiver un champ au milieu du seizième siècle ne nécessitait aucune maîtrise de la langue écrite, il faut aujourd'hui un minimum de savoir-lire pour prétendre à une vie simplement normale. L'environnement devient de plus en plus graphique, les messages nombreux, les affichages rapides (que l'on pense aux panneaux publicitaires rotatifs, qui intègrent sur une

même surface plusieurs publicités à la suite). Le moindre contrat d'assurance est devenu une épreuve de compréhension de texte, les "plats du jours" des restaurants s'enrichissent de "menus variables" à choix multiples, et les modes d'emploi pour appareils électroménagers de vrais problèmes d'explication de texte ! C'est de plus en plus "toujours plus et mieux pour les Français qui... lisent" !

Les chances de réussite sont bien entendues inégales, car chaque individu a des compétences plus ou moins efficaces, auxquelles s'ajoutent des environnements sociaux complexes. C'est évidemment pour les plus handicapés (et je prends ce terme à la lettre : "hand" et "cap", main et tête, écrire et lire) que mes recherches sont justifiées, même si le bon élève d'Épinal n'est pas au-dessus de toute erreur.

C'est donc tout naturellement que mes efforts se sont orientés vers ceux qui subissent le plus souvent l'échec scolaire, et vers l'étude des erreurs de lecture qu'ils commettent. Je devrais dire qu'ils produisent, car on verra que la notion de faute n'est pas appropriée, puisque les erreurs sont le résultat de mécanismes tout aussi réguliers (sinon plus !) que les lectures correctes.

Pour rester cohérent dans ma démarche, j'ai analysé principalement les erreurs de lecture d'enfants en moyenne ou grande difficulté. Elles sont les meilleurs révélateurs des mécanismes de lecture. L'idée n'est pas nouvelle, mais elle est appliquée ici de façon systématique.

Les travaux sur les processus d'apprentissage de la lecture chez l'enfant sont encore rares. La plupart portent en effet sur le monde des adultes. Essayer de transposer les résultats d'une population à une autre, c'est-à-dire des adultes vers les enfants, est risqué, voire aberrant. On ne peut de dicto décider que les mécanismes sont les mêmes, que les populations adulte et enfantine fonctionnent sur des règles identiques, que les méthodes d'investigations sont

équivalentes.

La tâche la plus délicate fut donc de créer pour ce type de recherches les moyens méthodologiques adéquats : démarche, choix de corpus, programmes informatiques, tests de vérification.

La littérature sur la question est très majoritairement anglophone, et la langue anglaise, contrairement à la langue française, est essentiellement monosyllabique, et plus complexe dans les différentes valeurs phoniques des grammes. Il n'aurait pas été sérieux de transposer littéralement les résultats sur l'anglais aux recherches sur le français. Là encore, je me suis heurté à une certaine carence bibliographique. Quelques chercheurs francophones de renom ont abordé ce domaine, mais leurs travaux sont encore difficiles d'accès (voir une liste sur mon site dans Bibliographie). Internet permet heureusement de passer outre les délais de publication, et l'on y trouve quelques articles intéressants (voir sur le site le lien Toilographie).

Les publications sont davantage tournées vers une recherche introspective, c'est-à-dire faite d'interprétation des faits à partir des performances des locuteurs adultes, en quelque sorte une recherche impressionniste.

Ma méthodologie s'appuie au contraire sur l'idée suivante, qu'on pourra taxer de postulat : les mécanismes psycholinguistiques de l'adulte sont différents de ceux de l'enfant, et la réflexion faite par l'adulte sur les performances de l'enfant sont faussées par ses propres mécanismes de lecteur expert (cf. Alain Bentolila, "Il est important d'établir une distinction claire entre apprendre à lire et savoir lire : le comportement du lecteur expert ne nous fournit pas directement un modèle d'apprentissage. Lorsque l'élève apprend à lire, il doit nécessairement découvrir comment fonctionne le code écrit et comprendre notamment le principe des mécanismes qui relient les unités graphiques et les unités phoniques de l'oral.

L’adulte qui sait lire maîtrise ces mécanismes avec une telle dextérité qu’il en oublie presque son parcours - parfois laborieux - d’apprentissage.”, Roll, Enseignants-Informations sur la lecture-Evaluation de la lecture).

À partir de cette position théorique, j’ai délibérément choisi de ne mesurer que les productions d’enfants en lecture, et d’élaborer des hypothèses sur les causes des erreurs rencontrées.

En d’autres termes (et cela fait suite à un constat lorsque j’enseignais le français à des enfants bandjabi, dans le Sud-Est du Gabon, voir “La connaissance du fonctionnement des langues maternelles peut-elle contribuer à l’acquisition du français ?”, conférence de Libreville, Gabon, 1987) : les erreurs de lecture sont les meilleurs indicateurs du fonctionnement cognitif.

Pour mettre mes hypothèses à l’épreuve des faits, j’ai fabriqué quelques outils, en particulier des tests (Test01, Test02, Test03). Chacun d’eux a nécessité un soin particulier, méthodique, afin que les interprétations abusives soient évitées. Chacun de ces tests est falsifiable, et a été réalisé de façon à ce qu’il soit discutable, au sens scientifique du terme. Ils sont disponibles dans cette étude.

Pour donner à la critique une vue d’ensemble de ces travaux, voici leur synthèse.

1. APPRENTISSAGE OU ENSEIGNEMENT ?

1.1. État des lieux.

Un constat général d’échec en lecture a été réalisé par différentes enquêtes institutionnelles. Leurs conclusions ont pu révéler, à quelques variantes près (il s’agit de différences de pourcentages), un taux élevé d’illettrisme, voire d’analphabétisme, d’abord chez les adultes

du contingent militaire, puis chez les élèves à l'entrée en sixième.

En septembre 1997, les évaluations pour l'entrée en 6ème analysées par la Direction de l'Évaluation et de la Prospective ont révélé que 12% des collégiens ne savaient pas lire, que 50% déchiffraient mais ne comprenaient pas vraiment ce qu'ils lisaient (relevant ainsi de l'illettrisme). L'échec en lecture, associé jusqu'alors à l'analphabétisme, s'étendait ainsi à l'illettrisme. On était loin des "10% d'élèves en échec".

1.2. Réaction

Le cri d'alarme a saisi les instances politiques, et les recherches sur ce problème, déjà entamées, ont proliféré.

La qualité de ces différentes recherches n'est pas mise en question ici. On ne peut que leur reconnaître des avancées cruciales en ce domaine : grâce à elles, on sait à peu près comment peut fonctionner l'apprentissage de la lecture. On doit par exemple à Frith Utha d'avoir dès 1985 proposé les trois étapes (logographique, alphabétique, orthographique) qui servent de réflexion à d'autres recherches, mieux cadrées par cette catégorisation.

Cependant, un certain flou a souvent régné à travers ces recherches, une confusion apparemment anodine, qui a engendré des pistes parfois contradictoires : l'amalgame entre apprentissage et enseignement, c'est-à-dire, au bout du compte, entre les mécanismes cognitifs et les pratiques pédagogiques.

Il est bien entendu que les deux notions sont liées, intimement, et qu'une meilleure connaissance de l'une (l'apprentissage) permet à l'autre (l'enseignement) une plus grande efficacité. Mais les différents auteurs, d'horizons différents (linguistes, psychologues, didacticiens, voire neurologues, orthophonistes, etc.), impliqués dans leurs travaux, restent chacun dans sa discipline. On trouve ainsi d'excellentes études faites par des théoriciens sur les mécanismes d'apprentissage de la lecture. Mais leur souci de les rendre utilisables par les

praticiens les fait glisser de la notion d'apprentissage vers celle d'enseignement. Savoir (ou admettre) qu'un cerveau d'enfant passe d'abord par une phase logographique avant une phase alphabétique puis orthographique relève de quel domaine ? S'il s'agit de l'apprentissage, alors il faut offrir aux acteurs directs de l'enseignement, les enseignants, des outils conformes à ces avancées théoriques. S'il s'agit d'enseignement, alors chaque enseignant peut simplement se référer à ce modèle et utiliser les supports pédagogiques de son choix.

1.3. Perspectives

Ma position est sans équivoque : mes travaux s'inscrivent d'une part dans une perspective de recherche fondamentale qui a pour but de traquer les mécanismes d'apprentissage, c'est-à-dire une tentative (encore une fois bien modeste, eu égard aux divers autres travaux sur la lecture) de compréhension des processus mis en oeuvre lorsque l'enfant lit, sous la lumière des erreurs qu'il commet, et d'autre part, dans une perspective de recherche appliquée pour l'enseignement, par l'étude détaillée des manuels présents sur le marché.

Cette position implique que l'enseignement est une conséquence des connaissances (relatives) que l'on a de l'apprentissage. Par exemple, la mise en évidence d'un fait linguistique particulier comme l'influence de l'oral sur les erreurs de lecture, doit être intégrée dans les pratiques pédagogiques. Certes, les enseignants n'ont pas attendu les chercheurs pour pratiquer l'oral dans leurs classes, mais savent-ils quel type d'oral, quel type de mots, quel type de sons, il vaut mieux que tel autre ? Savent-ils pourquoi il serait préférable d' utiliser telle comptine plutôt qu'une autre lorsqu'ils rencontrent un certain type d'erreur de lecture ?

On comprend bien mon souci : les recherches sur l'apprentissage doivent éclairer les mécanismes cognitifs en s'appuyant sur des faits linguistiques tangibles et réguliers (qu'on peut appeler un corpus d'erreurs), pour permettre ensuite des recherches appliquées sur l'enseignement, qui porteront sur les supports matériels présents en classe, en particulier et

surtout le manuel de lecture, dont on ne peut guère faire l'économie si l'on veut que l'enseignant ait la possibilité d'utiliser un soutien pédagogique correspondant à ce qu'on pense savoir des processus cognitifs.

1.4. Erreurs et Déterminisme : Règles sous-jacentes

Avant de passer au contenu de mes travaux, je crois nécessaire de poser les conditions méthodologiques qui les soutiennent.

Comme tout praticien peut s'en apercevoir chaque jour, les erreurs de lecture existent, et c'est bien elles qui sont le constituant majeur des problèmes de lecture ! Derrière ce truisme apparent se cache une conception clairement affirmée : un enfant qui lit mal n'est pas obligatoirement un enfant intellectuellement déficient (on le voit très bien avec les dyslexiques qui ont de bons résultats dans les autres matières, y compris en mathématiques). Un enfant qui lit mal met tout simplement en oeuvre des processus de lecture différents de ceux qu'il faudrait.

Pour illustrer cette idée, il suffit de se rappeler les erreurs (assez fréquentes à l'oral) résultant d'une analogie : "ils sontaient" au lieu de "ils étaient", "vous disez" à la place de "vous dites". Grammaticalement, il y a une erreur monstrueuse, mais logiquement, la forme est plus régulière... Autrement dit, voilà des erreurs mieux construites que leurs pendants légitimes !

À l'écrit, la lecture erronée de mots, de syllabes, ou de lettres, n'est pas un hasard. Comment expliquer qu'un même mot soit mal lu de la même manière par plusieurs dizaines d'enfants (observations faites avec différents tests) ?

Bien sûr, les réalisations orales possibles d'une même suite de lettres sont limitées, et par là-même les erreurs ont une marge de production également limitée. Mais pourtant, sur ces différentes erreurs, certaines l'emportent en nombre de réalisations, et très nettement. La

conclusion la plus plausible est qu'une règle sous-tend la production d'erreurs.

Au chercheur de trouver lesquelles (et c'est l'étude sur l'apprentissage), pour proposer ensuite des remédiations ou des recommandations (et c'est l'étude sur l'enseignement).

2. ÉLABORATION D'OUTILS

2.1. Corpus de référence

Mon premier travail a été d'étudier non pas les erreurs de lecture (il y en aura toujours !), ni les manuels (ils seront sans doute toujours là !), mais d'abord et surtout le terrain sur lequel tout le monde s'active, chercheurs, enseignants, lecteurs débutants : la langue écrite et parlée.

Il me fallait un état des lieux : quelles sont les lettres (= les grammes*) les plus fréquentes, quelles sont les combinaisons de lettres (=les polygrammes*) les plus rencontrées, quels sont les mots les plus employés, quels sont les phonèmes et les séquences phonétiques les plus probables ?

Ce travail avait déjà été réalisé, ici ou là, de façon parcellaire, et sans aucune indication méthodologique. Comment leurs auteurs (y compris des chercheurs incontournables et dont les travaux remarquables ont permis une avancée importante dans leur discipline, comme Charles Muller ou Pierre Léon) ont-ils comptabilisé les occurrences d'items ? Sur quel corpus se sont-ils appuyés ? Quels programmes (ou quels enquêteurs) ont-ils utilisés ? Où sont tous leurs résultats, pour permettre aux autres chercheurs d'avoir eux-aussi un droit de regard ?

Pour ma part, j'ai réuni une série de textes de tous horizons, de tous auteurs, de tous genres (des recettes de cuisine aux pièces de théâtre, en passant par des textes documentaires, des poésies, des romans, des articles, etc.).

Voici la liste abrégée de ces textes :

Productions d'enfants : Le Lutin ; Noisette et petite plume.

Recettes de cuisine ;

Contes et nouvelles : Barbe Bleue ; Cendrillon ; Chats et souris emménagent ; La Bergère et le ramoneur ; Grand ours ; Homo informaticus ; La cuisine de Muriel ; La légende de la nuit polaire ; La malle volante ; La marelle ; La petite fille et les allumettes ; La princesse de pierre ; La reine des neiges ; La rose qui guérit ; Le briquet ; Le chasseur solitaire ; Le diable et sa grand mère ; Le génie de la forêt ; Le lièvre et le grand génie de la brousse ; Le marabout vicieux ; Le mois de mars ; Le noël du Père Noël ; Le petit Chaperon rouge ; Le puits enchanté ; Le rêve de Tao ; Le rêve vendu ; Le rossignol ; Le sapin ; Lulu la luciole ; Peter Pan ; Sira et le sorcier ; Un papa qui avait le sens de l'humour ; Une mère peu ordinaire ; Voilà pourquoi l'eau de mer est salée.

Lexiques : Glottochronologie ; dictionnaire simplifié français pour l'Afrique.

Poésies : Apolinaire ; Baudelaire ; La Fontaine ; Hérédia ; Laforgue ; Rimbaud ; Verlaine.

Documentaires divers : Les aliments ; La chanson ; La civilisation rurale ; L'espace ; Les oiseaux ; La reproduction.

Romans : Arsène Lupin (La Cagliostro se fâche) ; Flaubert (Un coeur simple) ; Germinal ; La comédie humaine tome 1 ; Le Tour du monde en quatre-vingt jours ; Le château des Carpathes.

Histoires : Historiettes de Tallemant des Réaux.

Histoire : Déclaration des Droits de l'homme.

Réflexions : La Bruyère ; La Rochefoucault.

Théâtre : L'Avare ; La Thébaïde ; Alexandre ; Andromaque ; Les Plaideurs ; Britannicus ; Bérénice ; Bajazet ; Mithridate ; Iphigénie ; Phèdre ; Esther ; Athalie.

À ces textes qui représentaient moins de 4000 pages, j'ai ajouté d'autres textes pour un total dépassant 17000 pages (l'ensemble représente 42 Mo sur disque dur, et la liste complète est disponible sur mon site), pour un total in fine de plus de 7 millions de mots.

2.1.1. Grammes et polygrammes

Pour analyser ce corpus de base, j'ai utilisé un programme en Cobol*, écrit pour la circonstance, dont le but était de dénombrer statistiquement (en nombre brute d'occurrences et en fréquences) les grammes (=toutes les lettres de l'alphabet, y compris celles qui font partie du code Ascii étendu, à savoir : "é, ê, ï, ç, etc.)" et les polygrammes (=les graphies complexes, comme "ien", "oi", "eau", etc.). Ce travail est disponible dans sa totalité sur mon site.

Comme tout programme informatique, il n'est pas parfait. Mais le but de ce programme était de montrer les fréquences, c'est-à-dire les pourcentages de chaque item par rapport aux autres, ce qu'il arrive à faire de façon très précise. Les programmeurs reconnaissent que le traitement automatique de la langue écrite est régi par des exceptions dont il faut tenir compte dans l'algorithme, mais que l'intégration de ces exceptions alourdit considérablement le programme, et n'est jamais exhaustive. La nécessaire obligation est d'utiliser toujours le même programme pour le ou les mêmes textes, et non un programme différent pour une même étude sur des textes divers. Cette condition a toujours été respectée ici.

Voici la liste complète des grammes retenus : **a à â b c ç d e é è ê ë f g h i î ï j k l m n o ô p q r s t u ù û v w x y z,**

et la liste des polygrammes : **ai ail aill aim ain am an au ay ca ce ch ci co cu cy ean eau ei eil eill eim ein ell em en ent er err es ess est et ett eu euil ex ey ez ga ge gi gn go gu ien il ill im imm in inn oeu oi oin om on ou ouil oy ph qu tion un uy**.

2.1.2. Résultats

118 items sur 17290 pages ou 35 000 000 de caractères :

(nota : contrairement à la recommandation de Gaston Bachelard, qui pourchassait la précision décimale des calculs, j'ai opté délibérément pour une précision à plusieurs décimales. La raison est double : d'abord nous avons aujourd'hui des instruments qui nous permettent par une simple touche d'obtenir un calcul très précis sans aucun effort, ce qui n'était pas le cas en 1850, et ensuite parce que la précision obtenue pourra être utile pour d'autres domaines, en particulier le traitement automatique de la parole.)

item	classement	pourcentage	1 POUR ...
e	1	13,91186915	7,18810671
s	2	7,360535411	13,5859682
a	3	7,295998309	13,70614353
i	4	6,674865039	14,98157632
t	5	6,60317773	15,1442236
n	6	6,399030627	15,62736699
r	7	6,1274795	16,31992404
u	8	6,003583425	16,65671865
l	9	5,209300581	19,196435
e# (finale)	10	4,970548072	20,11850576
o	11	4,917626851	20,33501179

d	12	3,281950732	30,46968348
s# (finale)	13	3,084774297	32,41728256
m	14	2,685927255	37,23109023
t# (finale)	15	2,587129576	38,6528765
c	16	2,5843282	38,69477569
p	17	2,40482352	41,58309297
es	18	1,843479546	54,24524522
en	19	1,689041913	59,20516195
ai	20	1,667018477	59,98733749
v	21	1,595832737	62,66320878
é	22	1,591569401	62,83106468
ou	23	1,437795778	69,5509067
on	24	1,278328207	78,2271716
an	25	1,083408832	92,30125971
q	26	1,078914661	92,68573655
qu	27	1,07069007	93,39770937
f	28	0,991938049	100,8127474
er	29	0,942220029	106,1323225
b	30	0,849184693	117,7600125

g	31	0,834795934	119,7897545
h	32	0,802889311	124,5501698
et	33	0,794328441	125,8925085
eu	34	0,775106382	129,0145486
ent	35	0,716491213	139,5690529
in	36	0,691387118	144,6367706
il	37	0,689121509	145,1122897
ce	38	0,609300804	164,1225474
co	39	0,598434427	167,1026858
oi	40	0,572691971	174,6139373
un	41	0,569531517	175,582908
au	42	0,536259828	186,4767687
j	43	0,486775494	205,4335135
à	44	0,453700443	220,4097473
ch	45	0,435729456	229,500206
s,	46	0,423193083	236,2987582
em	47	0,413734519	241,7008865
om	48	0,376307788	265,7399164

x	49	0,370363056	270,0053324
ien	50	0,322625663	309,9567437
è	51	0,316310454	316,1450992
ell	52	0,278094322	359,5902258
y	53	0,256828939	389,3642214
t,	54	0,255441075	391,4797175
ge	55	0,234144345	427,0869756
x# (finale)	56	0,22646692	441,5655933
est	57	0,212539833	470,5000402
ê	58	0,212066762	471,5496143
z	59	0,189194079	528,5577665
s.	60	0,185110282	540,2185051
am	61	0,184845249	540,9930776
ill	62	0,184215438	542,8426695
ca	63	0,160242723	624,0532999
ez	64	0,156529403	638,8576084
es,	65	0,147999881	675,6762174
ci	66	0,143551307	696,6150441
ain	67	0,140815476	710,1492148

ett	68	0,136569239	732,2293101
ei	69	0,128398796	778,8235046
im	70	0,127937124	781,6339518
ess	71	0,110342314	906,2706423
tion	72	0,106988072	934,6836343
t.	73	0,106782885	936,4796637
ga	74	0,105674303	946,3038483
gn	75	0,099245102	1007,606403
eau	76	0,090929317	1099,75532
oin	77	0,086426596	1157,051241
â	78	0,078578182	1272,617887
cu	79	0,076959482	1299,38504
err	80	0,071735756	1394,004966
es.	81	0,071567616	1397,280014
gu	82	0,070892208	1410,592258
eil	83	0,068073733	1468,995395
ç	84	0,065845171	1518,714261
oy	85	0,06489618	1540,922756

ex	86	0,062385485	1602,93696
ô	87	0,056358109	1774,367567
î	88	0,056292563	1776,433605
er,	89	0,054374631	1839,092925
û	90	0,052063425	1920,734195
ù	91	0,045699769	2188,194874
ail	92	0,045015811	2221,441694
ein	93	0,044861921	2229,061936
gi	94	0,043294518	2309,761256
eill	95	0,042393974	2358,825827
ay	96	0,042257182	2366,461627
er.	97	0,037495127	2667,013225
go	98	0,036580334	2733,709333
x,	99	0,036144311	2766,68714
aill	100	0,035702588	2800,917385
oeu	101	0,033556671	2980,033376
et,	102	0,0322828	3097,624735
aim	103	0,028663524	3488,754524
ph	104	0,027133169	3685,525995

x.	105	0,018529552	5396,784528
euil	106	0,012764359	7834,314133
ean	107	0,012257091	8158,542897
ouil	108	0,010749534	9302,728791
k	109	0,010621292	9415,050443
ï	110	0,009495612	10531,18037
imm	111	0,008606467	11619,16987
uy	112	0,008452576	11830,71241
w	113	0,006551744	15263,11135
inn	114	0,006386454	15658,14056
ë	115	0,004163592	24017,72279
ey	116	0,003590778	27849,12143
cy	117	0,001430611	69900,18526
eim	118	0,000125392	797497,5682

2.1.3. Vocable

Pour faire l'inventaire du vocabulaire et du vocable, j'ai préféré le langage Java, et un programme (co-écrit aussi pour la circonstance) qui a permis d'avoir tous les mots du corpus général, classés par ordre décroissant, avec leur fréquence d'emploi. À volonté, on peut choisir avec la casse (c'est-à-dire avec prise en compte des lettres capitales et des minuscules, ce qui permet par exemple d'observer combien de fois "Le" commence une phrase, grâce à la majuscule) ou sans la casse (ce qui permet de ne pas compter pour des items différents les

mêmes mots écrits ou non en majuscules, comme “De”, “de”, “DE”).
Résultats :

Les 32 premiers items (pour le classement complet, voir sur le Site)

items	pourcent.
de	3,32
le	2,39
la	2,23
et	2,05
il	1,87
les	1,74
à	1,66
est	1,58
un	1,45
l'	1,36
en	1,19
pas	1,10
je	1,00
que	0,96
des	0,92
une	0,91
a	0,79
tu	0,79
c'	0,77
se	0,75
ne	0,71
qui	0,71
dans	0,70
on	0,67
pour	0,65
mais	0,62
du	0,62
elle	0,56
au	0,51
sur	0,48
son	0,46
tout	0,46

2.2. Corpus gigantesque ou corpus suffisant ?

Le traitement de nombreuses données, même numérisées, est assez coûteux en temps de calcul et en vérifications. Pour contourner ce problème, j’ai dans un premier temps envisagé d’effectuer la recherche de fréquences (grammes, polygrammes, vocables) sur une partie du

corpus. Les fréquences pour les mots les plus utilisés (mots outils et mots de base) se sont avérées suffisamment proches dans les deux corpus (7 millions de mots et 1 millions de mots environ) pour envisager le bien-fondé d'un corpus plus restreint. Mais alors quelle partie du corpus fallait-il supprimer ? Les textes étaient d'origines diverses, et mis à la suite sans ordre particulier. Enlever une quelconque partie relevait de l'arbitraire, et la conséquence eut été que certains vocables peu fréquents auraient pu disparaître.

L'O.N.L. m'a fourni aimablement un corpus de taille suffisante, et dont les textes étaient en rapport direct avec ma préoccupation : la lecture chez les enfants. Les fréquences pour les grammes-polygrammes-vocables de ce corpus de littérature de jeunesse devait donc être comparées à mon corpus général pour répondre à cette question : les fréquences visuelles entre les deux corpus étaient-elles identiques ?

Pour cela, j'ai écrit un petit programme en langage Perl. Voici pour illustration des exemples de comparaison pris au hasard. La similitude était très proche entre les deux corpus, ce qui m'autorisa l'utilisation définitive du corpus de littérature de jeunesse à la place du corpus général.

Exemples (il s'agit de "morceaux" de mots, non de mots entiers) :

"au" ..... corpus général : 0,562 %
corpus littérat. : 0,569 %

"z"..... corpus général : 0,163 %
corpus littérat. : 0,168 %

"e"..... corpus général : 14,18 %
corpus littérat. : 13,26 %

"dr"..... corpus général : 0,126 %
corpus littérat. : 0,139 %

"s"..... corpus général : 7,494 %

corpus littérat. : 7,365 %

"joli(...)"..... corpus général : 0,0043 %

corups littérat. : 0,0038 %

"dans(...)"..... corpus général : 0,173 %

corpus littérat. : 0,160

2.2.1. Littérature de jeunesse : vocable (programme en Java sur les mots entiers) (Les 201 premiers items)

item	sur 391257 caractères	Pourcentage	rang
de	12996	3,32160191	1
le	9378	2,39689002	2
la	8744	2,23484819	3
et	8052	2,05798235	4
il	7338	1,8754936	5
les	6817	1,74233304	6
à	6532	1,6694909	7
est	6218	1,58923674	8
un	5697	1,45607619	9
l	5353	1,36815444	10
en	4684	1,19716708	11
pas	4306	1,10055539	12
je	3922	1,00241018	13
que	3764	0,96202752	14
des	3630	0,92777893	15
une	3580	0,9149996	16
a	3121	0,79768541	17
tu	3103	0,79308485	18
c	3031	0,77468263	19
se	2958	0,75602481	20
ne	2801	0,71589773	21
qui	2794	0,71410863	22
dans	2777	0,70976366	23
qu	2706	0,69161702	24
on	2653	0,67807094	25
s	2589	0,6617134	26
pour	2549	0,65148994	27
mais	2435	0,62235308	28
du	2428	0,62056398	29
n	2296	0,58682656	30
elle	2228	0,56944668	31
ce	2120	0,54184334	32
au	2011	0,51398441	33
sur	1896	0,48459197	34
plus	1892	0,48356962	35
ratus	1852	0,47334616	36

son	1833	0,46849002	37
dit	1804	0,46107801	38
tout	1801	0,46031125	39
lui	1690	0,43194115	40
j	1593	0,40714927	41
ils	1585	0,40510457	42
avec	1584	0,40484899	43
ça	1409	0,36012135	44
fait	1291	0,32996215	45
ai	1270	0,32459483	46
bien	1269	0,32433925	47
sa	1191	0,3044035	48
moi	1184	0,30261439	49
par	1176	0,3005697	50
comme	1135	0,29009066	51
si	1134	0,28983507	52
nous	1114	0,28472334	53
était	1077	0,27526664	54
ses	987	0,25226386	55
faire	949	0,24255157	56
me	911	0,23283928	57
m	901	0,23028342	58
sont	863	0,22057113	59
avait	811	0,20728064	60
même	776	0,19833511	61
te	773	0,19756835	62
là	767	0,19603483	63
leur	763	0,19501249	64
être	739	0,18887841	65
alors	726	0,18555579	66
mon	715	0,18274433	67
grand	691	0,17661026	68
deux	676	0,17277646	69
quand	668	0,17073177	70
ou	666	0,1702206	71
as	651	0,1663868	72
aussi	644	0,16459769	73
suis	602	0,15386306	74
non	601	0,15360748	75
aux	591	0,15105161	76
ont	578	0,14772899	77
rat	557	0,14236167	78
où	551	0,14082815	79
puis	543	0,13878346	80
cette	538	0,13750553	81
demande	526	0,13443849	82
sans	518	0,1323938	83
ma	511	0,13060469	84
peut	494	0,12625972	85
rien	489	0,12498179	86
petit	484	0,12370386	87
encore	459	0,1173142	88

es	459	0,1173142	89
après	448	0,11450274	90
peu	446	0,11399157	91
oui	429	0,1096466	92
fois	425	0,10862425	93
faut	424	0,10836867	94
belo	413	0,10555722	95
père	396	0,10121225	96
dire	382	0,09763404	97
autre	363	0,09277789	98
temps	361	0,09226672	99
ces	352	0,08996644	100
jamais	350	0,08945527	101
répond	346	0,08843292	102
bon	345	0,08817734	103
monde	345	0,08817734	104
max	344	0,08792175	105
roi	339	0,08664382	106
maintenant	338	0,08638823	107
sous	337	0,08613264	108
peur	333	0,0851103	109
sais	330	0,08434354	110
mamie	329	0,08408795	111
autres	328	0,08383237	112
lili	327	0,08357678	113
leurs	322	0,08229885	114
gros	315	0,08050974	115
avoir	313	0,07999857	116
chez	313	0,07999857	117
mal	309	0,07897622	118
mina	307	0,07846505	119
enfants	305	0,07795388	120
mes	303	0,0774427	121
coup	295	0,07539801	122
maison	291	0,07437567	123
porte	290	0,07412008	124
devant	289	0,07386449	125
été	288	0,07360891	126
marou	272	0,06951952	127
petite	271	0,06926394	128
déjà	269	0,06875276	129
avant	267	0,06824159	130
papa	264	0,06747483	131
parents	262	0,06696366	132
parce	259	0,0661969	133
euh	258	0,06594131	134
étaient	257	0,06568573	135
école	256	0,06543014	136
comment	255	0,06517455	137
quoi	253	0,06466338	138
ah	249	0,06364103	139
ta	248	0,06338545	140

seul	247	0,06312986	141
aller	246	0,06287427	142
pourquoi	245	0,06261869	143
chien	240	0,06134076	144
car	239	0,06108517	145
ici	239	0,06108517	146
jour	231	0,05904048	147
eux	228	0,05827372	148
quelques	227	0,05801813	149
chose	226	0,05776254	150
peux	226	0,05776254	151
ralette	224	0,05725137	152
allez	223	0,05699579	153
maman	222	0,0567402	154
notre	219	0,05597344	155
personne	217	0,05546227	156
sûr	215	0,05495109	157
oh	208	0,05316199	158
peut-être	208	0,05316199	159
beaucoup	207	0,0529064	160
main	205	0,05239523	161
romain	205	0,05239523	162
crie	204	0,05213964	163
regarde	202	0,05162847	164
moment	201	0,05137288	165
elles	200	0,0511173	166
alexandre	199	0,05086171	167
hommes	199	0,05086171	168
entre	198	0,05060612	169
moins	196	0,05009495	170
pendant	195	0,04983936	171
rouge	194	0,04958378	172
depuis	192	0,0490726	173
monsieur	192	0,0490726	174
fromage	189	0,04830585	175
jeannette	188	0,04805026	176
fais	185	0,0472835	177
juste	185	0,0472835	178
grand-mère	184	0,04702791	179
mieux	181	0,04626115	180
tard	181	0,04626115	181
baptiste	179	0,04574998	182
grande	179	0,04574998	183
nuit	179	0,04574998	184
donc	178	0,04549439	185
crois	177	0,04523881	186
côté	176	0,04498322	187
derrière	176	0,04498322	188
maître	176	0,04498322	189
nouveau	176	0,04498322	190
contre	175	0,04472763	191
quelque	174	0,04447205	192

place	171	0,04370529	193
raldo	171	0,04370529	194
près	167	0,04268294	195
maîtresse	165	0,04217177	196
cela	163	0,0416606	197
enfin	163	0,0416606	198
jours	163	0,0416606	199
droit	162	0,04140501	200
jouer	161	0,04114942	201

3. NÉCESSITÉ DE TESTS

3.1. Pistage des erreurs et premières observations

Après ce travail réalisé sur le français écrit (corpus général et littérature de jeunesse), je me suis attaché à pister les erreurs de lecture.

Le choix de cette traque a été à chaque fois dicté par les observations sur le terrain. Après la simple observation d'une erreur récurrente, à savoir une permutation dans une matrice syllabique du type CVC transformée en CCV, puis d'une identification phonologique banale (une syllabe fermée se transforme en syllabe ouverte), et enfin d'une remarque neutre (ce qui est écrit est lu différemment à l'oral), j'ai postulé que l'oral pouvait influencer la lecture de l'écrit.

Pour cela, il me fallait deux preuves : d'une part, si un élément influence un autre, il doit nécessairement être quantitativement beaucoup plus important, pour avoir une véritable influence. Et en effet, environ 80% des syllabes du français oral sont ouvertes.

D'autre part, quelques erreurs de lecture CVC > CCV (exemple : "pal" lu "pla") devaient être vérifiées à grande échelle, d'où la nécessité d'un test dédié à cette tâche.

3.2. Bilinguisme ou bilingualité ?

Le Test01, effectué au départ avec moins d'une dizaine d'élèves français, a été ensuite étendu pendant plusieurs mois sur quatre écoles avec plus de deux cents élèves marocains francophones. Les résultats étaient identiques, ce qui m'a permis de considérer que les enfants marocains francophones, avec une langue maternelle qui elle aussi privilégie les syllabes ouvertes, procèdent de la même façon lorsqu'ils déchiffrent des mots.

Pour ces enfants la notion de bilinguisme ne s'applique pas, si l'on se réfère à la distinction que font Hamers et Blanc, Bilinguality and Bilingualism, 1989. Ces auteurs distinguent le bilinguisme qui réfère au fait qu'une société utilise plus d'une langue (il s'agit d'un concept d'aménagement linguistique) et la bilingualité qui renvoie à la connaissance et à l'utilisation de plus d'une langue par un individu (concept psycholinguistique).

Les élèves marocains francophones qui ont passé les tests ont une bilingualité dite enfantine (âge précoce d'acquisition) simultanée (école privée, effet de prestige, enseignement bilingue quotidien), équilibrée (compétences identiques en les deux langues), et endogène (le français est omniprésent au Maroc).

N'oublions pas également qu'en France la proportions d'enfants d'origine maghrébine est parfois très élevée, et que ces enfants font partie intégrante du public scolaire. Leurs erreurs de lecture sont les mêmes que ceux des élèves franco-français.

3.3. Passation des tests et protocole

Chaque test (Test01, Test02, Test03) a été élaboré en fonction d'erreurs observées, avec pour référence comparative un fait de langue (orale ou écrite, selon le cas), passés par des enseignants préparés à cela (par des réunions que j'ai effectuées pour les sensibiliser aux problèmes méthodologiques des passations de tests, en particulier la neutralité et le système de transcription qu'il fallait adopter), vérifiés après coup (à nouveau par des réunions avec les

mêmes enseignants pour vérifier que nous parlions le même langage et que les transcriptions étaient exactes), puis analysés item par item (avec utilisation de sous-totaux sous tableur).
En fin de compte, j'ai confronté les résultats des tests avec l'hypothèse de départ. Une aberration (au sens stricte du mot) comme le non fonctionnement d'une matrice sur les dix-sept que j'avais répertoriées pour le Test01 m'a amené à chercher une cause supplémentaire, étudiée par le Test03, pour comprendre ce type d'erreur.
Je n'entre pas davantage dans les détails de ce point de recherche sur l'influence de l'oral sur l'écrit. Nous y reviendrons plus loin, car ces quelques paragraphes avaient seulement pour but d'expliciter, même brièvement, ma méthodologie.

4. ÉLABORATION DES TESTS

4.1. Test01 : Pour une rection idiophonologique dans l'acte de décodage.

À l'oral l'influence de la langue maternelle sur l'articulation d'une autre langue est démontrée largement dans la littérature. Christophe Pallier (1994) a donné pour exemple le cas de l'épenthèse d'une voyelle [u] entre deux consonnes chez les Japonais dans la prononciation de pseudo-mots, la langue japonaise interdisant une suite de consonnes. Ce fait linguistique est probablement valable lorsqu'un apprenti-lecteur décode un mot : son bagage phonologique maternel entre en interférence et influence le décodage d'un autre système, celui de l'écriture de sa propre langue. Mais faut-il le démontrer .

Pour exemple, la prédominance du phonologique sur le visuel peut se constater dans la phrase suivante "FINISHED FILES ARE THE RESULT OF YEARS OF SCIENTIFIC STUDY COMBINED WITH THE EXPERIENCE OF YEARS". Si l'on compte le nombre de F, on en

trouve... (petit problème anonyme trouvé sur Internet). J'ai fait le test à un enfant non anglophone de cinq ans : la réponse fut juste. Aux adultes bilingues, la réponse est fausse.

La prononciation de F est parfois [v], d'où l'erreur de comptage.

Les erreurs de reconnaissance ne seraient pas obligatoirement d'ordre visuel, mais également phonologique. Il y a prédominance de l'auditif (même subvocalisé) sur le visuel. Il restait à démontrer ce point, et à élaborer pour cela une méthodologie en réalisant des tests, en français, pour repérer les phonèmes les plus influençables.

L'importance d'une influence phonologique dans l'acte de lecture n'est pas négligeable, car une conséquence pour l'enseignement serait, si ce point s'avérait exact, de privilégier les environnements phonologiques au détriment d'exercices répétitifs portant sur des lettres sorties de leur contexte.

4.1.2. Contraintes phonotactiques

Les fréquences des séquences consonantiques sont intégrées au système phonologique de l'enfant, à partir du moment où il parle sa langue. En conséquence, il est concevable que ces fréquences deviennent une contrainte probable. Par exemple, après [s] on attendra davantage une voyelle ou une consonne habituelle (p, t, l, etc.), qu'une consonne peu probable (comme r). Cela peut donc entraîner une production d'erreur chez l'enfant. De même, quatre-vingt pour cent des syllabes du français sont ouvertes (attaque + rime, sans coda). Le décodage dans un mot entraînerait donc une hypercorrection : "fort" (attaque + rime + coda) sera lu "fro" (branchante + rime).

Les contraintes phonotactiques de l'oral s'appliquent probablement au décodage, selon des facteurs internes à la phonologie de la langue : le décodeur choisira une réponse déterminée

par ces contraintes. Le temps de traitement nécessaire au décodage étant celui du temps de fixation de l'oeil entre deux saccades, dès que la réponse (sub)vocalisée s'exprime l'oeil peut être libéré pour chercher la suite. Il peut aussi faire des régressions (par exemple lors d'une double difficulté : attaque branchante + digramme dans "fraîche" lu d'abord "far..." puis "fèr", puis "frèch". La suite f + r est plus difficile que f + a, il y a élision de "r", mais le digramme avec accent circonflexe posant problème, le "a" est dissocié de "aî" et l'oeil retourne lire le "r", puis décode enfin "aî", récupère le "r", et finalement réussit la suite C+C+V+C). Chaque graphème subvocalisé est ainsi un stimulus inconditionnel qui entraînera un autre phonème, même si ce dernier ne correspond pas au graphème décodé !

4.1.3. Décodage ou déchiffrage ?

4.1.4. Le décodage

Une chaîne orale ininterrompue dite dans une langue étrangère inconnue de l'interlocuteur n'aura aucune signification pour ce dernier. Ce sera une suite de sons sans sens. Il suffit pour s'en convaincre d'écouter une chanson dans une langue qu'on ne connaît pas.

La chaîne orale du discours (c'est-à-dire les paroles) est une expression orale selon un certain code. Ce code diffère selon les langues, mais l'interlocuteur doit toujours, pour comprendre ce message oral, le décoder en le décomposant en éléments plus petits et identifiables : les mots, porteurs de sens.

Si cette chaîne orale est dite dans sa langue, ou dans une langue qu'il comprend, l'interlocuteur pourra la décoder par un découpage. Ce découpage utilisera aussi bien la reconnaissance de mots, que le rythme de la phrase, essentiellement grâce à la syllabe qui se prête naturellement au découpage (il est plus facile par exemple d'isoler des syllabes que des

phonèmes lorsqu'on entend des sons, et des enfants arrivent facilement à scander un texte en tapant le rythme sur chaque syllabe).

Le décodage est donc une notion qu'il faut réserver à l'oral. Pour l'écrit, on parlera de déchiffrage qui, lui, à l'inverse, ira des éléments les plus petits (les lettres) vers les éléments les plus complexes (syllabes, mots, phrases).

4.1.5. Le déchiffrage

Habituellement le déchiffrage est perçu comme l'attitude quelque peu hésitante d'un lecteur confronté à un texte écrit, et pour lequel la lecture s'effectue péniblement lettre après lettre.

Cette connotation négative a joué pour beaucoup dans le rejet du déchiffrage de certaines méthodes de lecture.

En fait, le déchiffrage correspond, pour l'écrit, à ce qu'est le décodage pour l'oral.

À l'inverse du décodage qui va découper la chaîne orale entendue en unités plus petites, le déchiffrage va des unités vues les plus petites (les lettres) vers les unités de plus en plus complexes (la syllabe, le morphème, le mot, la phrase) jusqu'à ce que le sens soit peu à peu découvert.

4.1.6. Du décodage dans le déchiffrage !

Cependant, et c'est un point clé de mon analyse, le déchiffrage implique une (sub) vocalisation, et ces chaînes sonores sont alors autodécodées par le lecteur. À ce moment là, le déchiffrage oralisé relève donc du décodage ! Ce décodage (ou "déchiffrage oralisé") engendre les mêmes contraintes que l'acte de parler.

Des articulations successives peuvent engendrer des oppositions articulatoires qui nécessitent des efforts que le décodeur va tâcher de minimiser : ainsi le pseudo-mot "bapa" est lu d'abord

"baba" puis rectifié en "bapa" (et non "papa"), en raison du voisement habituel d'une consonne sourde placée entre deux voyelles. Le décodage correct du premier graphème montre bien que "b" est connu et différencié de "p" au niveau visuel. Quand il y a confusion visuelle (en attaque de syllabe isolée), il peut y avoir une implication phonologique dans le décodage, selon l'environnement .

En quelque sorte, il y a du son dans l'image.

En effet, j'ai observé également : "don" lu correctement, mais "bi" lu [di], et, lorsque le mot complet fut présenté en entier et dans le bon ordre ("bidon"), le décodeur produisit [dibon]. Le "don" a été transformé en "bon", la suite d + b étant plus économique au niveau articulatoire, alors que "don" avait dans un premier temps été bien décodé. L'utilisation de coupes syllabiques permettrait de vérifier si, connaissant les lettres présentées en attaque de syllabes isolées, le décodeur ferait quand même une erreur pour les mêmes attaques, mais dans un mot plurisyllabique. Cela confirmerait bien l'hypothèse d'une rection idiophonologique.

Mais cela ne suffit pas à détecter chez l'enfant quel niveau de conscience phonique il utilise pour décoder : la conscience syllabique, ou la conscience phonémique ? La rection idiophonologique pourrait porter sur l'une ou l'autre, selon le niveau de performance du lecteur, et selon le mot déchiffré, car le décodeur n'aura pas la même contrainte et donc une production autre (par exemple, pour une conscience syllabique, le déplacement d'un phonème peut être entraîné par une nécessité de coupe syllabique acceptable phonologiquement).

Ce point sera développé par l'étude (voir plus loin) de la notion de déchiffrage

Je parlerai de rections idiophonologiques, dues aux imprégnations phonologiques de chaque enfant, parallèlement à la théorie des contraintes et de stratégies de réparation de Carole

Paradis, pour qui également, mais lors d'échanges verbaux et non en lecture, les contraintes et réparations peuvent s'appliquer aux processus de décodage.

4.2. Schèmes phonotactiques

4.2.1. Constat

À l'oral, 80% des syllabes en français sont ouvertes (c'est-à-dire terminées par une voyelle, comme " p**a** / r**i** ") ;

20% des syllabes sont fermées (c'est-à-dire terminées par une consonne, comme " pa**r** / mi ").

4.2.2. Hypothèse

L'apprenti-lecteur, lorsqu'il est au stade pré-orthographique du décodage, commet des erreurs dans l'environnement de syllabe fermée CVC (consonne+voyelle+consonne), et tend à reproduire par hypercorrection la structure CV de la syllabe ouverte (consonne+voyelle, ou encore consonne+consonne+voyelle).

4.2.3. Structures profondes

Si les résultats montrent une très nette tendance à commettre une erreur dans les syllabes fermées en les transformant en syllabes ouvertes (par exemple si un mot comme "quatorze" est lu "quatroze"), alors on pourra conclure qu'il s'agit non pas de hasard mais d'une rection idiophonologique dans l'acte de décodage, dont le mécanisme sous-jacent serait dépendant de schèmes phonotactiques.

Nota : Mon hypothèse relève de l'existence de " schèmes phonotactiques ", et à ce titre j'ai traité les données avec des matrices, quelle que soit la consonne de substitution. Par exemple, que "balsamine" soit lu "blasamine" ou "brasamine", le mécanisme est le même, et comptera pour un point dans le trigramme "BAL". Ce n'est donc pas la consonne qui est importante, mais le fait que cela soit une consonne.

4.2.4. Trigrammes

J’ai répertorié pour ce test un ensemble de trigrammes sur le critère de la légalité (ou prononçabilité). Chacun d’eux peut donc être lu avec ou sans permutation des consonnes (“bal” ou “bla” par exemple) :

BAL, BEL, BIL, BOL, BUL ; BAR, BER, BIR, BOR, BUR ; CAL, CEL, CIL, COL, CUL ; CAR, CER, CIR, COR, CUR ; CAR, DER, DIR, DOR, DUR ; FAL, FEL, FIL, FOL, FUL ; FAR, FER, FIR, FOR, FUR ; GAL, GEL, GIL, GOL, GUL ; GAR, GER, GIR, GOR, GUR ; PAL, PEL, PIL, POL, PUL ; PAR, PER, PIR, POR, PUR ; PAS, PES, PIS, POS, PUS ; SAL, SEL, SIL, SOL, SUL ; SAR, SER, SIR, SOR, SUR ; TAR, TER, TIR, TOR, TUR ; VAL, VEL, VIL, VOL, VUL ; VAR, VER, VIR, VOR, VUR.

Chacun de ces trigrammes étaient employé dans 20 mots ou pseudo-mots différents, ce qui représentait 17 matrices (Consonne+Voyelle+Consonne), composées chacune de 5 rimes (a, e, i, o, u), pour un total de 1700 mots à lire par élève, en plusieurs séances. On peut consulter sur le site la liste exhaustive des items, ainsi que les mots et pseudomots sélectionnés.

4.2.5. Analyse du Test01

L'analyse a été effectuée en répertoriant toutes les erreurs rencontrées, quelles qu'elles soient, puis à retenir celles qui sont en adéquation avec le sujet de cette étude (autrement dit les erreurs qui portent directement sur les trigrammes considérés).

Ensuite, chacune de ces adéquation a été observée : s'il s'agit d'une erreur du type CV+C > CVC (une syllabe ouverte lue comme une syllabe fermée) ou d'une erreur du type CVC > CV+C ou CCV(une syllabe fermée lue comme une syllabe ouverte), un point est attribué à la catégorie où s'effectue l'erreur.

L'ensemble de ce comptage est repris en sous-totaux, avec pourcentage.

J’ai répertorié 7482 erreurs (consultables en intégralité sur le site), dont 18,82% en adéquation avec les trigrammes.

Dans 71,22 % des cas, on a effectivement une hypercorrection qui réduit une syllabe fermée à une syllabe ouverte.

Sur les 17 matrices (les 17 types de trigrammes), une seule possède un résultat opposé à mon hypothèse. Il s'agit de la matrice SAR, pour laquelle la structure en syllabe fermée est plus élevée. Il faudra tenter de comprendre ce contre-exemple... mais peut-être que la difficulté articulatoire de SRA, SRE, SRI, SRO, SRU supérieure à SAR, SER, SIR, SOR, SUR explique tout simplement cette exception. La loi de l’économie phonétique fonctionne toujours !

En structure sous-jacente, on a donc le décodage suivant :

CVC > CV+C

ou bien

CVC > CCV

Tant que la perception des phonèmes (c'est-à-dire le déchiffrage subvocalisé) porte sur des suites prévisibles, l'erreur est rare. Lorsqu'il y a promiscuité phonétique, l'interférence ‘phonème prévu' vs ‘phonème phonologiquement induit' intervient et provoque l'erreur de déchiffrage.

On peut donc considérer tout à fait plausible l’existence d’une contrainte phonotactique très forte en structure sous-jacente lors de l’acte de déchiffrage chez l’enfant apprenti-lecteur (au stage pré-orthographique).

Cela peut se résumer par la matrice suivante :

C1+V1+C2 > C1+C2+V1

(nota : C1 = consonne en attaque de syllabe ; C2 = consonne en coda ; V1 = rime de syllabe).

4.3. TEST02 : pour un échantillonnage des erreurs de lecture

Le Test02, constitué de quinze phrases indépendantes à partir du vocabulaire de base (voir plus loin), a pour but de faire une tentative d'inventaire des erreurs de lecture, afin d'en déduire des règles de fonctionnement.

Ces phrases ont été conçues pour offrir un panel presque exhaustif des différentes graphies rencontrées en français.

En même temps, le système phonologique du français est résumé dans l'ensemble du test. À la lecture de ces phrases-test, le lecteur parcourt sans le savoir le système graphique et phonologique du français. Il ne reste plus alors qu'à inventorier les erreurs et à les classer pour en extraire une taxinomie.

4.3.1. Les quinze phrases du test02 :

1) jeanne et jean étaient deux personnes.

2) cette jeune fille était dans une chambre quand enfin le premier soleil entra.

3) le petit christophe, capitaine vers reims, à dix-huit kilomètres de sa femme, était vraiment simple.

4) le wagon était immobile encore un grand moment.

5) il y avait du brouillard : au regard on ne voyait presque rien.

6) chez le jockey l'expression asseyez-vous est un travail.

7) il a besoin de deux feuilles jaunes pour quinze phrases.

8) dans toute sa conversation il y a par exemple une question, un comment, un point, une expression.

9) dans beaucoup de pays la faim chez les moines reste une autre manière de conscience.

10) son coeur aime me faire un signe avec ses yeux.

11) vous avez comme moi appuyé votre nez et les mains.

12) leur soeur ayant plein de peine, ils n'étaient jamais au premier effet.

13) cette chose, la sienne, bien assez de gens que je connais de visage, presque dix, ne l'aimaient pas.

14) son voeu était d'avoir des ailes.

15) tu es ce temps un exemple.

4.3.2. Analyse

Le nombre d’erreurs répertoriées s’élève à 640, ce qui est suffisant pour avoir une base de données significative.

À partir de ces erreurs, que l’on peut consulter sur le site, j’ai observé des recoupements, à partir desquels un éclairage plus complet des mécanismes de lecture est apparu.

L'analyse du test02 (les quinze phrases) m'a permis de rendre compte que la notion de déchiffrage est plus complexe qu'il n'y paraît, et qu'il y aurait deux types de déchiffrages :

un déchiffrage systématique (au sens de “item après item”) : la **fluidité** (recherche de l'élément suivant le plus attractif au niveau perceptif, ce que devrait corroborer le test03 avec le second élément large). C’est la phase primitive de la lecture ;

un déchiffrage systémique : d'abord de **proximité** (recherche de l'unité syllabique), puis de **forme** (recherche d'une unité orthographique). C’est la phase générative de la lecture.

4.3.3. Synthèse

Il y aurait donc, pour résumer, des erreurs d'origine phonologique, et des erreurs d'origine visuo-perceptive, que je propose de classer comme ceci :

1) erreurs phonologiques :

a) par commutation (/Z/ et /g/, comme "gens" et "gan") ;

b) par permutation ("par" et "pra", voir les syllabes fermées vs syllabes ouvertes) ;

2) erreurs visuo-perceptives :

a) déchiffrage systématique : relevant d’un principe de fluidité (second élément large)

b) déchiffrage systémique : relevant de deux principes, proximité et forme (unités linguistiques).

Le modèle demande à être affiné, mais il me semble être valide. Il permettrait de mieux comprendre les erreurs de lecture, et par là les mécanismes d'apprentissage.

4.3.4. Classement des erreurs

Le classement des erreurs a été fait sous trois colonnes, dont voici un résumé :

déchiffrage systématique (fluidité)	**déchiffrage systémique (proximité)**	**déchiffrage systémique (forme)**
chambre > cha_me_be_ré	aimaient > aimé_an	ailes > ai_lé
vous > vousse	reims > rei_me	dix-huit > dichu
quand > quande	immobile > in_mobile	deux > dans
quand > quena		moines > moi_né
regard > regarde		point > pain
simple > si_me_ple		signe > singe
temps > tempe		avez > avec
encore > ne_sore		
connais > comais		
soeur > sor		
enfin > enfine		
faim > fai_me		
sienne > si_ne		
exemple > ex_pemple		
jean > je_an		
reims > re_mi		

aime > a_mi avoir > a_vor capitaine > capita_ne		

Dans la première colonne, la lecture est systématique, et les erreurs rencontrées relèvent de l'ignorance d'un digramme ("ai" lu "a") ou d'une permutation ("encore" lu "ne_sore", "c" lu "s" n'étant qu'une symbolisation imparfaite dans la correspondance entre signifiant graphique et signifiant sonore). Ces permutations ne sont pas ici d'origine phonologique (comme c'est le cas pour les syllabes ouvertes et les syllabes fermées, processus étudié avec le Test01), mais visuo-perceptive. Il semble que le gramme suivant interfère dans la reconnaissance de la chaîne des caractères, en donnant une priorité à celui qui a un poids visuel supérieur. Autre exemple, "quand" est lu "que_na", le gramme "n" ayant un poids visuel supérieur à "a" : a+n devient n+a.
Dans la deuxième colonne, la lecture est systémique, et passe du gramme au polygramme. Le lecteur cherche une proximité visuelle qui permettent une syllabation facile. "reims" est décomposé en "rei_me", "faim" est lu "fai_me". Ce type d'erreur par proximité entraîne un mauvais découpage du mot, mais pas forcément une erreur de correspondance entre un signifié graphique et sa lecture (on a bien "ai" dans "aim" de "faim").
Dans la troisième colonne, la lecture est également systémique, mais le lecteur cherche une forme connue, ce qui engendre une erreur totale : "es", digramme très fréquent ("les, mes, tes, ses, des"), prend la valeur phonique "é" ([e]) comme pour "moines" lu "moi_né" ; "avez" est reconnu comme la forme de "avec", et confondu comme le sont "dans" et "deux".
On voit bien que la perception visuelle a un poids important dans les erreurs de lecture. Elle constitue, avec les schèmes phonotactiques, l'ensemble des facteurs d'erreur.

4.4. TEST03

Le but du test03 est de vérifier l'existence d'une loi du second élément large (S.E.L.), autrement dit si les erreurs de permutation observées sur le terrain serait le résultat d'une attraction visuelle, d'un "poids visuel".

4.4.1. Poids visuel

Définition : le Poids Visuel ,noté PV, est la surface d'un gramme par rapport à un "oeil" de référence, la lettre "x", à laquelle j'affecte la valeur 1).

4.4.2. Hypothèses

4.4.2.1. Première hypothèse

Si un gramme (par exemple "m") a un poids visuel supérieur à celui d'un autre gramme (par exemple "s"), c'est-à-dire si PVm > PVs (le Poids Visuel de "m" est plus grand que le Poids Visuel de "s")

alors

l'attraction visuelle de "m" sera supérieure à celle de "s", et la suite "s" .... "m" sera permutée en "m"...."s".

Par exemple, le (pseudo)mot "somo" devrait être lu "moso", mais "moso" ne sera pas lu "somo".

Le nombre de fois où "somo" sera lu devrait provoquer un nombre d'erreurs bien supérieur à celui de la lecture du mot "moso". On devrait s'attendre à un alignement (toutes erreurs confondues pour "moso" et "somo") sur une lecture majoritaire "moso".

4.4.2.2. Deuxième hypothèse

Certaines chaînes de caractères possèdent deux grammes de poids visuel équivalent. C'est le cas par exemple de "g" et "b", de "v" et "r", etc.

Si la loi du SEL est valide, le nombre de permutations entre "g" et "b" d'une part, ou entre "v" et "r" d'autre part, devrait être sensiblement identique.

Les pseudo-mots "gaba" et "baga" seront lus correctement, ou, s'ils sont lus de façon incorrecte, les permutations g / b seront d'un nombre équivalent pour les deux mots (en d'autres termes, le nombre de fois où "gaba" sera lu "baga" devrait s'approcher du nombre de fois où "baga" sera lu "gaba").

4.4.2.3. Troisième hypothèse (falsification)

Si la loi du SEL devait s'avérer effective, on devrait pouvoir contrôler ses conséquences en modifiant, à l'intérieur des pseudo-mots, le corps des grammes, c'est-à-dire en échangeant par exemple une police de caractère par une autre, ou encore en substituant l'attribut gras à l'attribut normal, afin d'affecter au gramme considéré un poids visuel différent. La lecture de "moso" devrait être différente selon qu'on a "moso" ou "mo**s**o" (ici, j'ai ajouté l'attribut gras et augmenté d'un degré la taille de "s", entraînant ainsi un poids visuel supérieur à "m").

Cette falsification devrait donc amener le lecteur à commettre plutôt l'erreur "**s**omo", ce qui confirmerait la loi du second élément large par contrario. Cette étape est encore à faire sur la même population d'élèves que d'habitude, pour ne pas ajouter des paramètres non comparables.

L'observation des erreurs de lecture du Test02 a permis de montrer qu'il y a différents types de déchiffrage. Le premier, le déchiffrage systématique, semble être conditionné par une recherche de fluidité (un item après l'autre, tant qu'il n'y a pas d'interférences).

Cette fluidité peut être perturbée par la présence d'un élément plus éloigné mais plus attractif. Cette notion d'attraction, toujours selon les faits linguistiques observés, semble dépendre d'un "poids visuel" de chaque gramme.

4.5. Mesure du poids visuel

Cette notion de "poids visuel" correspond à la surface totale d'un gramme. Mais il fallait un outil pour mesurer cela.

Pour cela, j'ai dû calculer la surface (que je préfère pour des raisons théoriques appeler "poids visuel") de chaque gramme (on reste dans la même isotopie sémantique !) en imprimant sur papier millimétré chaque lettre de l'alphabet (police time news roman) en taille ... 200 ! avec l'atttribut contour.

Ensuite, j'ai compté un à un chaque millimètre carré...

Dans un premier temps, on a donc la surface brute (par exemple 510 mm2 pour le gramme "b", 365 mm2 pour " f ", etc.). Mais comme cette surface brute n'a pas de valeur représentative en soi, j'ai pris l' "oeil" de référence des typographes (à savoir la lettre "x") à laquelle j'ai donné la valeur étalon 1 ("x" fait 310 mm2 en time news roman taille 200). Avec une banale règle de trois, j'ai pu enfin classer par ordre décroissant les grammes selon leur poids visuel.

4.5.1. Classement des grammes par leur poids visuel

On a ainsi par ordre de poids visuel décroissant :

gramme	surface brute	poids visuel relatif
m	669	2,16
d	535	1,73
g	516	1,66
b	510	1,65
h	510	1,65
p	499	1,61
q	490	1,58
w	484	1,56
û	461	1,49
k	449	1,45
ô	432	1,39
ù	431	1,39
n	424	1,37
ê	406	1,31
ë	398	1,28
a	386	1,25
u	383	1,24
é	376	1,21
è	376	1,21
f	365	1,18
ç	362	1,17
o	354	1,14
y	354	1,14
s	334	1,08
e	328	1,06
x	310	1,00
z	301	0,97
c	300	0,97
j	296	0,95
l	263	0,85
t	260	0,84
ï	256	0,83
v	254	0,82
r	237	0,76

i	221	0,71

Ce résultat m'a donc permis de concevoir le test03, qui s'inscrit dans la première des trois étapes du déchiffrage (qui sont, d'après mes analyses : la fluidité dans le déchiffrage systématique, puis la proximité et la forme dans le déchiffrage systémique).

D'aspect visuel, cette étape recherche la fluidité (c'est le rôle de la vision, influencée par l'attraction visuelle des items graphiques).

L'hypothèse du Test03 (embrayée par des observations sur le terrain) pourrait intéresser plus particulièrement les dyslexiques. Un pseudo-mot comme "damu" aura plus de risques d'être lu "madu" que l'inverse, à cause du poids visuel (et attractif pour l'oeil) de "m" bien supérieur à "d" ("m" : 2,15 ; "d" : 1,75).

4.6. Élaboration du Test03

Pour ce test03, j'ai fabriqué des pseudo-mots de façon à éviter certains pièges (par exemple une proximité sémantique, le pseudo-mot devenant alors une amorce, ce qui fausserait le test).

Voici le test03 :

Le test03 comporte une seule série de pseudo-mots, à lire en colonne. Cette série est subdivisée en quatre groupes de quatre pseudo-mots, et d' un pseudo-mot isolé.

pseudo-mots
mano
mosa
muga
madu

gofa
gazo
guba
gadu
foca
fazi
fusu
fiso
suti
sala
voro
tava
mili
namo
soma
guma
damu
foga
zago
buga
dagu
cofa

zafi
sufu
sifo
tusi
lasa
rovo
vata
limi

4.7. Analyse du Test03

le Test03, dont l'objectif est de savoir s'il existe une attraction visuelle (proposition dans ce cas là d'une "loi" du second élément large), s'avère trop simple. Sur un total de 4930 mots lus (34 mots lus par 145 élèves de niveau CE1 en fin de premier trimestre), je n'ai relevé qu'une vingtaine d' erreurs en corrélation avec l'objectif, c'est-à-dire relevant de la permutation. Ce nombre peu élevé impliquerait de refaire le test en le complexifiant, ou en le proposant à des élèves de cours préparatoire en fin de deuxième trimestre.

Total des erreurs de permutation : 20

Nombre d'erreurs considérées comme neutres : 3 (15 %) ; (erreurs dont l'écart entre le premier gramme et le second est inférieur à 0,1, par exemple dans l'item "vata", "v" a pour valeur absolue 0,82, et "t" a pour valeur absolue "0,84". Bien que "vata" ait été lu comme prévu "tava", je n'ai pas comptabilisé cette erreur, même si elle va dans le sens de l'objectif du Test03, car l'écart en poids visuel entre "v" et "t", égal à 0.02, est insignifiant) ;

Nombre d'erreurs relevant de l'attraction visuelle large (ou SEL) : 13 (76,47 %) ;

Nombre d'erreurs relevant de l'attraction visuelle verticale (ou SEV) : 2 (11,76 %) ;

Nombre d'erreurs qui s'opposent au Test03 : 2 (11,76%).

En résumé, il y a erreur due à une attraction visuelle dans 88,23 % des cas, et il y a erreur due à une cause inconnue dans 11,76 % des cas.

Encore une fois, le faible nombre d'erreurs en corrélation avec cette recherche est trop faible pour en tirer une conclusion définitive, mais on peut noter tout de même que les résultats obtenus vont très nettement dans le sens d'une attraction visuelle, cette attraction visuelle reposant sur deux facteurs : le second élément large ou SEL et le second élément vertical ou SEV (rôle de l'empattement, ou encombrement vertical).

À noter également : les deux erreurs qui s'opposent au Test03 ne représentent qu'une seule occurrence chacune ("mano" lu contre toute attente "namo", et "gazo" lu "zago") alors que leur performance prévisible et observée est de six pour "namo" lu "mano", et de deux pour "zago" lu "gazo"). Ces deux erreurs apparaissent donc isolées, contrairement à celles qui suivent la théorie d'une attraction visuelle. On touche peut-être bien là à un mécanisme réel.

4.7.1. Classement des erreurs

Voici les erreurs relevées, avec dans la première colonne le mot à l'origine de l'erreur, dans la deuxième colonne l'erreur produite, dans les troisième quatrième cinquième et sixième colonnes la valeur 0 pour Non et 1 pour Oui pour indiquer de quel type d'erreur il s'agit (SEL, SEV, ou autre, c'est-à-dire opposé), dans la septième colonne le poids visuel (PV) de l'attaque de la première syllabe suivi du poids visuel de l'attaque de la seconde syllabe pour le mot à lire, dans la huitième colonne le poids visuel de l'attaque de la première syllabe suivi du poids visuel de l'attaque de la seconde syllabe pour l'erreur effective, et enfin dans la dernière

colonne l'écart entre ces deux valeurs de la septième colonne. Cet écart peut-être positif ou négatif. La loi du SEL implique que si l'écart est négatif, l'erreur de permutation est prévisible (par exemple namo ou mano).

à lire	lu	SEL	non SEL	SEV	PROCHE	PV donné	PV lu	écart
mano	noma	0	1	0	opposé	2,16-1,37	1,37-2,16	0,79
gazo	zago	0	1	0	opposé	1,66-0,97	0,97-1,66	0,69
suti	tuti	0	1	1		1,08-0,84	0,84-1,08	0,24
suti	tussu	0	1	1		1,08-0,84	0,84-1,08	0,24
dagu	gagu	proche	proche	proche	1	1,73-1,66	1,66	0,07
guba	buga	proche	proche	proche	1	1,66-1,65	1,65-1,66	0,01
vata	tava	proche	proche	proche	1	0,82-0,84	0,84-0,82	-0,02
sufu	fuga	1	0	0		1,08-1,18	1,18	-0,1
cofa	foga	1	0	0		0,97-1,18	1,18	-0,21
lasa	sala	1	0	0		0,85-1,08	1,08-0,85	-0,23
lasa	sala	1	0	0		0,85-1,08	1,08-0,85	-0,23
zago	gazo	1	0	0		0,97-1,66	1,66-0,97	-0,69
zago	gazo	1	0	0		0,97-1,66	1,66-0,97	-0,69
namo	mano	1	0	0		1,37-2,16	2,16-1,37	-0,79
namo	mano	1	0	0		1,37-2,16	2,16-1,37	-0,79
namo	mano	1	0	0		1,37-2,16	2,16-1,37	-0,79
namo	mano	1	0	0		1,37-2,16	2,16-1,37	-0,79
namo	mano	1	0	0		1,37-2,16	2,16-1,37	-0,79
namo	mano	1	0	0		1,37-2,16	2,16-1,37	-0,79
limi	mili	1	0	0		0,85-2,16	2,16-0,85	-1,31
		13	**4**	**2**	**3**			

5. OUVERTURES THÉORIQUES

La continuité de ces Tests a été l'élaboration, et la mise à l'épreuve, de théories sur l'apprentissage de la lecture ("schèmes phonotactiques", "loi du second élément large", "étapes du déchiffrage"), pour une compréhension toujours plus fine des processus cognitifs.

5.1. Pour une proposition d’un modèle de lecture

5.1.1. Supports somatiques

Les supports somatiques dans les actes cognitifs sont solidement établis, grâce aux études en résonnance magnétique fonctionnelle. Les localisations cérébrales sont également, et clairement, identifiées. Si l'aire de Broca est une des aires somato-sensorielles et motrices qui permet de gérer les bases de la parole (reconnaissance des phonèmes, activation morpho-phonatoire), l'aire de Wernicke gère quant à elle un stade plus évolué de la parole humaine, associant signifiants et signifiés.

Mais ces quasi-certitudes, aussi éclairantes soient-elles, souffrent d'être statiques pour le linguiste.

En effet, localiser des zones du cortex cérébral et découvrir leur rôle principal mais non exclusif (on sait que l'endommagement partiel d'une zone entraîne des stratégies compensatoires par d'autres zones normalement affectées à d'autres tâches) ne suffit pas à expliquer le fonctionnement dynamique des processus de lecture.

5.1.2. Pluridisciplinarité

Il faut donc, pour tenter de comprendre ce qui se passe, proposer un modèle unifié des processus de décodage chez l'apprenti-lecteur. La conséquence de ce cadre théorique est d'accepter l'apport de différentes disciplines.

L’essai de modèle que je propose va dans ce sens, mais, parce que le sujet d'étude est le langage, l'aspect linguistique est ici privilégié.

La littérature actuelle propose une spécialisation des hémisphères : le droit reconnaît les mots globaux, et le gauche effectue un traitement analytique.

Là encore, cela ne suffit pas à éclairer les processus en jeu : si l'activation de l'hémisphère droit est plus intense à la présentation d'un logographe, ou si l'activation de l'hémisphère gauche est plus intense dans le déchiffrement d'un pseudomot, on a affaire à des constats, effectués sans le souci de conception générale de la lecture, sans recherche d'une théorie globale de l'apprentissage de la lecture. Et c'est normal, puisque le problème récurrent "étude sur l'enseignement et/ou sur l'apprentissage ?" est une synthèse qui incombe à la linguistique, grâce aux recherches extra-linguistiques.

En effet, si la lecture est un processus spécifiquement langagier, elle fait appel à des processus cognitifs plus généraux : la perception visuelle et auditive, la mémoire de travail et la mémoire à long terme. Chacun de ces processus est étudié par une ou plusieurs disciplines relevant de la médecine et des sciences de la cognition. On voit que la linguistique a tout à gagner de l'apport de ces autres disciplines, sans pour autant perdre de son indépendance méthodologique.

On peut déjà ici distinguer deux actes, la lecture, et le discours, et concevoir pour chacun un ordre d'intervention de ces deux hémisphères. Pour le discours, c'est d'abord l'aire de Wernicke qui intervient (le sens est en attente de transfert), puis l'aire de Broca (le sens trouve son moyen d'expression). Pour la lecture, si l'on s'accorde sur le principe qu'un lecteur débutant oralise (intérieurement ou non), l'ordre d'intervention est différent : l'aire de Broca gère l'information phonologique, avant de l'envoyer à l'aire de Wernicke qui lui associera du sens. Cela sera repris plus loin, car cela n'est toujours pas suffisant.

5.2. Étapes de l’acte de lecture primaire

La notion de lecture est prise ici dans son acception stricte. Elle s’arrête, par l’objet même de cette étude, à la reconnaissance des signifiés. C’est une lecture de bas niveau. La lecture secondaire, plus élaborée sémantiquement, concerne le sens littéral (ce que dit un texte) et étendu (présuppositions, connotations, interprétations, plaisir, etc.).

Mais cela n'est toujours pas suffisant. Puisqu'il s'agit de lecture, le point de départ est un signifiant visuel, mot ou texte. Dès ce point, l'acte de lecture se met en place, selon trois étapes qui, si elles s'approchent des étapes "logographique, phonologique, orthographique" devenues classiques, en diffèrent dans leur formulation et les conséquences théoriques qui en résultent.

1) **une étape biologiquement nécessaire** : un traitement de surface en deux dimensions permet de voir des traces graphiques (les grammes) ;

2) **une étape culturellement associative** : à un gramme clairement identifié est associé un son pré-défini socialement. L'acquisition relève de la mémorisation et de la tâche répétitive.

3) **une étape cognitivement régulatrice** : les sons fournis à l'étape précédente doivent être agencés selon deux contraintes. La première est linéaire (c'est la suite des lettres dans l'ordre où elles sont vues), la seconde est paradigmatique (c'est la priorité ou probabilité d'un son sur un autre dans un environnement donné).

Ce modèle, présenté en ces termes, permet la prise en compte de processus quasi-invisibles autrement. En effet, se poser la question "comment fonctionne l'acte de lecture ?" cache d'emblée le traitement de l'erreur qui, seule, peut nous permettre de comprendre les

mécanismes. Il sera plus fertile de se demander "comment fonctionnent les erreurs de lecture ?".

Pour avoir un embryon de réponse, voici la suite logique de cette analyse, sur les trois étapes.

5.2.1. L'étape biologiquement nécessaire

Cette étape permet la reconnaissance des formes. Ici, il s'agit, avec des tests linguistiques, si possible consolidés par une observation des fixations oculaires, de littéralement voir pourquoi une erreur se produit.

Il existe une loi du second élément lourd qui, à l'oral, impose de toujours mettre en seconde position l'élément phonétiquement le plus lourd (on peu commencer à dire "tac-tic, tac-tic", on finira toujours par dire "tic-tac, tic-tac" car [a] est plus lourd que [i] nécessitant un plus grand effort articulatoire).

Pour la lecture, je propose une loi du second élément large pour expliquer que le regard cherche naturellement les letttres les plus larges (ou les plus hautes), c'est-à-dire celles qui ont une chasse (une largeur) ou un poids visuel plus importants. Si, dans un environnement donné, une lettre qui chasse beaucoup est située après une autre plus étroite, et qu'elle ne fait pas partie de son groupe grammatique, elle perturbe le déchiffrage et fait commettre l'erreur. La priorité visuelle d'une lettre sur une autre vaut si celle-ci est beaucoup plus large ou haute que celle-là.

5.2.2. L'étape culturellement associative

Le signe très arbitraire d'une langue alphabétique nécessite un apprentissage mnémonique, où un phonème est attribué à un gramme (ou à un polygramme, par exemple /e/ pour "ez", "et", "es"). Il s'agit de mémoire à court terme (le lecteur a retenu l'appariement inculqué phonème /

(poly)gramme) et de mémoire immédiate (le lecteur retient cet appariement au moment même où il lit pour le combiner aux appariements suivants).

Cette symbolisation a fait basculer le processus de lecture dans l'hémisphère gauche, plus apte aux traitements analytiques. Les erreurs de lecture relèvent à ce moment-là de la mémoire. L'empan mnémonique diffère d'un enfant à un autre, et la difficulté dans cette étape est double : le choix d'un paradigme associatif (à une entrée visuelle il faut accoler une valeur phonétique dans un réservoir de paires gramme / phonème apprises par tâches répétitives relevant du "par-coeur"), et le maintien, en mémoire, de ce choix. A cette étape, une didactique de la phonologie est cruciale pour que l'apprenti-lecteur accède à l'étape suivante.

5.2.3. L'étape cognitivement régulatrice

Cette étape, pour laquelle j'ai conçu les test01 et test02, est la conséquence de ce qui est vu, de ce qui est su, de ce qui est dit. En d'autres termes, on ne lit que ce qu'on sait dire, et on ne dit que ce qu'on a entendu. Le lecteur lit ce qu'il sait lire, mais ce qu'il sait lire est le résultat de ce qu'il sait dire. L'erreur viendrait, à ce stade, des conflits phonologiques sous-jacents, d'une guerre entre Langue et parole ! Pour savoir lire, le futur lecteur doit devenir l'arbitre des conflits (dont les conséquences sont les erreurs des trois étapes) sur un terrain qui le mène du visuel vers l'oral, en utilisant plusieurs outils cérébraux.

5.3. Le décryptage

Jusque là, le sens est absent (du moins sa reconstitution par le déchiffrage, car l'enfant peut très bien savoir de quoi parle un texte, ou ce que peut désigner le mot).

Lorsque l'étape est menée à son terme, l'aire de Wernicke intervient : puisque l'enfant sait déjà parler, dès qu'une quantité minimale mais suffisante de déchiffrage lui permet de décoder un

ensemble de sons proches donnant une chaîne (sub)sonore similaire à un morphème connu, le sens de ce morphème est associé à cette chaîne (sub)sonore ("sub" parce qu'il s'agit, en lecture silencieuse, de parole intériorisée), et il y a accès au sens. C’est le décryptage.

5.3.1. Triptyque de la lecture primaire

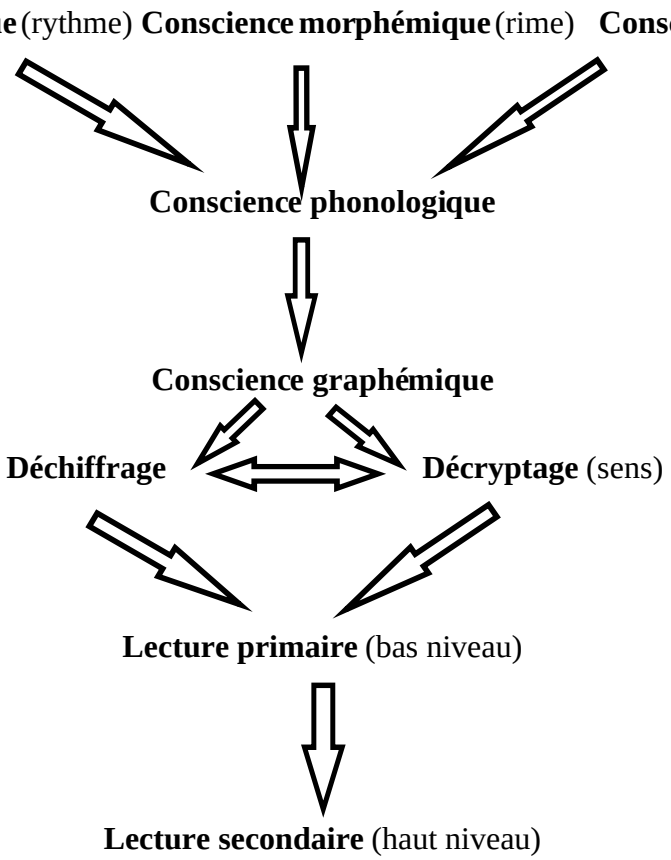
À l'habituelle représentation de la lecture en décodage puis sens, je proposerai un triptyque :

1) déchiffrage (i.e. grapho-phonétisme puis conscience phonologique) ;

2) décodage (i.e. conscience phonologique puis hologrammisme) ;

3) décryptage (i.e. décodage puis sémantisation).

5.3.2. Schéma de la lecture primaire



5.4. L’empan mnémonique

L'entraînement à la lecture permet alors, pour certaines chaînes grammatiques, une économie de décodage. Cette performance relève du développement de l'empan mnémonique. Au lieu d'associer une paire gramme / phonème dans un paradigme unaire (UN gramme ou polygramme pour UN phonème), le lecteur utilise un paradigme pluriel (plusieurs grammes ou polygrammes avec plusieurs phonèmes) pour aller jusqu'au mot entier, toujours dans le sens du plus simple vers le plus complexe. Ici, l'erreur est provoquée, comme à l'étape 1, par des similarités visuelles ('dent', 'lent') et comme à l'étape 2, par des faiblesses mnémoniques. On le voit donc, à chaque étape son type d'erreurs, à chaque type d'erreurs une cause.

5.5. Avantages du modèle

Ce modèle, s'il est probablement lacunaire, offre plusieurs avantages :

a) il réunit dynamiquement les processus de lecture chez l'apprenti-lecteur ;

b) il spécifie ces processus étape par étape ;

c) il met en évidence le type d'erreur propre à chaque étape ;

d) il pointe la cause pour chaque type d'erreurs ;

e) il montre que le terme "lecture" est une notion qui recoupe plusieurs activités orientées d'abord de l'immédiat vers l'analyse, de la surface à deux dimensions vers un traitement à une dimension supplémentaire (le temps, en raison de la mémoire), et d'un mécanisme cérébral vers un autre.

5.5.1. Modèle et modélisation

Ce modèle permettra, lorsqu'une base de données d'erreurs sera suffisante pour en extraire des lois prédictibles, d'envisager un simulateur de lecture, programme informatique volontairement générateur d'erreurs, dont la tâche sera de vérifier différentes hypothèses. Chaque sous-programme correspondra à une hypothèse (par exemple la loi du second élément large). L'ensemble des sous-programmes prédira les erreurs attendues. La vérification in fine ne nécessitera plus qu'un échantillon restreint de lecteurs testés. Mon idée est qu'il faudra bien un jour que nous utilisions des test réels uniquement pour finaliser des théories dont les résultats auront été simulés auparavant par le programme : le gain en temps et en énergie sera considérable, sans sacrifier à la rigueur scientifique.

6. MANUELS DE LECTURE

6.1. Outils d'analyse

Le travail réalisé sur le corpus de base servant de référence, il me fallait effectuer une étude statistique sur les manuels de lecture les plus utilisés en France.

Avec la pratique, je me suis doté d'un outil supplémentaire, le langage de programmation Perl*, à la fois plus facile et plus efficace pour l'étude des corpus textuels.

La réalisation de petits programmes très souples, mais très précis, a rendu possible la reconnaissance des environnements d'un mot ("mot1" + "mot2" + mot recherché + "mot3" + "mot4"), mais aussi, et de façon plus exacte qu'avec le Cobol (lourd et lent), le comptage des grammes, des polygrammes, du nombre de mots d'une lettre, de deux lettres, de trois lettres, etc., et même certaines fréquences phonétiques pour les progressions des manuels.

6.2. Étude comparative

De toutes ces données statistiques sur un corpus de base d'une part, et sur les manuels de lecture d'autre part, j'ai pu montrer des différences très nettes entre ce qui existe à travers les genres d' écrits qu'on peut rencontrer dans la vie, et les manuels de lecture qu'on ne rencontre normalement qu'en classe.

Les manuels de lecture ayant pour vocation d'apprendre à lire, il me semblait évident qu'ils devaient coller au plus près du fait linguistique incontournable : les fréquences du français écrit, mais aussi oral.

Le détail de cette étude est disponible intégralement sur le site, et comporte également l'analyse comparative des catégories grammaticales, comme suit :

occurrences, vocables, hapax, écarts pondérés

les coordinations : mais ou et donc or ni car

les interjections : bravo chut ha ah ho oh o ouf heu

les négations : aucun nulle jamais ni non pas personne rien ne

les pronoms personnels sujet : je tu il elle on nous vous elles ils

les déterminants: un une le la l' les

les pron pers et adj poss sing plur 1ère personne : je j' me m' moi mon ma mes nous notre nos

les pron pers et adj poss sing plur 2ème personne : tu te t' toi ton ta tes vous votre vos

les pron pers et adj poss sing plur 3ème personne : il elle ils elles leur leurs se soi lui eux son sa ses

les quantificateurs : rien nulle aucun assez autant beaucoup combien davantage mieux moins plusieurs seulement surtout tant tellement trop plus peu

les pronoms relatifs : lequel laquelle dont qui

les adverbes de lieu : aillleurs dedans dehors dessous dessus devant ici loin

les prépositions : afin au autour aux avant avec chez contre dans de depuis des devant du entre hors jusqu' par pendant pour sans selon sous sur suivant travers vers.

6.3. Les progressions dans les manuels de lecture : pour un principe de cohérence.

Le contenu de tout manuel doit proposer des progressions (puisqu'il s'agit d'un apprentissage dans le temps) dont le principe de base serait d'être cohérentes, c'est-à-dire fondées sur les faits de langue, écrite et orale. Nous nous intéressons ici aux promesses des manuels dont on se demandera, sur ce point et en dépit de qualités pédagogiques indéniables, si elles sont toujours tenues.

La cohérence, principe fondateur d'une progression adaptée sur cet aspect linguistique des manuels de lecture, repose principalement sur trois points.

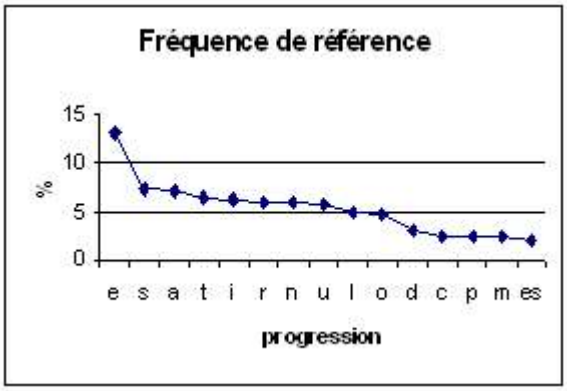
6.3.1. Les fréquences visuelles (les “lettres” les plus fréquentes).

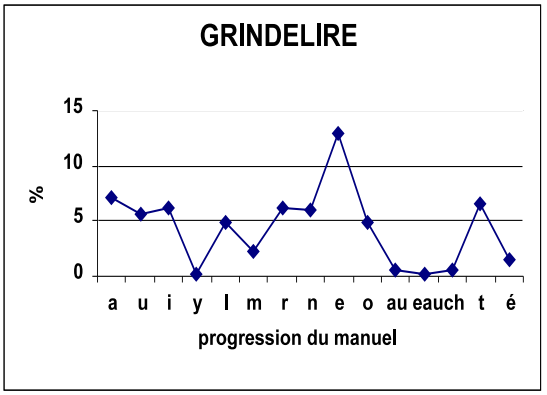
Le matériau de base avec lequel l'apprenti-lecteur doit se familiariser est l'ensemble des lettres du français (les grammes). Certaines ont beaucoup de chances d'apparaître (e, s, a, t, i, r, ...) d'autres beaucoup moins (â, k, V, w, ë, ...). De la même manière, les groupes de lettres (les polygrammes) n'ont pas les mêmes fréquences. Lorsqu'un enfant de C.P. découvre une ligne et demie de texte (une centaine de caractères), il rencontrera de façon quasi-certaine "en", "ai", "ou", "on ". Mais il lui faudra parcourir environ 400 lignes (plus de 25000 caractères !) pour trouver "ey", environ 1000 lignes (69000 caractères !) pour "cy", et 79000 caractères pour "eim"! Est-il donc bien utile de donner à apprendre des graphies rarissimes qui pourraient être apprises l'année suivante dans la continuité du cycle des apprentissages fondamentaux ? Il s'agit là d'un fait de langue qu'aucun manuel de C.P. ne doit ignorer. Certains d'ailleurs le rappellent clairement (les auteurs de Ratus, dans leur introduction, évoquent cet argument lorsqu'ils écrivent "Etant donné la forte fréquence de ce graphème...").

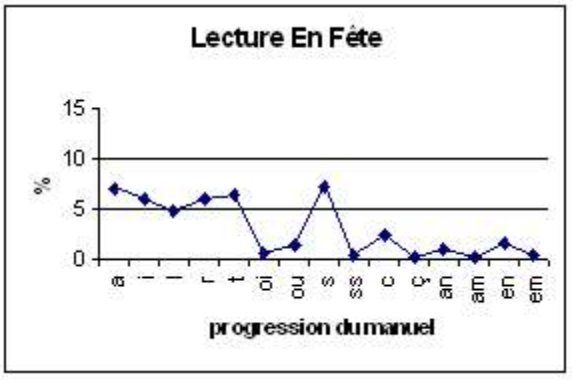
Sous peine de se mettre " à côté de la langue écrite ", la progression d'un manuel de lecture actuel se doit de respecter les fréquences statistiques des grammes et polygrammes. De ce

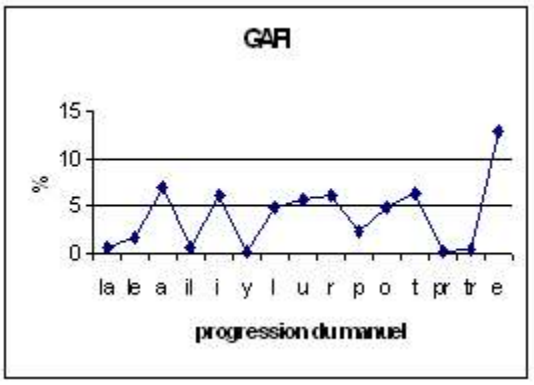
point de vue, Grindelire l'affirme en disant qu' il (l'ouvrage) "offre [...] la rigueur d'une progression pour étudier le code".

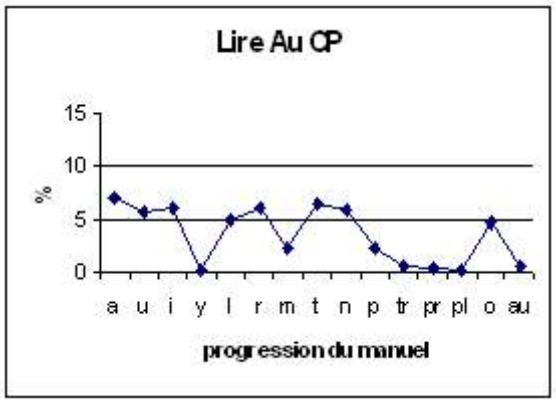
Voici, sous forme de graphique, une étude succincte des progressions visuelles des manuels, comparée à une progression de référence réalisée à partir d'un corpus de littérature de jeunesse réuni par l'ONL. Les enseignants pourront les interpréter en fonction de leur point d'intérêt ou du manuel qu'ils utilisent. (Exemple de lecture d'un de ces graphiques (Grindelire) : " a " est étudié en premier dans la progression, mais on le rencontre en troisième position par ordre de fréquence dans le corpus de référence (littérature de jeunesse). " u " est étudié en deuxième, alors qu'il est en huitième place dans le corpus. " e " est proposé en neuvième, bien qu'il soit en premier, et de loin, dans le corpus de référence. Certains manuels vont jusqu'à l'ignorer (peut-être parce qu'il paraît trop évident, ou parce qu'il a trop de valeurs différentes, comme dans " le ", " femme ", " des ", " en ", " eur ", " est ", " elle ", etc.).

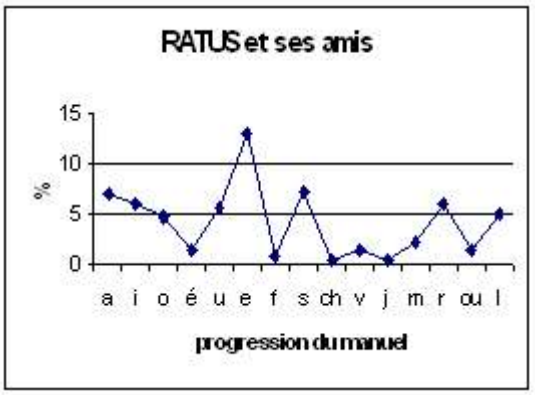










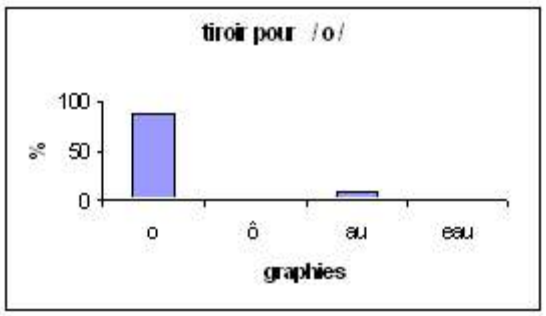


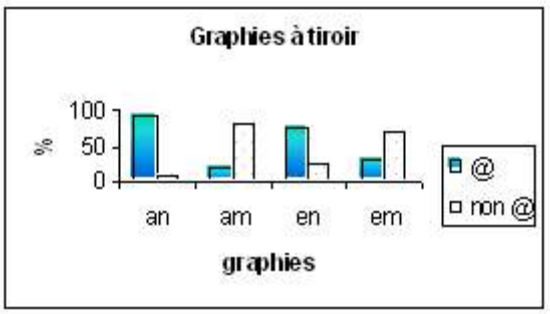
6.3.2. Les fréquences phonologiques (les phonèmes les plus fréquents).

La transcription de l'oral à l'écrit n'est pas toujours régulière. Cela relève de l'orthographe. Mais un manuel de C.P. doit veiller à ne pas tomber dans le piège de la dictée, qui relève davantage du C.E.1, et privilégier "la simplicité des graphies" (Lire au CP), "pour aller du plus simple au plus complexe" (Ratus). En conséquence, mieux vaut-il éviter :

- les " tiroirs à graphies " qui présentent d'emblée, dans la même leçon, les différentes façon d'écrire un son (par exemple "o, au, eau") alors que ces graphies ont des différences de fréquences extrêmes (voir graphique) ;
- les graphies qui ne se prononcent pas de la même façon dans les textes de tous les jours ("an" se prononce /@/ dans 90% des cas, comme "manger", mais "am" se prononce /@/ dans 20% des cas seulement ; "en" se prononce /@/ dans 75% des cas, comme "dent", mais "em" se prononce /@/ dans 0% des cas seulement).

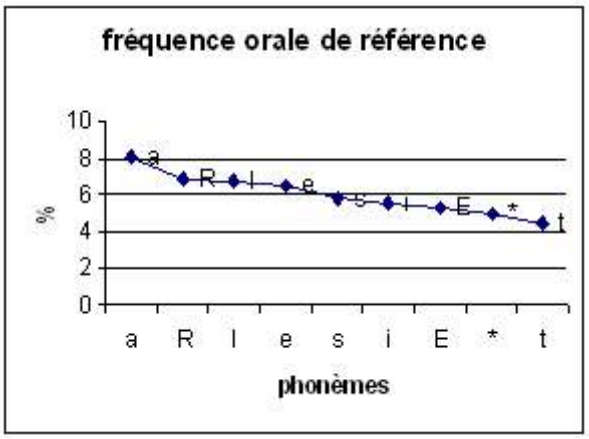
Les manuels qui ne prennent pas ces précautions risquent donc d'apprendre à lire de façon erronée !

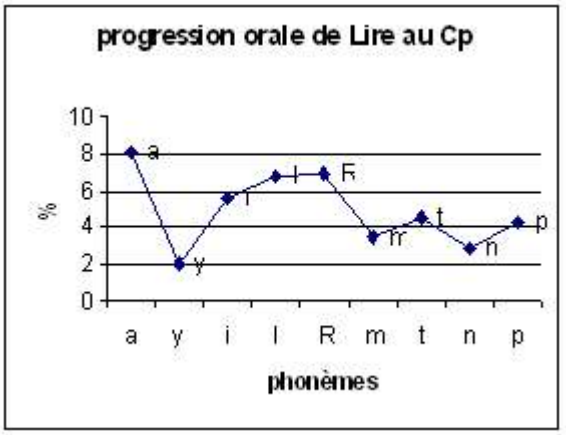


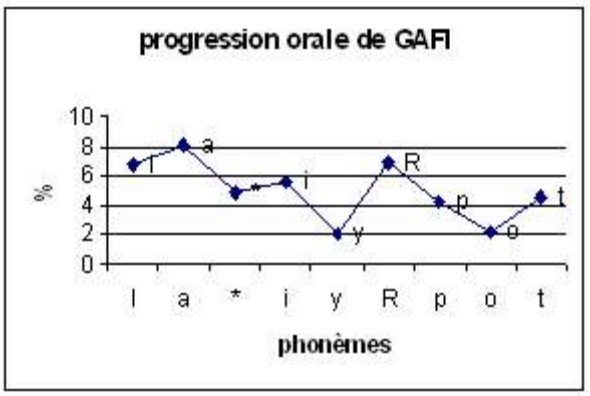


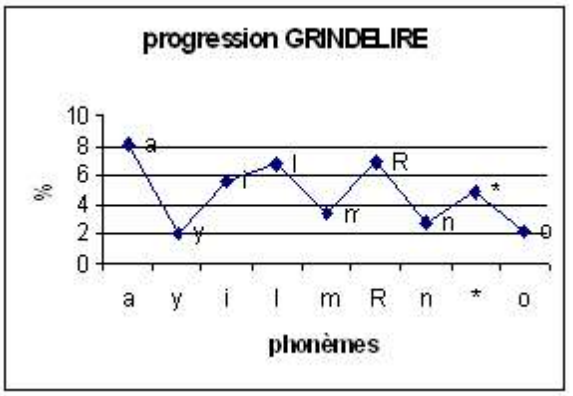
Rappel : par commodité, le son de " dent " est noté / @ /

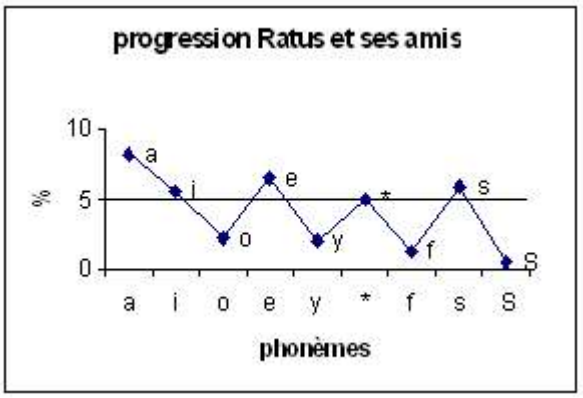
Les progressions des phonèmes doivent également refléter les faits de la langue orale, et les manuels devraient proposer une progression qui présente de façon décroissante les sons les plus fréquents de la langue. Voici, pour nos 5 manuels, ce qu'il en est : (nota : le schwa, ou e caduc, est représenté par *).

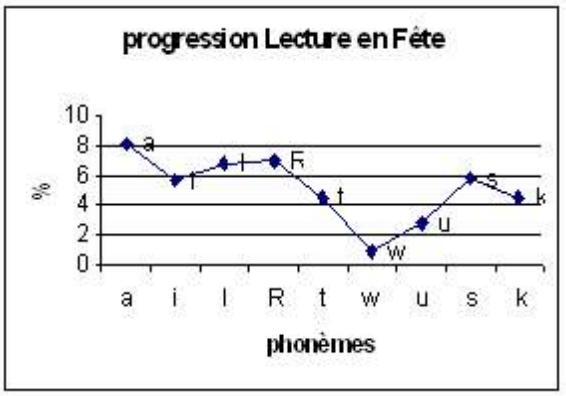










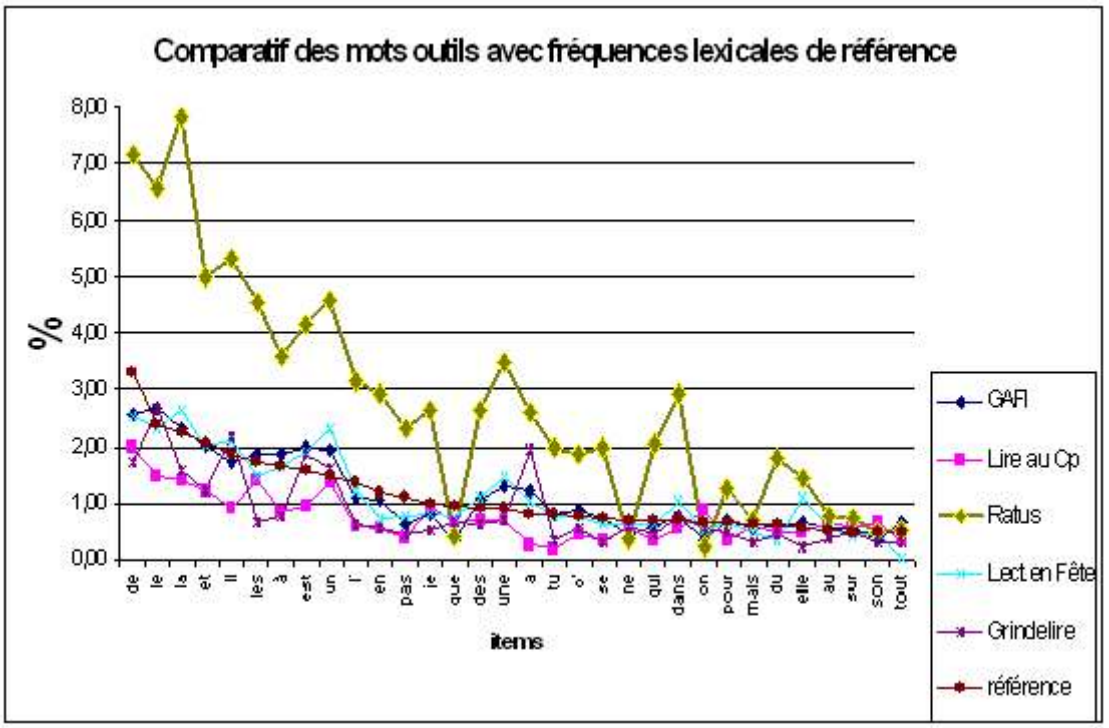


6.3.3. Les fréquences lexicales

La suite attendue des deux premiers points se situe au niveau des mots. Mais quels mots ? Les mots retenus seront de deux types :

6.3.3.1. Les mots outils

Ils permettent un repérage facile dans les textes de manuels, grâce à leur faible longueur (d'une moyenne de 3 lettres). Ils sont incontournables.



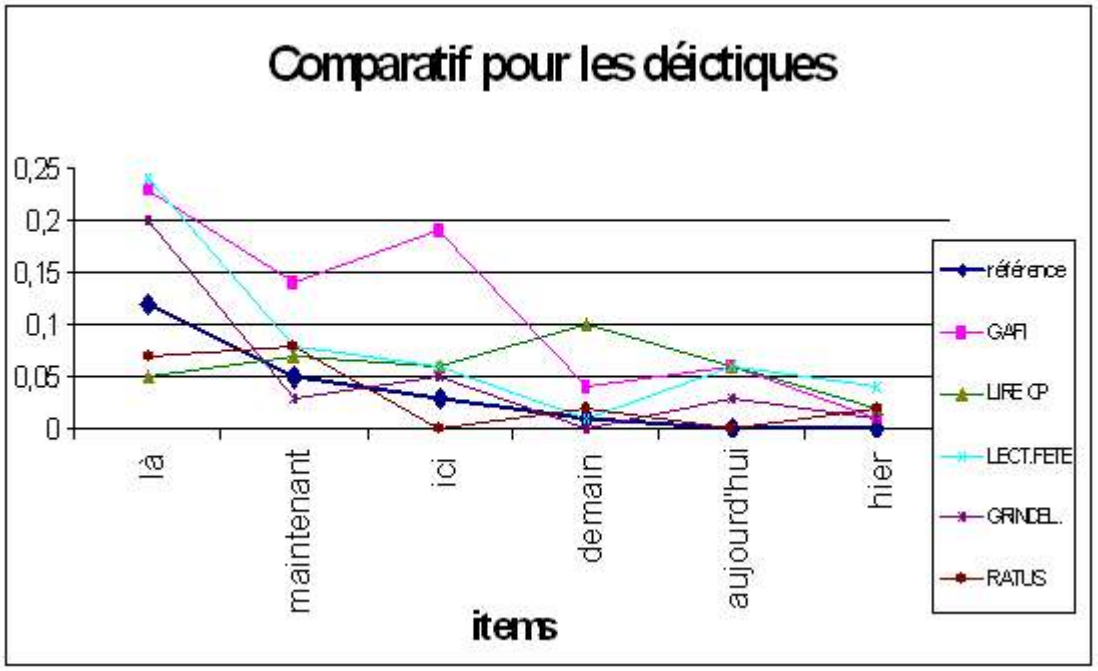
(le chevauchement des courbes montre ici que les manuels se sont presque tous alignés sur les fréquences du français écrit)

6.3.3.2. **Le vocable** (les mots différents d'un texte)

La combinatoire, qui permet d'assurer un apprentissage évolutif, reste la priorité (le sens est le but, non le moyen de la lecture). Pour cette raison, le vocabulaire doit représenter le mieux possible le système phonologique du français. J'ai constitué ce vocabulaire de base (voir pour plus de détails 7.2.).

6.3.4. Les déictiques

Mais il n'en reste pas moins que le manuel de lecture doit rester attrayant ("Des scénarios vivants...", Gafi), et proche du discours ("On insiste plus particulièrement [...] sur la séquentialité chronologique et les indicateurs qui la marquent (d'abord, puis, alors, enfin, pendant ce temps, dès que...), sur les relations logiques et les marques qui les manifestent : mais, ainsi, donc", Gafi). Pour cette raison, le manuel doit rester vivant et proche de l'enfant. Les déictiques, ces mots propres aux situations de communication, sont des révélateurs d'un texte vivant.



6.3.4. La lisibilité

En fin d'année, les textes à lire doivent, pour être accessibles, avoir un degré de lisibilité " assez facile " (selon la formule de Flesch, qui prend en compte la moyenne de mots par phrase, et la moyenne de syllabes par mot, on a : 0-30 très difficile comme chez Proust, 30-50

difficile, 50-60 assez difficile, 60-70 standard, 70-80 assez facile, 80-90 facile, 90-100 très facile comme une bande dessinée).

Le corpus de littérature de jeunesse a un degré de 48,10 (difficile/assez difficile). Pour comparaison, en prenant le dernier texte de chaque manuel (hormis les textes d'auteurs), on a ceci : Gafi 56,78 (assez difficile), Grindelire 43,41 (difficile), Lecture en Fête 63,11 (standard), Lire au CP 49,16 (difficile/assez difficile), Ratus 57,10 (assez difficile). Quant à l'intelligibilité (la compréhension), bien qu'elle s'instaure dès la maternelle, elle ne devient la priorité qu'à partir du C.E.1, lorsque l'enfant est (quasi)débarrassé des efforts du déchiffrage. On découvre donc que les manuels présentent des textes trop ambitieux en fin d'année de Cours Préparatoire. Peut-être leur faudrait-il proposer une alternative aux enfants moins avancés, avec des textes plus lisibles qui pourraient côtoyer les textes déjà présents.

6.4. La cohérence

Ce bref comparatif sur l'intérêt à porter aux fréquences de la langue, lors de l'apprentissage, ne cherche pas à être exhaustif, on le voit. Il a pour but de présenter certains indicateurs de cohérence sur les progressions des manuels de lecture et montre qu'en général chaque manuel étudié a mis en pratique, tout au plus et plus ou moins fidèlement, un ou deux faits de langue qui serviront de support à l'apprentissage de la lecture. Parfois, la volonté de bien faire amène les auteurs à organiser le manuel avec des progressions très scolaires, de niveau CE1, inadaptées pour un enfant en deuxième année de cycle 2.

Autant l'utilisation du manuel en classe doit rester pédagogique et dépendante des élèves, autant la conception de l'outil doit s'appuyer sur des principes rigoureux : le principe de cohérence est fondamental si l'on cherche à utiliser comme support un manuel à la fois proche de la langue orale et des réalités de la langue écrite. L'enseignant devra donc y être vigilant.

7. DÉBOUCHÉS PRATIQUES

L'autre versant de mes travaux, indispensable lorsqu'on souhaite une utilité sociale, est la proposition de pistes pour l'élaboration de supports pédagogiques. Mais il s'agit bien de suggestions, dont les éditeurs et les pédagogues pourront s'inspirer, s'ils le décident. Elles découlent directement des résultats obtenus par mes recherches en cours.

Il s'agit d'un lexique des 2800 premiers items du français écrit, d'un vocabulaire de base "orthophonologique" de 141 mots, et d'un choix de longueur des mots.

7.1. Les 2800 premiers items

Lexique des 2800 premiers items du français écrit (la version complète est disponible en téléchargement sur le site). J'ai effectué ce classement sur le corpus de littérature de jeunesse de L'ONL. Voici à titre indicatif les 110 premiers items (les ponctuations sont retenues) :

Rang	ITEM	Pourcentage
1	,	8,63708
2	.	4,09138
3	de	3,91286
4	et	2,37177
5	la	2,37110
6	le	2,06198
7	-	2,01529
8	à	1,74571
9	il	1,72144
10	l'	1,54773
11	les	1,48860
12	un	1,32331
13	que	1,27948
14	d'	1,23699
15	en	1,11818
16	je	1,02410
17	une	0,95690
18	des	0,90807
19	qui	0,89773
20	elle	0,89702
21	qu'	0,89292
22	;	0,87269
23	ne	0,80605
24	vous	0,77905
25	!	0,77821

26	dans	0,74687
27	est	0,70864
28	pas	0,69447
29	se	0,67550
30	ce	0,67061
31	du	0,62236
32	s'	0,60928
33	pour	0,60047
34	n'	0,57607
35	lui	0,54960
36	son	0,54092
37	au	0,52499
38	:	0,51719
39	était	0,51297
40	plus	0,48919
41	?	0,48675
42	sur	0,45398
43	on	0,45120
44	sa	0,44350
45	avait	0,44298
46	mais	0,42987
47	par	0,40260
48	comme	0,40064
49	a	0,39068
50	avec	0,38822
51	me	0,38777
52	« »	0,36948
53	si	0,36876
54	c'	0,36162
55	tout	0,35528
56	j'	0,34696
57	ses	0,33819
58	nous	0,33573
59	cette	0,31271
60	...	0,30963
61	bien	0,30460
62	y	0,28965
63	dit	0,28315
64	mon	0,26630
65	m'	0,25198
66	ils	0,24429
67	moi	0,23758
68	sans	0,23325
69	même	0,22519
70	tu	0,21389
71	ces	0,20858
72	où	0,20374
73	être	0,19460
74	ai	0,19382
75	deux	0,19317
76	aux	0,18362
77	ma	0,18218
78	là	0,17763
79	ou	0,16906
80	leur	0,16637
81	fait	0,16138

82	encore	0,15574
83	faire	0,15193
84	quand	0,15150
85	homme	0,14971
86	tous	0,14529
87	dont	0,13585
88	autre	0,13126
89	puis	0,13117
90	rien	0,12506
91	point	0,12284
92	peu	0,11969
93	cela	0,11767
94	donc	0,11264
95	sous	0,11020
96	femme	0,10725
97	votre	0,10704
98	avoir	0,10676
99	yeux	0,10637
100	m	0,10633
101	après	0,10405
102	mes	0,10388
103	aussi	0,10374
104	peut	0,10173
105	leurs	0,10144
106	toujours	0,10061
107	sont	0,10056
108	jamais	0,10008
109	fut	0,09913
110	non	0,09894

7.2. Vocabulaire “orthophonologique” de base

7.2.1. Critères de sélection

Le vocabulaire de base que je propose comporte 141 items. Il résulte de mon étude du corpus général. De ce corpus général, j’ai extrait tous les mots différents.

Ma méthodologie a consisté, à partir de ces mots différents (plus de cent mille) :

1) à utiliser toutes les lettres de l’alphabet en prenant pour chacune le mot le plus fréquent (par exemple pour la lettre "a" on a le mot "la", pour la lettre "b" on a le mot "bien", ... , pour la lettre "v" on a le mot "vous") ;

2) à employer différents environnements permettant de lire les valeurs phoniques d’une même lettre (par exemple la lettre "g" dans le mot "grand" et dans le mot "gens"), tout en respectant la condition 1) ("grand" et "gens" sont les mots les plus fréquents parmi ceux qui comportent la lettre "g" dans le corpus général) ;

3) à utiliser les polygrammes les plus fréquents tout en respectant la condition 2) (par exemple le polygramme "en" dans les mots "en", "rien", "étaient", "entre", "moment").

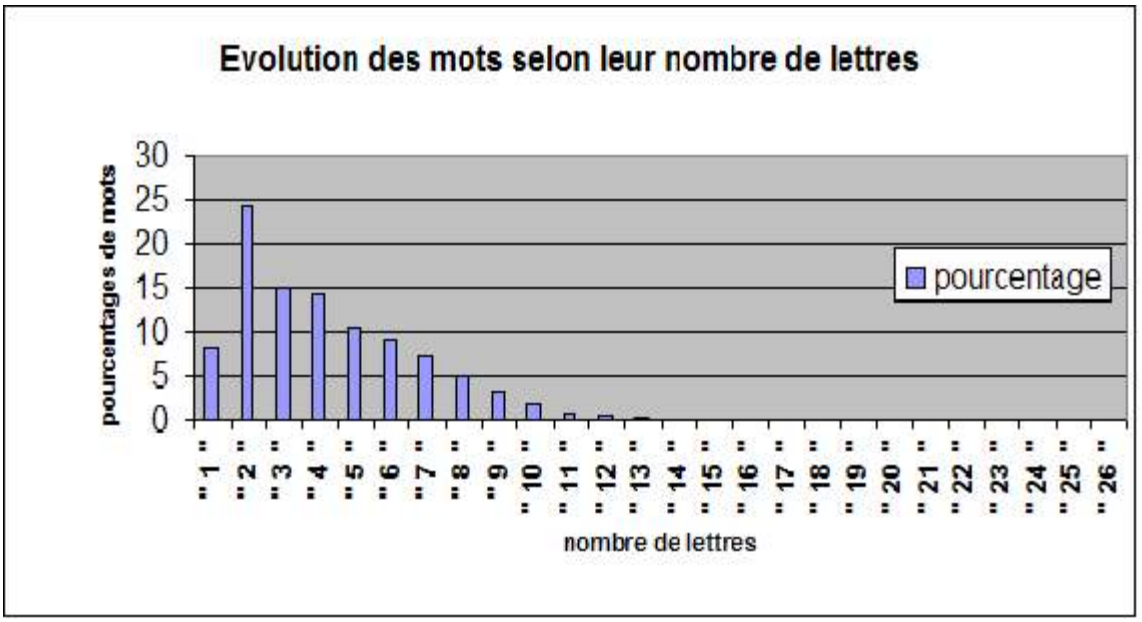
7.2.2. Liste

la	bien	ce	avec	comme
de	de	faire	grand	gens
visage	regard	chez	chose	Christophe
il	je	kilomètre	la	comme
me	ne	comme	encore	chose
pour	pas	que	quand	pour
rien	se	était	une	vous
avec	votre	wagon	dix [s]	expression [ks]
deux []	dix-huit [z]	exemple [gz]	y [i]	yeux [j]
pays [i]	voyait [j]	chez [ ]	avez [ ]	assez [ ]
nez [ ]	quinze [z]	était	travail	ailes
aime	faim	main	capitaine	jamais
chambre	dans	manière	au	autre
pays	ayant	Jean	Jeanne	beaucoup
peine	soleil	Reims	plein	femme
temps	premier	encore	rien	étaient
étaient	entre	moment	vers	premier
les	es (tu es)	toutes	presque	besoin
(il) est	reste	et puis	cette	petit
effet	deux	leur	feuille	exemple
expression	asseyez-vous	jockey	chez	signe
bien	étaient	conscience	sienne	il
fille	aime	sentiment	simple	vraiment
immobile	point	enfin	coeur	soeur
voeu	moi	point	point	moines
comme	comment	son	personne	vous
brouillard	voyait	phrase	que	conversation
question	un	une	jeune	jaune

appuyé				

7.3. Choix de longueur des mots

Proposition d'un tableau récapitulatif pour le choix de la longueur des mots dans les manuels.



Si la longueur moyenne des mots en français écrit est de 4 à 5 lettres, il serait faux de croire que la plupart des mots comportent 4 à 5 lettres. Une étude plus fine, grâce à un programme en Perl, me permets de montrer la répartition des mots selon leur nombre de lettres. On voit nettement que les mots de 2, 3, 4 et 5 lettres sont les plus fréquents, avec une nette dominance pour les mots de 2 lettres. C'est bien sûr la haute fréquence de "de" "la" "et" "le" "il","un", qui explique ce fait de langue (à eux seuls, ces six items pèsent entre 10% et 13% des corpus).

Il conviendrait donc, dans les manuels de lecture, de privilégier à la fois ces "petits mots" en raison de leur fréquence élevée, et les mots les plus courts (et toutefois fréquents) en raison cette fois de l'empan visuel (voir Pour un modèle de lecture).

7.3.1. Résumé

Pour résumer, les manuels de lecture devraient présenter :

1) les petits mots les plus fréquents ;

2) essentiellement des mots de une à cinq lettres ;

3) des mots dont le choix permet la plus grande représentation du système phonologique du français (voir pour exemple le contenu du Test02 et ses justifications théoriques).

8. RECHERCHES EN COURS ET PISTES DE RECHERCHES

Ces autres recherches sont la continuité de mes travaux. Elles concernent toujours l'apprentissage de la lecture.

8.1. Coupe syllabique

Un travail supplémentaire s'impose, et consiste à établir deux listes de pseudo-mots, l'une où ils sont coupés de façon non conforme à la phonologie, et l'autre où ils sont écrits selon les règles phonologiques du français (exemples : t*rottoire vs tro*ttoire ; att*raper vs a*ttraper ; chap*eau vs cha*peau). Si les pseudo-mots mal coupés sont lus plus difficilement, alors l'influence phonologique sur le décodage est encore plus importante qu'on ne l'a vu.

Dans les deux premiers exemples, le principe de l'attaque maximale (: entre deux voyelles séparées par des consonnes la frontière syllabique est celle qui déplace le nombre de consonnes dans l'attaque de la seconde syllabe, à condition que ce digramme consonantique soit acceptable) doit entraîner la structure branchante suivie d'une voyelle "tr + V".

8.2. Rendement phonologique et commutations

En français, les traits distinctifs minimaux (phèmes) voisé vs non-voisé ont un rendement élevé dans des paires minimales (poule vs boule ; file vs ville ; quand vs gant). L'erreur de décodage consiste en une commutation (p pour b, f pour v, etc.). Il semble que plus ce

rendement est phonologiquement élevé, plus la commutation est fréquente. Plus il est faible, plus elle est rare. Si cela était possible, il faudrait trouver des locuteurs ne connaissant pas certains mots (même à l'oral, par exemple grâce à des questions du type "comment s'appelle...."), et leur faire lire le mot qu'il ne connaissent pas dans la paire minimale. Dans un premier temps, relever les éventuelles erreurs. Dans un second temps, rendre familier le mot inconnu, le faire lire dans des listes, puis comparer les taux d'erreurs dans les deux cas. Si les résultats sont proches, on pourra en conclure que l'erreur est essentiellement d'ordre phonologique. Dans le cas contraire, il s'agirait d'erreurs dues à une amorce orthographique (ce cas devrait se trouver chez les décrypteurs).

8.3. Enchaînements et conscience syllabique

Autre facteur d'erreur : le phénomène d'enchaînement ("leur_ami") et non le phénomène de liaison ("les_amis") pourrait perturber la conscience syllabique et provoquer des non-mots ("leura...mi"), ce qui pourrait être à l'origine de la non compréhension. Le désordre dans la coupe syllabique entraînerait un désordre dans le décodage puisque la distribution graphémique (: au niveau visuel) s'accompagne par conséquent d'une nouvelle donne phonémique (: au niveau de la subvocalisation).

8.4. Déficiences auditives

Dans le cas de déficiences auditives, rédiger un texte-test à faire lire à deux populations d'enfants : a) déficients auditifs; b) non déficients.

Si le décodage est effectivement touché par des facteurs idiophonologiques, les erreurs de décodage seront identiques pour tous les a), identiques pour tous les b), et différentes entre a) et b), chaque population ayant des contraintes phonotactiques différentes.

Si mon hypothèse est juste, les erreurs produites par les enfants sourds débutants lecteurs ne

doivent pas être identiques à celles des enfants entendants, puisque leur répertoire phonologique est incomplet et déformé. Voir les études de Laurence Paire-Ficout et Nathalie Bedoin sur le code phonologique précoce chez le lecteur sourd. Il devrait y avoir par exemple une plus grande réalisation de consonnes labialisées que chez le décodeur entendant. Il faudrait analyser le système phonologique d'un enfant sourd, pour être en mesure de comparer ses erreurs de décodage à celles d'un enfant entendant.

8.5. Mémoire de travail

Des tests devront être mesurés (temps de réponse) pour savoir si (et dans ce cas comment) la mémoire de travail entre en jeu dans le décodage, dans la mesure où, définie dans la littérature comme l'ensemble des processus permettant d'éviter un débordement de traitement, elle engendre un choix de décodage à partir d'un seuil limite d'empan visuel. On devrait alors trouver des modifications de traitement dans le décodage, dues à l'abandon d'un graphème non décodé d'une part, et à un saut visuel sur un autre graphème. Cela devrait se confirmer dans des mots à écriture inconnue. Par exemple pour le prénom Valérie , à un stade où le décodeur ignore la lettre " é ", on obtient / vali /. Le " é " est littéralement supprimé, le décodeur cherche la structure CV ("l" + V ?) et va chercher le " i ". Le retour, fréquent lors des hésitations importantes (il y a balayage en essuie-glace pendant l'aporie verbale), associe le "r" préposé, auquel s'ajoute un schwa (habituel en français oral hésitant).

8.6. Environnements polygrammiques conflictuels

Les erreurs purement visuelles dans les environnements polygrammiques (par exemple les deux valeurs de "a" dans "ans" et "années", ou bien l'inversion de e+n+i dans "chenille" lu d'abord "chan" puis "chien" et enfin "che...ni..." complexifient le décodage et s'additionnent aux erreurs audiophonologiques. Dans ces cas, le facteur idiophonologique pourra être conflictuel. Par exemple, "oi" se décodera d'abord en "o + i" (et non [wa]), et le mot

"coiffeur" a été lu [kof] : élision du "i" car digramme difficile, puis refus de syllabe fermée avec suppression de "eu". Au lieu d'avoir une structure C+(C+V)+C+V+C [kw f], on a [C+V+C+C+V], c'est-à-dire une banale suite "attaque+rime+coda (coda constituée d'une branchante et d'une voyelle finale, le schwa, en vertu du principe de sonorité). Une sorte de retour à un confort articulatoire. Il faudra avoir suffisamment de faits observés pour en déduire des règles prédictibles.

8.7. Contraintes idiophonologiques et représentations visuo-perceptives

Pour vérifier si un phonème considéré est bien la cause d'une erreur observée, il faudra constituer une liste de mots placés de façon aléatoire. Pour certains de ces mots, on remplacera le gramme (ou le polygramme) par un symbole auquel l'observateur associera explicitement le phonème correspondant (par exemple un losange pour le [m]). Ces mêmes mots seront réécrits entièrement (sans le symbole). Toutes les erreurs engendrées par la présence de [m] devraient exister aussi dans les mots avec symbole, ce qui permettrait de voir si les contraintes idiophonologiques dominent les représentations visuo-symboliques. Pour vérifier si mon hypothèse est viable, il faut s'assurer que le décodeur est bien en situation d'aporie verbale (ce qui signifie qu'il cherche à décoder) et non en situation piégée par l'environnement visuel. Il est donc nécessaire d'établir des pseudo-mots où l'analogie graphique est forte mais sans qu'il y ait d'analogie phonologique (p et d par exemple, proches visuellement mais éloignés phonétiquement). De cette façon, on devra s'attendre à des erreurs plus nombreuses par contrainte idiophonologique que par confusion visuelle, à chaque fois que le premier phonème est lu correctement (puisque en conséquence il va activer le processus phonotactique, et ignorer le rapprochement visuel des graphies proches). Le choix des graphèmes pour piéger le décodeur se fera sur des indices oculaires de surface (tels que la hauteur relative, la symétrie horizontale ou verticale, l'empattement, etc.) et devra tenir

compte des fréquences des graphèmes en français.

8.8. Dyslexiques

Établir une corrélation entre les erreurs banales de décodage et les erreurs que l'on rencontre chez l'enfant dyslexique (confusion de lettres de formes voisines, élisions et inversions de lettres, mais aussi dévoisement, passage des constrictives aux occlusives). Les performances que l'on peut attendre devraient être contradictoires, ce qui pourrait montrer que les causes d'erreur ayant une source différente, les erreurs prédictibles du décodeur non dyslexique seraient véritablement dues à son système idiophonologique, alors que le décodeur dyslexique a (selon la littérature) des causes neuropsychologiques (prédominance de l'hémisphère droit sur le gauche), ce qui en soi n'explique pas la cause précise de l'erreur.

Une typologie des erreurs de décodage, si elle est suffisamment fonctionnelle, pourrait servir à étudier les cas de dyslexie et proposer, peut-être, des applications pratiques pour les enseignants du Primaire et les orthophonistes.

9. UTILITÉS

Utilité à court, moyen, et long terme : a) au niveau linguistique, une compréhension plus affinée des processus de décodage ; b) au niveau didactique, une mise en oeuvre de protocoles cliniques pour des orthophonistes ; c) au niveau pédagogique, une contribution à la constitution de manuels de lecture ; d) au niveau social, une contribution à la lutte contre l'illettrisme.

10. TERMINOLOGIE

La terminologie est parfois source de malentendus. Des recherches sur la lecture effectuées par une discipline comme l'optométrie peuvent être différemment interprétées, en raison de la confusion entre la vue et la vision. Un enfant peut bénéficier d'une bonne vue (et n'a pas besoin de verres correcteurs), mais sa vision n'est pas conséquemment suffisante : l'empan visuel, la motricité oculaire, sont parmi d'autres facteurs les moyens d'appréhender confortablement la lecture.

Ces déficits notionnels m'ont amené à proposer quelques définitions, de façon à éviter des confusions.

10.1. Lettre, ou gramme ? Digraphe ou Digramme ?

La littérature linguistique sur l'apprentissage de la lecture, ainsi que les ouvrages de pédagogie, utilisent la plupart du temps le mot "lettre". Parfois, les termes "graphe"ou "digramme" apparaissent.

Dans mes travaux, j'emploi "gramme" pour désigner "la lettre lue".

De même, si un "digraphe" signifie deux (= di) lettres (= graphes) associées et contiguës pour représenter généralement un phonèmes, je suggère, comme il s'agit de lecture, d'employer la notion de digramme, car une lettre ne se prononce pas, elle s'écrit. On prononce un son.

L'origine "graphein" incite à réserver le mot "digraphe" pour l'écriture (on parlera par exemple de calligraphie). Cette recommandation terminologique permettrait la distinction suivante : une erreur d'écriture (pouvant aller jusqu'à la dysorthographie) concernera des digraphes (ou des lettres), et une erreur de lecture (pouvant aller jusqu'à la dyslexie) concernera des digrammes.

Un digramme (deux grammes, c'est-à-dire deux caractères à lire, dans le sens de support

visuel symbolique) se lira généralement d'une seule émission de voix, sauf pour les diphtongues ("oi" par exemple, qui représente une consonne suivie d'une voyelle). Excepté les langues écrites où un seul gramme est associé à un seul phonème (le cas extrême est l'écriture serbe), le digramme est le propre des langues analytiques.

Sur cette voie, on parlera de “polygramme” pour une séquence égale ou supérieure à trois grammes, en quelque sorte une grappe de grammes.

Voici leurs fréquences (exprimées en pourcentages) sur le corpus de 7 millions de mots (on retrouve approximativement les mêmes valeurs pour des corpus plus réduits, mais il faut bien noter qu'il s'agit de statistique, et un type de texte pris isolément pourrait sensiblement varier, mais en respectant toutefois assez précisément l'ordre décroisssant ci-après) :

"es" : 1.8434 % ; "en" : 1.6890% ; "ai" : 1.6670% ; "ou" : 1.4377% ; "on" : 1.2783% ; "an" : 1.0834% ; "qu" : 1.0706% ; "er" : 0.9422% ; "et" : 0.7943% ; "eu" : 0.7751% ; "in" : 0.6913% ; "il" : 0.6891% ; "ce" : 0.6093% ; "co" : 0.5984% ; "oi" : 0.5726% ; "un" : 0.5695% ; "au" : 0.5362% ; "ch" : 0.4357% ; "em" : 0.4137% ; "om" : 0.3763% ; "ge" : 0.2341% ; "am" : 0.1848% ; "ca" : 0.1602% ; "ez" : 0.1565% ; "ci" : 0.1435% ; "ei" : 0.1283% ; "im" : 0.1279% ; "ga" : 0.1056% ; "gn" : 0.0992% ; "cu" : 0.0769% ; "gu" : 0.0708% ; "oy" : 0.0648% ; "ex" : 0.0623% ; "gi" : 0.0432% ; "ay" : 0.0422% ; "go" : 0.0365% ; "ph" : 0.0271% ; "uy" : 0.0084% ; "ey" : 0.0035% ; "cy" : 0.0014%.

Cela revient à dire que la probabilité de rencontrer le digramme "ai" dans un texte est inférieure à deux pour cent, ou bien que sur un texte qui comporterait mille caractères, c'est-à-dire environ 200 mots, "ai" apparaîtra généralement entre trois et quatre fois. De même, sur un même texte de 200 mots (la moitié de cette page), "cy" n'a aucune probabilité d'apparaître (il lui faudrait un texte de vingt mille mots pour avoir cette chance !). Le digramme "ai" aura presque deux mille fois plus de chances d'apparaître que "cy".

10.2. Attaque

Les mots à l'oral sont émis d'une façon continue. Mais si l'on observe leur prononciation, on peut remarquer une construction en syllabes, chaque syllabe correspondant à peu près à une émission de voix.

Chaque syllabe n'est pas identique aux autres, mais toutes n'ont qu'une façon générale d'être construite : un début, et une suite.

Ce début de la syllabe s'appelle l'attaque, suivie de la rime.

L'attaque peut être simplement constituée d'une consonne (le /p/ de /pal/), ou de deux (les /pl/ de /pli/).

Quand une attaque est constituée de deux consonnes, on l'appelle branchante.

Parfois, l'attaque et la rime (/p/ + /a/ => /pa/) sont suivies d'une coda (le /l/ de /pal/).

Cette notion d'attaque est importante pour commencer le développement de la conscience phonologique de l'enfant, en lui proposant une liste de mots presque identiques, dans laquelle s'est glissé un intrus dont l'attaque est différente (ex. : pal, pull, pomme, sol, pou), ou encore en lui demandant de supprimer l'attaque (ex. : dire les mots pal, pull, pomme, pou en enlevant /p/).

10.3. Décodage

Une chaîne orale ininterrompue dite dans une langue étrangère inconnue de l'interlocuteur n'aura aucune signification pour ce dernier. Ce sera une suite de sons sans sens. Il suffit pour s'en convaincre d'écouter une chanson dans une langue qu'on ne connaît pas.

La chaîne orale du discours (c'est-à-dire les paroles) est une expression orale selon un certain code. Ce code diffère selon les langues, mais l'interlocuteur doit toujours, pour comprendre ce

message oral, le décoder en le décomposant en éléments plus petits et identifiables : les mots, porteurs de sens.

Si cette chaîne orale est dite dans sa langue, ou dans une langue qu'il comprend, l'interlocuteur pourra la décoder par un découpage. Ce découpage utilisera aussi bien la reconnaissance de mots, que le rythme de la phrase, essentiellement grâce à la syllabe qui se prête naturellement au découpage (il est plus facile par exemple d'isoler des syllabes que des phonèmes lorsqu'on entend des sons, et des enfants arrivent facilement à scander un texte en tapant le rythme sur chaque syllabe).

Le décodage est donc une notion qu'il faut réserver à l'oralisation (la lecture d'un mot nécessite une sub-vocalisation, et relève donc d'une forme particulière d'oralisation). Pour l'écrit, on parlera de déchiffrage qui, lui, à l'inverse, ira des éléments les plus petits (les lettres) vers les éléments les plus complexes (syllabes, mots, phrases).

10.4. Fréquence des mots

Nous utilisons des mots pour parler, et certains de ces mots sont plus souvent utilisés que d'autres. Dans ce cas, ils apparaissent avec une fréquence plus élevée, qu'on exprime habituellement en pourcentage. Par exemple, sur cent mots écrits, le mot outil "de" apparaît presque quatre fois.

Une fréquence élevée indique une probabilité élevée de rencontrer un mot (à l'oral ou bien à l'écrit).

Les fréquences sont utiles car elles permettent de trier dans l'ensemble des mots ceux qui sont les plus nécessaires, tant à l'oral pour l'apprentissage du vocabulaire, qu'à l'écrit pour l'apprentissage de la lecture.

Inversement, les mots rares, c'est-à-dire les moins fréquents, ne présentent qu'un intérêt limité et occasionnel (le mot "lanterne" est sûrement moins utile que le mot "temps", par exemple).

Un mot qui n'apparaît qu'une fois dans un texte s'appelle un hapax.

Il existe une loi mathématique (appelée loi de Zipf) qui montre que la fréquence des mots d'un texte est inversement proportionnelle à leur rang. Cela signifie que, si l'on fait un classement des mots d'un texte en utilisant leurs fréquences, on obtient la formule suivante : rang d'un mot multiplié par sa fréquence = constante. Pour simplifier, le centième mot sera cent fois plus rare que le premier.

En fait, cette loi n'est valable que si l'on ne tient pas compte des extrêmes (les mots les plus fréquents et les mots les plus rares).

10.5. Lisibilité :

(notion qui s'exprime conventionnellement en degrés). Un texte écrit possède des particularités de surface : des phrases longues, moyennes, ou brèves, des mots longs, moyens, ou courts. Ces particularités d'ordre visuel déterminent la complexité du texte.

Des phrases très longues dotées de mots longs (comme chez Proust) offrent un degré de lisibilité plus faible que des phrases brèves dotées de mots courts (comme les bandes dessinées, ou les histoires pour enfants).

Cette notion de lisibilité est associée directement à la mémoire de travail, qui possède ses propres contraintes. Par exemple, on considère en moyenne qu'une personne retiendra un maximum de sept mots en mémoire immédiate : une phrase écrite de quatorze mots

risquera de n’être retenue qu’à moitié, ce qui la rendra peu lisible.

Inversement, une phrase de quatre ou cinq mots sera lisible à cent pour cent (à condition que les mots eux-mêmes soient d’une longueur moyenne ou petite).

Il ne faut pas confondre la lisibilité et l’intelligibilité. La lisibilité concerne la forme (ce qui est vu) alors que l’intelligibilité relève du fond (le contenu, le sens).

Sur le plan pédagogique, l’enseignant a tout intérêt à vérifier si les textes proposés aux élèves sont lisibles. Il existe pour cela des méthodes de calcul qui prennent en compte le nombre de mots par phrase et le nombre de lettres par mot (voir par exemple les formules de Flesh). Mais à n’en pas douter le bon sens de l’enseignant permettra plus rapidement, à vue d’oeil, d’apprécier la lisibilité d’un texte.

10.6. Saccade :

pour lire, l’oeil se déplace le long des lignes, et s’arrête à certains endroits. Ces déplacements successifs entrecoupés de pauses s’appellent “saccades”.

C’est pendant les pauses, ou fixations, que la lecture est effective.

Les saccades sont d’autant plus nombreuses que le lecteur est débutant, car son regard possède un empan faible (quelques mots, voire seulement quelques lettres).

Elles sont d’autant plus rapides que le lecteur est expert (il reconnaît d’emblée les mots familiers).

En définitive, pour un même nombre de mots, un lecteur débutant fait plus de saccades (avec souvent des retours), et lentement, alors qu’un lecteur expert fait moins de saccades (et moins de retours) et cela plus rapidement.

10.7. Empan visuel :

si l’on ouvre la main, doigts écartés, la distance maximale en ligne droite entre le bout du pouce et le bout de l’auriculaire s’appelle “empan”.

En lecture, l’empan visuel est la distance maximale que le regard peu capter lorsqu’il se fixe sur un mot.

L’empan visuel atteint (pour la moyenne des individus) un ensemble de sept lettres contiguës.

Un mot comme “un” est lu avec une seule fixation, car il nécessite un empan visuel faible (deux lettres).

Un mot comme “indéniablement” entraîne plusieurs fixations puisque son nombre de lettres dépasse l’empan visuel maximal (dans le meilleur des cas, ce mot nécessite deux fixations ; dans la pratique, trois ou davantage !).

Pour en savoir plus : l’empan visuel correspond à la fovéa et à sa périphérie immédiate. La fovéa est le point central et le plus net de la vision. Plus les lettres environnantes en sont éloignées, plus elles sont illisibles.

10.8. Vocabulaire vs Vocable

La littérature sur la question est souvent contradictoire. Plutôt que de proposer, de dicto, une définition personnelle qui serait prétentieuse, je préfère proposer ce raisonnement suivant, qui a le mérite de pouvoir être discuté : le dictionnaire offre des entrées (les lemmes) chacune différente des autres, classées par ordre alphabétique. Or, les déclinaisons (formes fléchies et autres dérivés) ne sont pas incluses dans l'ordre alphabétique, et souvent ignorées (par exemple, le mot "capsule" ne sera pas juste

avant le mot "capsules", puisque son pluriel ne sera pas indiqué). De ce fait, les mots épicènes (dont la forme ne varie pas au féminin et au masculin) ne sont pas non plus deux mots différents. "Artiste" (un ou une) vaudra pour une entrée, mais pourtant un locuteur qui emploiera "un artiste" et "une artiste" aura un vocabulaire de deux mots, et dans la phrase "Je vous présente M. Untel, un artiste de talent, et Mme Unetèle, une artiste tout aussi talentueuse", il y a deux sens pour "artiste", deux fois le mot "artiste", pour une seule forme commune ("artiste"). Le locuteur a fait preuve d'un vocabulaire riche (dans la mesure où il utilise les deux genres possibles du mot "artiste") mais n'utilise pour cela qu'un vocable : "artiste".

Pour résumer, il semble plus cohérent d'utiliser "vocabulaire" pour "l'ensemble de tous les mots rencontrés dans un texte, quelle que soit leur forme", et "vocable" pour "l'ensemble de tous les mots différents d'un texte".

Le nombre de mots total et *le nombre de mots différents* couvriraient la même notion. Il y a redondance.

Pour éviter cela, je propose ceci :

Le vocabulaire serait le nombre total de mots (un mot qui apparaîtrait une fois au singulier, une fois au pluriel, correspondrait à deux mots de vocabulaire).

Les vocables, eux, seraient les mots différents (un mot qui apparaîtrait cent fois de façon identique correspondrait à un seul vocable).

Il reste (hélas) quasi impossible de différencier à l'heure actuelle les sens différents d'un vocable ("droit" civique ou "droit" comme un i), et le décompte informatique ne retiendra qu'un mot.

11. BIBLIOGRAPHIE

ALEGRIA & MORAÏS (Université libre de Bruxelles, Belgique) : Analyse segmentale et acquisition de la lecture

BENTOLILA, Alain : *De l'illettrisme en général et de l'école en particulier*, Plon, 1997

BENTOLILA, Alain : *Le propre de l'homme : parler, lire, écrire*, Plon, 2000

BOUCHARD (Marie-Joëlle) : " Apprendre à lire comme on apprend à parler ", Paris, 1991, 174 p, pp.145-146, Collection *Pédagogies pour demain. Didactiques.* Isbn 2-01-018322-3 ;

BOURBEAU, Nicole, (1988), C'est pas lisable ! La lisibilité des textes didactiques, Guide pratique, Sherbrooke, Collège de Sherbrooke, 166p.

BOURQUE, G. (1989), Des mesures de lisibilité, Communication présentée au 57e Congrès de l'ACFAS. Montréal: [inédit].

BYRNE Brian (University of New england, Australie) : Etude expérimentale de la découverte des principes alphabétiques par l'enfant.

CARBONNEL S., GILLET P., MARTORY .-D., VALDOIS S., 1996, Neuropsychologie : approche cognitive des troubles de la lecture et de l'écriture chez l'enfant et l'adulte, Solal.

CHALL, J. S. (1958), Readibility: An appraisal of research and application, Colombus: Ohio State University Press.

CHAUVEAU G, 1997, Comment l'enfant devient lecteur, Retz.

COEFFÉ, C., HUMBERT, R., JACOBS, A.M, & O'REGAN, J.K. L'analyse des mouvements oculaires en temps réel. Informatique et Sciences Humaines. 1985, 65, 67-72.

O'REGAN, J.K. Compte rendu de la Troisième Conférence Européenne sur les mouvements des yeux. Bulletin de la Société d'optique Physiologique, 1987

CONTENT, A. (1993). Le rôle de la médiation phonologique dans l'acquisition de la lecture. In J.P. Jaffré, & M. Fayol (Eds), *Lecture-Ecriture : acquisition,* Les actes de la Villette (pp. 80-96), Paris : Nathan Pédagogie

Corpus :

Corpus de SZKLARCZYK (1961) : Essai sur la structure phonologique du français, Ph. D., University of Pennsylvania, Philadelphia ;

corpus de WIOLAND : F. Wioland, Etude statistique des phonèmes et diphonèmes dans le français parlé, dans Revue Acoustique, 16 (1971), p. 258-262 ;

corpus de BAUDOT : J.A. Baudot : *Information, redondance et répartition des lettres et des phonèmes en français,* Université de Montréal, 1968 .

Corpus de HATON et LAMOTTE (1971) : *Essai sur la structure phonologique du français*, Ph. D., University of Pennsylvania, Philadelphia, 1961 ;
Corpus de DE KOCK (1971 et 1974), à partir de L. Warnant, Dictionnaire de la prononciation française, Gembloux, 1965, 1390900 phonèmes.

DALE, E. et CHALL, J. S. (1948), A formula for predicting readibility, Columbus: Bureau of Educational Research, Ohio State University.

EHRLICH, Marien France, et TARDIEU, Hubert, (1985), Lire, comprendre, mémoriser les textes sur écran vidéo, Communication et langages, #65, p.91n106.

FERNBACH, Nicole, (1990), La lisibilité dans la rédaction juridique au Québec, Ottawa, Le Centre de promotion de la lisibilité, Centre Canadien d'information juridique, 128p.

GÉLINASnCHEBAT, C., MACOT, M., PRÉFONTAINE, C., et DAOUST, F. (1991), *La lisibilité de documents d'information du ministère de la Main d'oeuvre, de la Sécurité du revenu et de la Formation professionnelle*, Avis professionnel présenté au ministère de la Main d'oeuvre, de la Sécurité du revenu et de la Formation professionnelle, Gouvernement du Québec, 50 p.

GOMBERT (Jean-Émile) : " L'apprentissage de la lecture : apports de la psychologie cognitive " in *L'enfant apprenti lecteur*, sous la direction de Gérard Chauveau,

Martine Rémond et Eliane Rogovas-Chauveau, L'Harmattan, Paris, 1993, Collection CRESAS n°10

GOMBERT (Jean Emile) *La construction des connaissances phonologiques chez l'enfant*, Revue Parole, 1999 -9/10, pp89-100

HAMERS et BLANC, Bilinuality and Bilingualism, 1989

HJELMSLEV (Louis) : *Nouveaux essais*, Paris, Presses Universitaires de France, 1985, 207 p, Collection *Formes sémiotiques*, Isbn 2-13-038827-2

JAMET E., 1998, Comment lisons-nous ?, Sciences humaines, N˚82, PP. 20-25.

LALANDE (Jean-Noël) : *L'apprentissage de la langue écrite, du b-a ba à la b.d.*, Paris, 1985, 184 p, Presses Universitaires de France, Isbn 2-13-039105-2

LEON (Pierre Roger) : *Phonétisme et prononciation du français*, Paris, 1992, 192 p, pp. 75-76, Collection *Faculté Linguistique*, Isbn 2-09-190290-X

LIBERMAN Isabelle & SHANKWEILER Donald (University of Connecticut et Haskins Laboratories) : Phonologie et apprentissage de la lecture : une introduction.

MANN Virginia A. (University of California, Irvine) : Les habiletés phonologiques prédicteurs valides de futures capacités en lecture.

MAUFFREY (Annick) et COHEN (Isday) : *Eléments pour une pédagogie différenciée*, Paris, 1995, 231 p, collection *Formation des enseignants. Professeurs des écoles*

MAZAUX (Jean-Michel) et GUEGAN (Jacqueline) : " Cerveau, langage et lecture " in *La lecture*, tome 1, de la neurobiologie à la pédagogie, L'Harmattan, Paris, 1990, p. 219

MURONI (Jean-Marc), "La connaissance du fonctionnement des langues maternelles peut-elle contribuer à l'acquisition du français ?", in *Petit dictionnaire bantou du Gabon : ndjabi/français/ndjabi*, L'Harmattan, Paris, 1989, 1992, 2000.

O'REGAN, J.K., Pynte, J., & COEFFÉ, C. Comment le regard explore un mot isolé. Bulletin de Psychologie, 1986, 39, 429-432.
O'REGAN, J.K., & LÉVY-SCHOEN, A., Les mouvements des yeux au cours de la lecture. L'Année Psychologique, 1978, 78, 459-492.

O'REGAN, J.K., L'utilisation d'un petit ordinateur dans l'étude des mouvements oculaires. Informatique et Sciences Humaines, 1974, 22, 7-12.

PAIRE-FICOUT, L., & BEDOIN, N. (1996). Rôle du code phonologique précoce en lecture silencieuse chez des sourds et des entendants. Actes de Colloque : Perception Cognition Handicap. Lyon

PALLIER, C., BOSCH, L., & SEBASTIAN-GALLÉS, N. (1997) A limit on behavioral plasticity in speech perception. *Cognition*, 64(3), B9-B17

PALLIER, C., COLOME, A., & SEBASTIAN-GALLÉS, (1999) : Phonological representations and repetition priming. *Proceedings of Eurospeech '99*, Budapest, Hungary, Sept. 5-9, 1999, vol. 4 (pp. 1907-1910)

PALLIER, C. (2000) : Word recognition : do we need phonological representations ? In Cutler, A. McQueen, J.M., & Zondervan, R. *Proceedings of the Workshop on Spoken Word Access Processes*, pp. 159-162, May 29-31, 2000, Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

PEEREMAN, R. & HOLENDER, D. (1990). La reconnaissance des mots dans les écritures non-alphabétiques, *European Bulletin of Cognitive Psycyhology*, 10 (3), 289-339.

PERFETTI Charles A. (University of Pittsburgh, USA) : Représentation et prise de conscience au cours de l'apprentissage de la lecture

PEYTARD (Jean) et GENOUVRIER (Emile) : *Linguistique et enseignement du français*, Larousse, Paris, 1970, p. 42

RIEBEN L. et PERFETTI C., 1989, L'apprenti lecteur, Delachaux et Niestlé.

SCHOLES R.J. (1998) - *Arguments contre la conscience phonique*,Actes de Lecture, Sept. 1998, 63, 15-22

SELVA, Thierry, CHANIER Thierry, Sciences et Techniques Educatives (STE), vol 7, 2, 2000. Editions Hermès : Paris pp 385-412

SPRENGER-CHAROLLES Liliane et CASALIS. S, Lire. Lecture et écriture : acquisition et troubles du développement. PUF 1996

TUNMER William E. (University of Western Australia, Australie) : Conscience phonologique et acquisition de la langue écrite

WEHRHEIM (Sylvia) et de VALS (Marie) : *Apprentissage de la lecture : activité de l'intelligence*, Toulous, 1976, 203 p, éditeur E. Privat, Collection Pragma.

WOOLDRIDGE Terence Russon, Université de Toronto, extrait de La Lexicographie française du XVIII au Xxème siècle, Paris, Klincksieck, 1988 : 305-13

12. TOILOGRAPHIE

On pourra trouver sur mon site les liens suivants, mis à jour le plus régulièrement possible, à l'adresse suivante : **http://linguistiques.muroni.free.fr**

12.1. LINGUISTIQUE

1. Prosodie et Phonologie, études comparées

http://www.icp.inpg.fr/jep2000/html/jep2000/node16.html

La phonologie (Queen's University Kingston)

http://qsilver.queensu.ca/french/Cours/215/chap3.html

L'émergence de la conscience phonologique, par Bruno De Cara

http://www.exco.ucl.ac.be/sblu/activites/emergence.htm

Rôle de la syllabe dans la perception de la parole (Christophe Pallier, École de Hautes Études en Sciences sociales, 1994)

http://www.ehess.fr/centres/lsc/persons/pallier/thesis/abstract.txt

La syllabation automatique du français parlé (N. Delbecque, E. Bas)

http://bach.arts.kuleuven.ac.be/elicop/elilap3.html

Les trois étapes de la lecture

http://www.offratel.nc/magui/Troisetap.htm

Richaudeau François, Les Actes de Lecture (revue de l'AFL) : réponse au livre de José Morais "L'art de la lecture&quot

http://lecture.org/actes/AL61/AL61LU0.html

Site officiel des sciences du langage

http://www.talou.com/

Marges.linguistiques

http://marges.linguistiques.free.fr/

Jean-Emile Gombert : lecture et phonologie

http://adapt-scol-franco.educ.infinit.net/themes/dile/documents/gombert.pdf

Les stratégies cognitives des bons et mauvais lecteurs (Jean-Paul Martinez)

http://www.er.uqam.ca/nobel/lire/textes/stratcognit.pdf

Les difficultés de lecture (Jean-Paul Martinez)

http://www.er.uqam.ca/nobel/lire/textes/difficultelect.pdf

Comment les enfants entrent dans la culture écrite (Jacques Bernardin) : note de lecture par Jean Foucambert (décembre 97)

http://www.lecture.org/actes/AL60/AL60LU3.html

Émergence de la parole (Revue Parole, sommaire 1999, octobre)

http://www.umh.ac.be/RPA/rpa1999-9-10.html

Fonctionnement de la parole : aspects physiologiques, acoustiques et perceptifs (Danielle Duez)

http://www.lpl.univ-aix.fr/lpl/presentation/equipes/fp.htm

Vocabulaire, comparaison de fréquences

http://hypermedia.univ-paris8.fr/jean/infolit/cours/infolit3.html

Développement du langage

http://www.fsj.ualberta.ca/beaudoin/ling/matern.htm

Les sons et l'écriture

http://www.limsi.fr/Individu/habert/Cours/PX/ProprietesDesLanguesArticle/node7.html

Quelques aspects du français d'aujourd'hui

http://www.france.sk/culturel/pedagaspects.htm

De la fréquence relative des phonèmes en français et de la relativité de ces fréquences

http://bach.arts.kuleuven.ac.be/elicop/elilap1.html

La syllabation automatique du français parlé

http://bach.arts.kuleuven.ac.be/elicop/elilap3.html

11.2. (NEURO)PSYCHOLOGIE

Revue canadienne de psychologie expérimentale (septembre 99) : mémoire implicite ; voisins orthographiques ; momentum représentationnel.

http://www.cpa.ca/cjep/cjep533.html

Technologies d'apprentissage et troubles d'apprentissage (bibliographie)

http://olt-bta.hrdc-drhc.gc.ca/publicat/bibldis_f.html

Étude des mécanismes d'accès à la signification de mots écrits chez des lecteurs sourds (Docteur Laurence Paire-Ficourt)

http://www.multimania.com/pch/theses98.htm

NeuroPsychologie Cognitive : Nathalie Bedoin, maître de conférences

http://unpc.univ-lyon2.fr/~bedoin/bedoin.html

L'apprentissage de la lecture et la langue des signes pour l'enfant sourd sévère ou profond (Gwënola Paranthoën)

http://perso.club-internet.fr/ltzgw/Memoire/Paranthoen.htm

Dysfonctionnements développementaux : essai de définition opérationnelle (Abdelhamid Khomsi, Docteur en Linguistique, Professeur de Psychologie)

http://www.coridys.asso.fr/pages/base_doc/txt_khomsi/txt.html

Transcodage numérique et transcodage lexical (Nathalie Duranteau)

http://palissy.humana.univ-nantes.fr/labos/cybele/doc-msh/duranteau.htm

Hyperactivité : troubles d'apprentissage

http://web.wanadoo.be/scarlett/hyperactivite/apprentissage.htm

Comment les enfants apprennent-ils à lire ?

http://www.chez.com/psychologue/Psychologue/lire.htm

Psychologie cognitive de la lecture : quelques aperçus...

http://perso.libertysurf.fr/resolution/

11.3. NEUROLINGUISTIQUE

Cerveau et langage : perception et compréhension du langage

http://www.isc.cnrs.fr/naz/naz2.htm

Sciences cognitives : le courrier du CNRS n°79

http://www.cnrs.fr/cw/fr/tous/sommaire/som79.html

Coridys, bibliographie

http://www.coridys.asso.fr/pages/biblio/biblio.html

Revue de neuropsychologie, sommaire du volume 6, 1996

http://criugm.qc.ca/revueneuropsy/vol6.html

11.4. ÉDUCATION NATIONALE

Principe alphabétique et lecture

http://www.lecture.org/principe.html

11.5. DIVERS

Notions optométriques

http://pages.globetrotter.net/assoqc/infovis.html

Apprendre à lire autrement (Association)

http://site.voila.fr/semeria/index.jhtml

La lecture en couleurs : une technologie appropriée (William Bernhardt)

http://perso.wanadoo.fr/une.education.pour.demain/articlesapfond/lecture/bhardt.htm

Pédagogie : les problèmes de l'éducation scolaire

http://www.arfe-cursus.com/

Questionnaire sur la lecture (Lise Bessette et Abdelkébir El Bina, groupe LIRE 2000, sous forme de quizz)

http://www.er.uqam.ca/nobel/lire/jeuquest/quizlect.html

Une approche phonétique en identification automatique des langues : la modélisation acoustique des systèmes vocaliques

www.ddl.ish-lyon.cnrs.fr/membres/~Pellegrino/resume.htm

Bien voir pour bien lire

http://www.inform-optique.com/

---

13. REMERCIEMENTS

Mes remerciements les plus sincères vont

à Monsieur Alain BENTOLILA, pour son regard expert et bienveillant,

à Monsieur Bruno GERMAIN, pour sa patience et ses remarques constructives,

à tous les enseignants qui ont participé patiemment à mes recherches (écoles La Ribambelle, Lalla Amina, Jeanne d'Arc, Subrini),

et à tous les élèves... qui ont tant à nous apprendre !

Synthèse pour l'habilitation à diriger des recherches

Décembre 2002

Jean-marc MURONI