



La lecture du CP au CM2 avec le test de l'alouette : que peut-on dire du niveau des élèves à quarante ans de distance ? : comment se déroule cet apprentissage (continu/discontinu) ?

Amélie Izard

► **To cite this version:**

Amélie Izard. La lecture du CP au CM2 avec le test de l'alouette : que peut-on dire du niveau des élèves à quarante ans de distance ? : comment se déroule cet apprentissage (continu/discontinu) ?. Éducation. Université Toulouse le Mirail - Toulouse II, 2013. Français. <NNT : 2013TOU20087>. <tel-00961020>

HAL Id: tel-00961020

<https://tel.archives-ouvertes.fr/tel-00961020>

Submitted on 19 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 2 Le Mirail (UT2 Le Mirail)

Cotutelle internationale avec :

Présentée et soutenue par :

Amélie IZARD

Le mercredi 2 octobre 2013

Titre :

La lecture du CP au CM2 avec le test de L'alouette :
Que peut-on dire du niveau des élèves à quarante ans de distance?
Comment se déroule cet apprentissage (continu/discontinu)?

ED CLESCO : Sciences de l'éducation

Unité de recherche :

Unité Mixte de Recherche en "Education, Formation, Travail, Savoirs" (EFTS)

Directeur(s) de Thèse :

Jacques FIJALKOW, professeur émérite en psychologie linguistique
Université Toulouse II

Rapporteurs :

Isabelle DELCAMBRE, professeur émérite en sciences de l'éducation, Université Lille 3
Jean-Marie BESSE, professeur émérite en psychologie cognitive, Université Lyon 2

Autre(s) membre(s) du jury :

Christine MIAS, professeur en sciences de l'éducation, Université Toulouse II
Michel GRANDATY, professeur à l'école interne IUFM, Université Toulouse II
Monique SENECHAL, professeur en psychologie du développement, Université de Carleton

REMERCIEMENTS

Je souhaiterais avant tout remercier mon directeur de recherche, **Jacques FIJALKOW**, professeur émérite en psychologie linguistique et anciennement directeur de l'équipe EURED (Equipe universitaire de Recherche en Education et Didactique) au sein de l'université de Toulouse II, pour m'avoir encouragée à effectuer ce doctorat. Je vous remercie donc pour la confiance que vous m'avez accordée, pour votre soutien et votre disponibilité.

Je remercie également **l'école doctorale CLESCO** (Comportement, Langage, Education, Socialisation, Cognition) pour m'avoir fait confiance en me donnant l'occasion de bénéficier d'une allocation de recherche pendant trois ans, qui m'a ainsi permis de me consacrer pleinement à ce travail de recherche.

Merci aux membres de l'ancienne **équipe EURED** avec qui j'ai eu plaisir à échanger lors de nos séminaires hebdomadaires. Merci à ceux qui ont pris le temps à chaque fois de lire mes travaux et d'y apporter des corrections. Merci à vous pour les échanges constructifs.

Par la même occasion, j'adresse une mention particulière à **Serge RAGANO** et **Laurence PASA**, pour leurs conseils avisés et leur disponibilité.

J'adresse aussi mes remerciements à **Monique SENECHAL**, professeur en psychologie du développement à l'Université de Carlton, Ottawa, **Jean-Marie BESSE**, professeur en psychologie cognitive à l'Université de Lyon 2, **Isabelle DELCAMBRE**, professeur émérite en Sciences de l'éducation à l'Université de Lille 3, **Christine MIAS**, professeur en sciences de l'éducation à l'Université de Toulouse II et **Michel GRANDATY**, professeur à l'école interne IUFM à l'Université Toulouse II, pour l'intérêt qu'ils ont porté à mon travail.

Cette recherche n'aurait pu se réaliser sans la participation des **enfants**. C'est pourquoi, je leur adresse un grand merci pour avoir bien voulu participer à cette expérience. Un grand merci aussi aux **enseignants** pour m'avoir accueillie dans vos classes.

Je tiens à remercier chaleureusement ma famille pour m'avoir soutenue durant cette longue et rude étape.

De la même façon, je remercie mes amis qui par leur soutien m'ont permis de mener à terme ce travail. Je remercie en particulier **Aurélie**, **Gwen**, **Jess**, **Mandi**, **Marie**, **Sarah**, **Vincent T.** et **Vincent F.** pour leur présence. Merci pour vos encouragements, votre compréhension et merci pour tous les bons moments passés ensemble. Tout simplement merci d'avoir été et d'être toujours là.

Merci à mes collègues **Carine** et **Darine**, maintenant devenues amies. Vos conseils d'anciennes « thésardes » m'ont précieusement aidée. Merci aussi pour votre bonne humeur et pour tous ces moments partagés.

Enfin, un remerciement tout particulier à **Mathieu** pour avoir été à mes côtés durant cette dernière année qui n'a pas été des plus faciles... Je te remercie donc pour ta patience et ta compréhension. Merci pour ta capacité d'adaptation qui m'a permis de finaliser cette recherche. Merci pour ton soutien et pour avoir toujours cru en moi.

Résumé

Dans un contexte idéologique où la question de la baisse du niveau revient de façon récurrente dans l'espace public, nous nous sommes servie du test de *L'alouette* (Lefavrais, 1967) pour mener une étude comparative du niveau en lecture des élèves du CP au CM2. Les résultats indiquent qu'à quarante ans d'écart, le niveau en lecture a diminué de six mois. Au-delà de cette baisse relative, apparaît un fait tout aussi intéressant : celui d'un accroissement des écarts. On constate alors un décrochage des élèves les plus en difficulté et un décollage des élèves les plus performants. Le test de *L'alouette* étant basé sur une conception traditionnelle de la lecture, nous y avons vu là, l'occasion de mettre à l'épreuve les modèles dominants en lecture : les modèles développementaux. De cette façon, nous nous sommes demandée dans quelle mesure le test de *L'alouette* nous informe quant à l'apprentissage de la lecture et plus particulièrement en ce qui concerne le caractère discontinu ou continu de la lecture. Les résultats obtenus semblent aller dans le sens d'une discontinuité certes, mais qui ne serait pas absolue. Ainsi, dès le début de l'apprentissage les élèves utiliseraient une procédure lexicale en plus de la procédure par médiation phonologique. L'une ou l'autre étant dominante à certains moments de l'apprentissage.

Mots clés : apprentissage de la lecture, niveau, discontinuité, continuité

Abstract

The fall of the level of ability in literacy has often been brought to public attention, so we used the *Alouette* test (Lefavrais, 1967) to make a comparative study on the levels of literacy in students aged from 2nd year to 6th year. The study indicates that in forty years the level of literacy in students has reduced by six months. In addition to this decline, there also appeared to be widening gaps between the two extremes of ability. Those who experience the most difficulty become disengaged whereas those with the highest level of ability take off. As the *Alouette* test is based on the traditional concept of literacy, this allowed us to test the other dominant developmental theories. In this way we were able to question whether the *Alouette* test could tell us about the process in which children learn how to read whether continuously or discontinuously. The results seem to show that the tendency is for the discontinuity, but not entirely. Hence, as soon as children learn to read they would use a lexical process in addition to using phonological mediation, both being dominant at different times during their education.

Key words: reading acquisition, level, discontinuity, continuity

Table des matières

Remerciements.....	1
Résumé	3
Table des matières.....	4
Listes des tableaux, figures et cartographies.....	6
INTRODUCTION GÉNÉRALE.....	15
ÉTAT DE LA QUESTION	23
Chapitre 1 : Le niveau en lecture des élèves.....	24
Chapitre 2 : L'apprentissage de la lecture.	50
QUESTIONS DE RECHERCHE.....	76
MÉTHODOLOGIE	79
ANALYSE DES RÉSULTATS	89
Partie 1 : Le niveau en lecture des élèves du CP au CM2.	90
Chapitre 3 : Quarante ans plus tard où en est la lecture ? Comparaison des résultats au test de <i>L'alouette</i> d'enfants du CP au CM2.	91
Chapitre 4 : Évolution de la dispersion à 40 ans de distance.....	103
Chapitre 5 : Le niveau en lecture en 2009.	113
CONCLUSION	123
Partie 2 : L'apprentissage de la lecture des élèves du CP au CM2.	126
<i>Études préliminaires.....</i>	<i>129</i>
Chapitre 6 : La performance lexicale.....	130
Chapitre 7 : Types de réponses erronées et processus de lecture.....	138

Chapitre 8 : Etude des mots qui ont posé le plus de difficultés de lecture.	146
<i>L'influence des variables linguistiques</i>	165
Chapitre 9 : L'influence de la longueur et de la fréquence d'un mot.	166
Chapitre 10 : L'influence de la structure graphémique d'un mot.	195
CONCLUSION	224
<i>L'influence des variables liées au test de L'alouette</i>	228
Chapitre 11 : L'influence des dessins lors de la lecture du texte de <i>L'alouette</i>	235
Chapitre 12 : L'influence des similitudes lors de la lecture du texte de <i>L'alouette</i>	259
Chapitre 13 : L'influence du co-texte sémantique lors de la lecture du texte de <i>L'alouette</i>	284
Chapitre 14 : L'influence des mots surdéterminés lors de la lecture du texte de <i>L'alouette</i>	303
CONCLUSION	319
Chapitre 15 : Processus de lecture et évolution du CP au CM2	321
CONCLUSION GÉNÉRALE-DISCUSSION.....	334
BIBLIOGRAPHIE	352
ANNEXES	CDRom

Liste des tableaux, figures et cartographies

Cartographies

Cartographie 1 : test de <i>L'alouette</i> (Lefavrais, 1965)	81
Cartographie 2 : mots longs peu fréquents dans le texte de <i>L'alouette</i>	170
Cartographie 3 : mots simples et complexes dans le texte de <i>L'alouette</i>	199
Cartographie 4 : test de <i>L'alouette</i> (Lefavrais, 1965)	230
Cartographie 5 : erreurs dues aux dessins présents autour du texte de <i>L'alouette</i>	238
Cartographie 6 : mots sur lesquels les élèves ont commis une erreur du fait des similitudes dans le texte de <i>L'alouette</i>	263
Cartographie 7 : mots sur lesquels les élèves ont commis une erreur à cause du co-texte sémantique dans le texte de <i>L'alouette</i>	287
Cartographie 8 : mots surdéterminés dans le test de <i>L'alouette</i>	306

Figures

Figure 1 : les différents types d'évaluations en France en 2010	26
Figure 2 : modèle de Gough (1972)	54
Figure 3 : modèle de Rumelhart (1977)	55
Figure 4 : détail de l'échantillonnage	87

Graphiques

Graphique 1 : médianes des âges théoriques et effectifs de l'ensemble des élèves du public hors EP, selon le niveau scolaire en 1965 et 2009.....	97
--	----

Graphique 2 : médianes des âges théoriques et effectifs de l'ensemble des élèves du public avec écoles en EP, selon le niveau scolaire en 1965 et 2009	100
Graphique 3 : médianes, quartiles et valeurs extrêmes des âges de lecture en 1965 et 2009	107
Graphique 4 : écarts-types en 1965 et 2009 selon le niveau scolaire.....	107
Graphique 5 : étendue en 1965 et 2009 selon le niveau de scolarisation.....	108
Graphique 6 : âges de lecture en mois selon le genre et la classe fréquentée en 2009	117
Graphique 7 : âges de lecture en mois selon le genre et l'établissement fréquenté en 2009	118
Graphique 8 : âges moyens de lecture selon la classe et le type d'établissement.....	120
Graphique 9 : pourcentages d'erreurs selon la longueur et la fréquence des mots en cycle 2 et cycle 3.....	175
Graphique 10 : pourcentages d'erreurs commises selon la longueur et la fréquence des mots.....	176
Graphique 11 : processus de traitement des mots en CP selon leur fréquence et leur longueur	180
Graphique 12 : processus de traitement des mots en CE1 selon leur fréquence et leur longueur	182
Graphique 13 : processus de traitement des mots en CE2 selon leur fréquence et leur longueur	184
Graphique 14 : processus de traitement des mots en CM1 selon leur fréquence et leur longueur	185
Graphique 15 : processus de traitement des mots en CM2 selon leur fréquence et leur longueur	187
Graphique 16 : proportion d'erreurs de décodage par classe selon la longueur et la fréquence des mots	187
Graphique 17 : évolution des processus de lecture des mots longs peu fréquents (ou rares) selon la classe	188

Graphique 18 : évolution de la proportion d'erreurs entraînant une verbalisation de non-mots ou de mots lors de la lecture de mots longs peu fréquents (ou rares)	190
Graphique 19 : évolution des processus de lecture des mots courts fréquents (ou très fréquents) selon la classe	190
Graphique 20 : évolution de la proportion d'erreurs entraînant une verbalisation de non-mots ou de mots lors de la lecture de mots courts fréquents (ou très fréquents)	191
Graphique 21 : pourcentages d'erreurs commises selon la complexité des mots et le cycle de scolarisation	201
Graphique 22 : pourcentages d'erreurs commises selon la complexité des mots et la classe	202
Graphique 23 : processus hypothétiques de lecture selon la complexité des mots en CP	207
Graphique 24 : processus hypothétiques de lecture selon la complexité des mots en CE1 ...	209
Graphique 25 : processus hypothétiques de lecture selon la complexité des mots en CE2 ...	211
Graphique 26 : processus hypothétiques de lecture selon la complexité des mots en CM1	214
Graphique 27 : processus hypothétiques de lecture selon la complexité des mots en CM2	215
Graphique 28 : proportions d'erreurs de décodage selon la structure graphémique des mots du CP au CM2	216
Graphique 29 : évolution des processus de lecture des mots complexes selon la classe	217
Graphique 30 : évolution de la proportion d'erreurs entraînant une verbalisation de non-mots ou de mots lors de la lecture de mots complexes	219
Graphique 31 : évolution des processus de lecture des mots simples selon la classe	219
Graphique 32 : évolution de la proportion d'erreurs entraînant une verbalisation de non-mots ou de mots lors de la lecture de mots simples	220
Graphique 33 : évolution de la proportion d'élèves ayant été influencés au moins une fois par un dessin entre le CP et le CM2	242

Graphique 34 : évolution de la proportion d'élèves ayant été influencés au moins une fois par un dessin entre le CP et le CM2 dans le début du texte.....	247
Graphique 35 : évolution de la proportion d'élèves ayant été influencés au moins une fois par un dessin entre le CE1 et le CM2 dans le reste du texte.....	247
Graphique 36 : évolution de la proportion d'élèves qui a commis au moins une erreur due aux similitudes	268
Graphique 37 : évolution de la proportion d'élèves qui a commis plusieurs erreurs dues aux similitudes.....	269
Graphique 38 : évolution de la proportion d'élèves ayant commis au moins une erreur due à une similitude dans le début du texte entre le CP et le CM2	273
Graphique 39 : évolution de la proportion d'élèves ayant commis au moins une erreur due à une similitude dans le reste du texte entre le CE1 et le CM2	273
Graphique 40 : évolution de la proportion d'élèves ayant commis au moins une erreur due au co-texte sémantique entre le CP et le CM2.....	292
Graphique 41 : évolution de la proportion d'élèves ayant commis plusieurs erreurs dues au co-texte sémantique entre le CP et le CM2	293
Graphique 42 : évolution de la proportion d'élèves ayant commis au moins une erreur due au co-texte sémantique dans le début du texte entre le CP et le CM2	296
Graphique 43 : évolution de la proportion d'élèves ayant commis au moins une erreur due au co-texte sémantique dans le reste du texte entre le CP et le CM2	296
Graphique 44 : évolution de la proportion d'élèves qui a commis au moins une erreur due aux mots surdéterminés.....	311
Graphique 45 : évolution de la proportion d'élèves qui a commis plusieurs erreurs dues aux mots surdéterminés	311
Graphique 46 : évolution de la proportion d'élèves qui a commis au moins une erreur due aux mots surdéterminés dans le reste du texte.....	313
Graphique 47 : évolution de la proportion d'élèves qui a commis plusieurs erreurs dues aux mots surdéterminés dans le reste du texte.....	314
Graphique 48 : évolution des processus de lecture entre le CP et le CM2	334

Tableaux

Tableau 1 : les différents types d'évaluation en lecture en France en 2010.....	32
Tableau 2 : répartition des élèves selon la classe et le type d'établissement	80
Tableau 3 : nombre d'élèves par type d'établissement selon le niveau scolaire	86
Tableau 4 : nombre d'élèves par type d'établissement selon le niveau scolaire	93
Tableau 5 : médianes en mois des âges théoriques et effectifs de l'ensemble des élèves du public hors EP, pour la lecture du texte ou des « syllabes » en 1965 et 2009	95
Tableau 6 : médianes des âges théoriques et effectifs de l'ensemble des élèves du public (avec écoles en EP), pour la lecture du texte ou des syllabes en 1965 et 2009	98
Tableau 7 : nombre d'élèves par type d'établissement selon le niveau scolaire	105
Tableau 8 : nombre d'élèves par type d'établissement selon le niveau scolaire	115
Tableau 9 : indices moyens de performance lexicale en fonction du nombre de mots parcourus et correctement lus et selon le niveau scolaire.....	131
Tableau 10 : mots qui ont entraîné le plus de difficultés de lecture et pourcentage d'élèves qui ont commis une erreur sur ces mots (par rapport au nombre d'élèves qui ont lu ce mot).....	147
Tableau 11 : mots sur lesquels au moins la moitié des élèves de CP ont commis une erreur de lecture	148
Tableau 12 : mots sur lesquels au moins la moitié des élèves de CE1 ont commis une erreur de lecture	151
Tableau 13 : mots sur lesquels au moins la moitié des élèves de CE2 ont commis une erreur de lecture	156
Tableau 14 : mots sur lesquels au moins la moitié des élèves de CM1 ont commis une erreur de lecture	160
Tableau 15 : mots sur lesquels au moins la moitié des élèves de CM2 ont commis une erreur de lecture	162

Tableau 16 : pourcentages d’erreurs commises selon la longueur et la fréquence des mots et le cycle de scolarisation	172
Tableau 17 : pourcentages d’erreurs commises selon la longueur et la fréquence des mots et la classe	174
Tableau 18 : pourcentages d’erreurs commises selon la complexité des mots et le cycle de scolarisation	200
Tableau 19 : pourcentages d’erreurs commises selon la complexité des mots et la classe	200
Tableau 20 : pourcentage et nombre d’élèves ayant été influencés ou non par une image selon le cycle de scolarisation.....	239
Tableau 21 : pourcentage et nombre d’élèves ayant été influencés une ou plusieurs fois par un dessin selon le cycle de scolarisation	239
Tableau 22 : pourcentage et nombre d’élèves ayant été influencés ou non par un dessin selon la classe en cycle 2	240
Tableau 23 : pourcentage et nombre d’élèves ayant été influencés une ou plusieurs fois par un dessin selon la classe en cycle 2	241
Tableau 24 : pourcentage et nombre d’élèves ayant été influencés ou non par un dessin selon la classe en cycle 3	241
Tableau 25 : pourcentage (et nombre) d’élèves ayant été influencés une ou plusieurs fois par un dessin selon la classe en cycle 3	242
Tableau 26 : pourcentage et nombre d’élèves ayant été influencés ou non par un dessin dans le début du texte selon la classe.....	244
Tableau 27 : pourcentage et nombre d’élèves ayant été influencés une ou plusieurs fois par un dessin dans le début du texte selon la classe	244
Tableau 28 : pourcentage et nombre d’élèves ayant été influencés ou non par un dessin dans le reste du texte selon la classe	245
Tableau 29 : pourcentage et nombre d’élèves ayant été influencés une ou plusieurs fois par un dessin dans le reste du texte selon la classe.....	246
Tableau 30 : mots sur lesquels le plus grand pourcentage d’élèves (nombre) a commis une erreur due aux dessins selon le niveau de scolarisation.....	248

Tableau 31 : erreurs susceptibles d'être dues à la présence de similitudes.....	262
Tableau 32 : pourcentage et nombre d'élèves ayant commis ou non des erreurs dues à une similitude selon le cycle de scolarisation.....	264
Tableau 33 : pourcentage et nombre d'élèves ayant commis une ou plusieurs erreur(s) due(s) à une similitude selon le cycle de scolarisation.....	265
Tableau 34 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude selon la classe en cycle 2	265
Tableau 35 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude selon la classe en cycle 2	266
Tableau 36 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude selon la classe en cycle 3	267
Tableau 37 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude selon la classe en cycle 3	267
Tableau 38 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude dans le début du texte selon la classe	270
Tableau 39 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude dans le début du texte selon la classe.....	271
Tableau 40 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude dans le reste du texte selon la classe.....	272
Tableau 41 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude dans le reste du texte selon la classe.....	272
Tableau 42 : mots sur lesquels le plus grand nombre d'élèves a commis une erreur due à une similitude selon le niveau de scolarisation.....	274
Tableau 43 : types d'erreurs commis à cause du co-texte sémantique.....	286
Tableau 44 : pourcentage et nombre d'élèves ayant commis ou non des erreurs dues au co-texte sémantique selon le cycle de scolarisation	288
Tableau 45 : pourcentage et nombre d'élèves ayant commis une ou plusieurs erreur(s) due(s) au co-texte sémantique selon le cycle de scolarisation.....	288

Tableau 46: pourcentage (et nombre d'élèves) ayant été influencés ou non par le co-texte sémantique selon la classe en cycle 2	289
Tableau 47 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par le co-texte sémantique selon la classe en cycle 2	290
Tableau 48 : pourcentage et nombre d'élèves ayant été influencés ou non par le co-texte sémantique selon la classe en cycle 3	291
Tableau 49 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par la sémantique selon la classe en cycle 3	291
Tableau 50 : pourcentage et nombre d'élèves ayant été influencés ou non par le co-texte sémantique dans le début du texte selon la classe	294
Tableau 51 : pourcentage et nombre d'élèves ayant été influencés ou non par le co-texte sémantique dans le reste du texte selon la classe.....	295
Tableau 52 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par le co-texte sémantique dans le reste du texte selon la classe.....	295
Tableau 53 : mots dont la lecture a été la plus influencée par le co-texte sémantique pour chaque niveau scolaire.....	297
Tableau 54 : mots dont la lecture peut être influencée par plusieurs variables du test de <i>L'alouette</i>	305
Tableau 55 : pourcentage et nombre d'élèves ayant commis ou non des erreurs dues aux mots surdéterminés selon le cycle de scolarisation	307
Tableau 56 : pourcentage et nombre d'élèves ayant commis une ou plusieurs erreur(s) due(s) aux mots surdéterminés selon le cycle de scolarisation.....	307
Tableau 57 : pourcentage et nombre d'élèves ayant été influencés ou non par les mots surdéterminés selon la classe en cycle 2.....	308
Tableau 58 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par les mots surdéterminés selon la classe en cycle 2	309
Tableau 59 : pourcentage et nombre d'élèves ayant été influencés ou non par les mots surdéterminés selon la classe en cycle 3	309
Tableau 60 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par les mots surdéterminés selon la classe en cycle 3	310

Tableau 61 : pourcentage et nombre d'élèves ayant été influencés ou non par les mots surdéterminés dans le reste du texte selon la classe	312
Tableau 62 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par les mots surdéterminés dans le reste du texte selon la classe	313
Tableau 63 : mots dont la lecture a été la plus influencée par les mots surdéterminés pour chaque niveau scolaire.....	314
Tableau 64 : proportion d'élèves qui ont eu recours aux différents processus selon la classe.....	324

INTRODUCTION GENERALE

Plan de l'introduction

I. La place de la lecture dans la société depuis le XVe siècle	15
1. La lecture de l'Ancien Régime à la Révolution: outil d'instruction religieuse et morale	16
<i>D'une lecture religieuse.....</i>	<i>16</i>
<i>... à une lecture distraction</i>	<i>16</i>
2. La lecture sous la IIIe République : un pré-requis nécessaire à l'éducation civique et morale.....	17
<i>La lecture obligatoire pour tous.....</i>	<i>17</i>
<i>La lecture à l'école devient une priorité.....</i>	<i>18</i>
3. La lecture au XXe siècle : entre valorisation et inquiétudes	18
II. Objectifs de la recherche	20

La lecture en France, aujourd'hui à la base de toutes les acquisitions scolaires fait débat depuis maintenant plusieurs années. Son statut, tout comme sa finalité ont changé en même temps que la société a évolué. Bien que l'on puisse ramener le début de la lecture à la naissance de l'écriture, cette pratique a commencé à réellement se développer lorsque le phénomène d'alphabétisation a été entrepris grâce à l'invention de l'imprimerie. Le passage du livre manuscrit au livre imprimé au milieu du XVe siècle a ainsi permis à un plus grand nombre de personnes d'accéder à la lecture. Grâce aux techniques de fabrication et de diffusion, la lecture, au rythme de l'alphabétisation, s'est développée au cours du temps.

I. La place de la lecture dans la société depuis le XVe siècle

1. La lecture de l'Ancien Régime à la Révolution : outil d'instruction religieuse et morale

D'une lecture religieuse...

Le phénomène d'alphabétisation de masse commence donc sous l'Ancien Régime lors des conflits de religion (la Réforme protestante au XVe et XVIe siècle et la Contre-Réforme catholique au XVIe siècle). En effet, tous les laïcs doivent être alphabétisés afin que l'on puisse leur inculquer la religion, et l'apprentissage de la lecture doit leur permettre d'accéder aux textes religieux. L'instruction en tant qu'instrument de propagande se développe et son objectif est l'endoctrinement. La lecture, alors considérée comme un outil de diffusion, va peu à peu se développer grâce à la création d'écoles. Ainsi, dans ce contexte, de nouveaux ordres religieux dédiés à l'instruction chrétienne sont créés (le plus connu est celui des Frères des Ecoles chrétiennes créée en 1684 par Jean-Baptiste de la Salle). En 1685, suite à l'édit de Nantes, Louis XIV souhaitant assurer la mainmise de la religion d'Etat, remplace les écoles protestantes par de nombreuses écoles catholiques. Durant cette période, l'alphabétisation se développe avant tout dans les villes et la lecture est réservée aux clercs et à la noblesse. Pour la majorité de la population qui est paysanne et analphabète, l'accès au livre s'effectue donc grâce à un intermédiaire.

... à une lecture distraction

La Révolution rompt avec tout ce qui avait été mis en place sous l'Ancien Régime. Les privilèges du clergé et de la noblesse sont supprimés. L'éducation devient progressivement

une affaire d'Etat et non plus d'Eglise, et la lecture va peu à peu se démocratiser. Ainsi, afin d'encourager tous les Français à lire, le nombre d'écoles va considérablement augmenter, des manuels scolaires vont être rédigés puis diffusés et des bibliothèques vont se développer dans les communes. La lecture, outre le fait qu'elle permet de s'instruire est maintenant pratiquée pour se distraire (romans, textes philosophiques, *etc.*). L'alphabétisation des masses va donc se développer durant le XIX^{ème} siècle et on parle d'une révolution de la lecture « Ils lisent trop, ils lisent n'importe comment, ils lisent n'importe quoi »¹.

Ainsi, en 1833, Guizot estimant que « *l'instruction primaire est désormais une des garanties de l'ordre et de la stabilité sociale* »², encourage la création d'écoles dans les communes de plus de 500 habitants. Il organise l'enseignement primaire au profit des classes populaires autour de deux principes : la liberté de l'enseignement et son organisation qui s'effectue en deux temps : d'abord élémentaire, puis supérieur. Afin d'éduquer l'enfant à la citoyenneté, il précise que l'enseignement élémentaire des écoles, qu'elles soient publiques ou privées, comprend nécessairement « *l'instruction morale et religieuse, la lecture, l'écriture, les éléments de la langue française et du calcul, le système légal des poids et mesures.* » (cf. Art. 1). L'éducation religieuse n'est donc plus la seule finalité de l'enseignement. En 1867, Duruy contribue à développer l'enseignement primaire en encourageant la gratuité de l'éducation pour les pauvres.

2. La lecture sous la III^e République : un pré-requis nécessaire à l'éducation civique et morale

La lecture obligatoire pour tous

Sous la III^{ème} République, les lois Ferry instituent un enseignement laïc, gratuit et obligatoire pour tous,³ rompant ainsi avec l'éducation antérieure (*i.e.* doctrinale). L'instruction publique devient une priorité. L'objectif n'est plus d'inculquer la religion mais d'instruire et de former le citoyen. Il est désormais question « *d'assurer à l'enfant tout le savoir pratique dont il aura besoin dans la vie* » et de « *former son esprit* ». La lecture, bien qu'elle soit devenue obligatoire, n'occupe pas encore une place centrale dans l'instruction :

¹ De F. Barbier et C. Bertho Lavenir « *Histoire des médias, de Diderot à Internet* » cité par A-M Chartier et J. Hébrard dans « *Discours sur la lecture (1880-2000)* »

² Circulaire du 18 juillet 1833

³ Loi Ferry sur la gratuité des écoles primaires publiques le 16 juin 1881.

Loi Ferry sur l'enseignement primaire laïque et obligatoire le 28 mars 1882.

elle est considérée comme un outil nécessaire pour initier les enfants à l'acquisition des connaissances et à l'éducation intellectuelle, morale et civile. Ainsi, peu de place est consacrée à son enseignement et on privilégie les savoirs historiques, géographiques et scientifiques. Ainsi, alors que sous l'Ancien Régime, la finalité de la lecture était religieuse, elle devient sociale avec la III^e République. Désormais, on ne lit plus pour apprendre le catéchisme mais pour s'instruire, s'éduquer et comprendre.

A la fin du XIX^{ème} siècle, la lecture considérée comme un outil n'occupe donc pas encore une place centrale dans l'instruction : « *Elles se situent (avec l'écriture) dans une zone intermédiaire entre le psychique et le physiologique, le corporel et le cognitif, le geste et l'activité mentale, le savoir-faire et le savoir* » (Chauveau G., 2011).

Le rattachement de l'apprentissage de la lecture à l'école élémentaire en 1887 constitue le point de départ de la scolarisation et la lecture, considérée jusqu'alors comme un outil, va davantage être envisagée comme un pré-requis scolaire.

La lecture à l'école devient une priorité

A partir de 1923, l'apprentissage de la lecture devient une priorité de l'école: « *L'enfant ne peut rien apprendre s'il ne sait pas lire* » (I.O, 1923) Avec l'écriture, elle devient une base indispensable à toute éducation scolaire : « *L'enseignement de la lecture sert à deux fins. Il met entre les mains de l'enfant l'un des deux outils - l'autre étant l'écriture - indispensables à toute éducation scolaire. Il lui donne le moyen de s'initier à la connaissance de la langue et de la littérature française.* » (I.O, 1923)

3. La lecture au XX^e siècle : entre valorisation et inquiétudes

Jusqu'aux années 1960, la lecture, devenue un sujet de réflexion pédagogique, va être valorisée.

En 1959, la réforme scolaire prolonge l'âge de la scolarité obligatoire jusqu'à 16 ans et, au début des années soixante, le système scolaire rapproche les deux composantes de l'enseignement français (l'école du peuple et l'école des notables) pour les fondre en une seule institution. L'arrivée en 6^e de tous les enfants met en évidence une grande inégalité dans les acquisitions scolaires selon le milieu social et scolaire auxquels ils appartiennent. Dès lors, la notion d'échec scolaire apparaît et on la met rapidement en relation avec les difficultés

d'apprentissage de la lecture au cours préparatoire. Les discours changent : on passe du « ils lisent trop » à « ils ne lisent plus ». L'école passe donc au cœur des débats et sa finalité change. L'objectif de l'instruction n'est plus d'inculquer à l'enfant tous les savoirs dont il aura besoin dans la vie, mais de le préparer à une longue scolarité pour une meilleure insertion professionnelle. Le savoir-lire devient donc un indicateur de l'aptitude à une scolarité secondaire. Par conséquent, la finalité même de la lecture change. Désormais, on ne lit plus pour se distraire ou s'éduquer mais pour se former. La maîtrise de la lecture devient un enjeu national et apparaît comme la « *pierre de touche de la démocratisation de l'enseignement* » (Chartier et Hébrard, 2000).

A partir de 1964, le thème de la lecture à l'école suscite inquiétudes et questionnements. La lecture devient difficilement définissable et on s'interroge sur sa finalité : que faut-il lire ? Pourquoi faut-il lire ?

Parallèlement, la dyslexie devient la « maladie » du siècle et relance la controverse pédagogique (méthode globale vs méthode syllabique).

L'arrivée de la linguistique dans les années 1970 dans le champ de la pédagogie de l'apprentissage fait émerger de nouvelles théories sur la lecture. Alors que jusqu'ici on s'intéressait aux finalités de la lecture, on s'interroge maintenant sur les moyens et les méthodes les plus efficaces pour favoriser cet apprentissage. L'acte de lire devient un objet de recherche scientifique à part entière et de nombreuses disciplines vont s'en emparer (neurologie, psychologie, sociologie, linguistique, etc.).

Dans ce contexte, où la maîtrise de la lecture est devenue capitale, la nécessité de repérer les échecs des enfants, s'impose dès les premières années de l'apprentissage. Dans cette optique, de nombreux dispositifs d'évaluations, nationales et internationales, vont être mis en place.

En 1981, l'association caritative ATD Quart Monde fait apparaître une nouvelle notion : l'illettrisme. A l'époque, elle qualifie ceux qui ont suivi en totalité ou en partie le cycle de l'école primaire française sans y acquérir les compétences requises. En 1984, la notion est officialisée par le Gouvernement (rapport « Des illettrés en France ») et le Groupe interministériel permanent de lutte contre l'illettrisme (GPLI) est créé afin de résoudre le problème. Dans la période qui suit, on cherche à mesurer l'ampleur de ce phénomène et on s'interroge sur ses causes.

Par ailleurs, l'arrivée de nouveaux médias au XXe siècle (la radio, la télévision et Internet) et l'augmentation du temps de loisirs des Français portent eux aussi un coup à la lecture.

A partir du milieu du XXe siècle, la lecture, dont on ne cesse de prôner l'importance, connaît donc une « crise » qui engendre de nombreux débats tant sur sa finalité que sur son enseignement : Qu'est-ce que savoir lire ? Qu'est-ce que lire ? Comment faut-il l'enseigner ?

II. Objectifs de la recherche

La lecture n'a donc rien d'un invariant. Ce bref historique nous permet de voir que sa place et son statut au sein de notre société ont évolué au cours des siècles en même temps que la société a changé. Tantôt considérée comme un outil, comme un pré-requis et/ou comme une pratique, elle fut successivement employée à des fins religieuses, morales, culturelles et sociales.

A partir des années soixante, elle devient un champ de recherche à part entière. L'instruction de masse et l'unification de l'école font apparaître un nouveau sujet de discorde, celui d'un échec scolaire dans l'apprentissage de la lecture.

Aujourd'hui, c'est l'insertion professionnelle des jeunes qui prime et, entre une valorisation inconditionnelle de la lecture et une inquiétude sociale relative des non-lecteurs dont certains estiment que le nombre s'accroît, la lecture continue de faire débat.

Dans nos sociétés, où l'écrit est omniprésent, l'acquisition de la langue écrite est devenue plus que nécessaire. C'est un enjeu majeur car, dans les pays développés, une bonne maîtrise de la langue est considérée comme une condition essentielle d'intégration sociale, culturelle et économique. La non-maîtrise de cette habileté préoccupe donc de plus en plus. Elle est devenue un sujet de débat récurrent dans l'espace public et au sein de l'Education nationale. De fait, régulièrement le système éducatif français est pointé du doigt par des politiques, des journalistes, des enseignants, des parents, et la lecture n'est pas épargnée. Les nombreuses enquêtes menées à ce sujet le prouvent et régulièrement les polémiques médiatiques évoquent la question du niveau de compétence dans les savoirs de base parmi lesquels le savoir lire-écrire est central (*cf.* chapitre1).

Ainsi, afin de nous faire notre propre idée quant à l'évolution du niveau en lecture, nous avons fait le choix d'utiliser le test de *L'alouette* qui nous permet d'effectuer une comparaison à 40 ans d'écart (cf. chapitres 3 et 4). Au-delà d'une baisse du niveau, les études menées à ce sujet font état d'une inégalité dans les performances en lecture. Partant de ce constat, nous avons donc tenté de voir si ces inégalités apparaissent aussi dans notre échantillon (cf. chapitre 5).

Néanmoins, pour juger du niveau en lecture, il faut d'abord définir ce qu'on entend par apprendre à lire et savoir lire (cf. chapitre 2). De la même façon qu'il existe plusieurs manières de mesurer le niveau en lecture, il existe plusieurs façons de définir l'acte lexique et de concevoir l'apprentissage de la lecture.

Dans les modèles les plus reconnus, l'acquisition de la lecture, considérée à travers les processus d'identification du mot, reposerait sur une progression prenant en compte différents aspects de la langue. Cette progression est alors envisagée de manière discontinue puisque l'enfant traverserait une série de stades (logographique, alphabétique, orthographique) durant cet apprentissage. L'intérêt de ces modèles réside dans le fait qu'ils ont permis de mettre en évidence le rôle majeur de la phase alphabétique dans l'apprentissage de la lecture.

Dans ce contexte, le test de *L'alouette* (Lefavrais, 1967) est particulièrement représentatif puisqu'il repose sur une conception traditionnelle de la lecture et mesure essentiellement le niveau de décodage. Les erreurs commises lors de la lecture de ce texte traduiraient alors une insuffisance dans les capacités de décodage. Nous y voyons donc ici l'occasion de tester les modèles développementaux. Par conséquent, nous nous sommes demandée dans quelle mesure les erreurs commises lors de la lecture du texte de *L'alouette* nous informent quant au caractère discontinu ou continu de la lecture. Dans cette optique, nous avons dans un premier temps mené trois études préliminaires (cf. chapitres 6,7 et 8). La première consiste à analyser la performance lexicale des élèves. En effet, avant d'étudier les difficultés rencontrées en lecture, nous avons tenu à voir ce que les élèves savent lire (cf. chapitre 6). Ce n'est qu'à partir des chapitres suivants que nous nous intéressons aux difficultés rencontrées lors de la lecture du texte de *L'alouette* (cf. chapitres 7,8, 9, 10). Notre analyse portant sur les erreurs, nous avons ensuite réalisé une typologie des erreurs et processus de lecture identifiés lors de la lecture du texte (cf. chapitre 7). Une fois les types d'erreurs identifiés et les processus de lecture répertoriés, nous nous sommes demandée quels sont les mots du texte qui ont entraîné le plus de difficulté de lecture et pour quelles raisons (cf. chapitre 8).

L'utilisation de ce test présente une différence majeure par rapport aux études portant sur l'identification du mot : alors que les recherches menées sur l'apprentissage de la lecture se font la plupart du temps sur des mots isolés, le test de Lefavrais présente les mots en contexte. Celui-ci rendrait disponibles d'autres types d'indices pour identifier les mots susceptibles de discuter les modèles de référence.

C'est pourquoi, dans un premier temps, nous nous proposons de voir si les effets des variables linguistiques « classiques » observés dans les recherches portant sur l'identification du mot isolé apparaissent en situation de lecture de texte (*cf.* chapitres 9 et 10).

Dans un second temps, nous nous proposons de profiter du fait que le texte de Lefavrais ne soit pas constitué de façon anodine. En effet, l'introduction de faux indices parmi les indices pertinents amènerait les enfants à tomber dans ce que Lefavrais appelle des pièges et qui traduisent alors une défaillance dans le décodage, selon sa définition de l'acte lexique et de son apprentissage. Nous nous demandons donc alors de quelle façon les erreurs liées aux variables du test de *L'alouette* nous informent sur l'apprentissage de la lecture, et plus précisément sur le caractère discontinu ou continu de la lecture (*cf.* chapitres 11, 12, 13, 14).

Enfin, dans un dernier chapitre, pour avoir une vue d'ensemble, nous essaierons de voir quel est le processus de lecture dominant à chaque niveau scolaire et de quelle façon il évolue (*cf.* chapitre 15).

Nous terminerons par une conclusion sur les différents résultats obtenus dans cette recherche que nous essaierons d'expliquer, puis nous énoncerons les perspectives de recherches que permet ce travail.

ETAT DE LA QUESTION

Chapitre 1



LE NIVEAU EN LECTURE DES ELEVES

Plan du chapitre

I. Le niveau en lecture des élèves : évolution des dispositifs d'évaluation.....	25
1. Les évaluations françaises.....	26
a) Les évaluations nationales dans le système éducatif : entre évaluations-bilan et évaluations-diagnostiques.....	26
b) L'évaluation nationale hors du système éducatif : la Journée Défense et Citoyenneté.....	29
2. Les évaluations internationales	30
II. Les résultats en lecture aux dernières évaluations.....	33
1. La maîtrise de la lecture en référence aux programmes nationaux en fin d'école primaire.....	33
2. Les compétences en lecture en fin de scolarité obligatoire	34
3. Les capacités à utiliser la langue écrite chez les jeunes de 15 ans	35
III. Une baisse du niveau en lecture ?	37
1. Evolution des compétences en lecture à court terme	37
a) Evolution des compétences de base en lecture en fin d'école primaire	37
b) Evolution des compétences en lecture en fin de scolarité obligatoire.....	38
c) Evolution des capacités à utiliser la langue écrite chez les jeunes de 15-17 ans	39
2. Evolution des compétences en lecture à long terme : les études comparatives françaises	40
IV. Des inégalités dans les performances en lecture	42
1. Une inégalité commune à tous les pays : le genre.....	43
2. Des inégalités caractéristiques de notre pays.....	44
a) Des inégalités selon le milieu social.....	44
b) Des inégalités selon le type d'établissement fréquenté	44
c) Des inégalités entre les plus faibles et les meilleurs.....	45
CONCLUSION.....	47

I. Le niveau en lecture des élèves : évolution des dispositifs d'évaluation

La « massification » de l'école dans les années 1960, faisant émerger les notions d'échec scolaire et d'inégalité des chances, amène l'école à redéfinir ses objectifs et, avec l'apparition de la notion d'illettrisme dans les années 1980, la lecture passe au centre des préoccupations. Dans ce contexte, la nécessité de pouvoir faire état des acquis des élèves à différentes étapes de leur parcours scolaire s'impose peu à peu et les dispositifs d'évaluation se multiplient. Ainsi, dans le cadre de l'évaluation permanente du système éducatif, le ministère de l'Education nationale développe une politique d'évaluation afin d'« insuffler au sein du système éducatif tout entier une nouvelle culture de l'évaluation fondée sur la recherche de la performance »¹. Ainsi, entre les évaluations françaises d'un côté et les évaluations internationales de l'autre, des évaluations portant sur la maîtrise de la langue, notamment en lecture, existent depuis maintenant plusieurs années.

Depuis la fin des années 70, à la demande du ministère de l'Education nationale, la DEPP² conçoit et met en œuvre des évaluations, des enquêtes et des études permettant d'apprécier les performances et les compétences en lecture des élèves du primaire (et du début du collège)³.

Plus largement, au-delà de ces publics scolaires, qui constituent la population spécifique de notre recherche, la DEPP contribue à l'évaluation du niveau en lecture des jeunes à l'issue de la scolarité obligatoire⁴.

Parallèlement à ces enquêtes nationales, elle participe aussi à des études européennes destinées à comparer les performances et les modes de fonctionnement des différents systèmes éducatifs, aussi bien pour les enfants du primaire que pour les élèves plus âgés⁵.

¹ Ministère de l'Education nationale (2012)

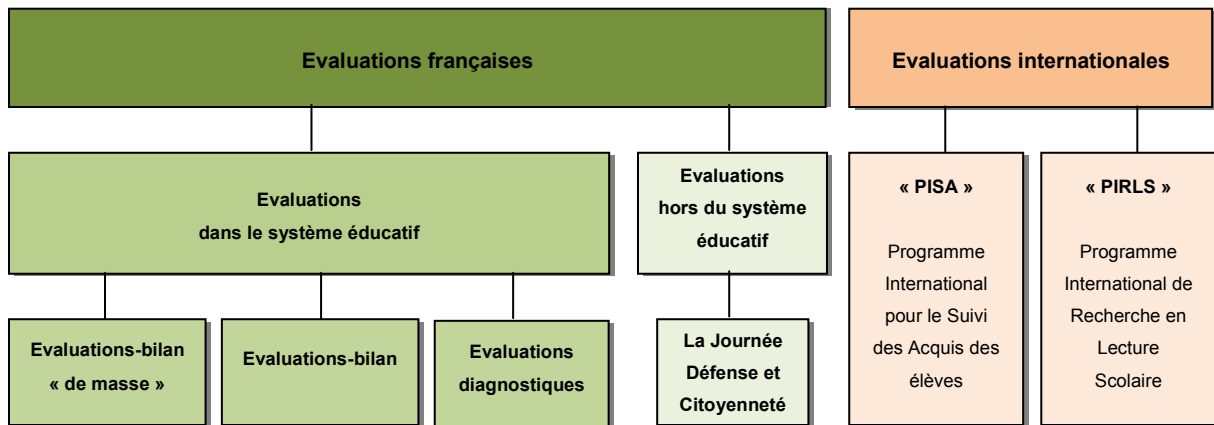
² Anciennement le service de la prévision, des statistiques et de l'évaluation du MEN qui devient, le 10 février 1987, la Direction de l'Evaluation et de la Prospective (DEP). En 2006, elle devient la Direction de l'Evaluation, de la Prospective et de la Performance (DEPP).

³ La DEPP publie aussi, de façon périodique, des ouvrages regroupant un certain nombre d'indicateurs statistiques (tels que les coûts, les activités et les résultats) permettant de faire état du système éducatif en France (« Repères et références statistiques », « L'état de l'école » ou encore « La géographie de l'école »). Par ailleurs, en collaboration avec des inspecteurs généraux, des équipes de recherche ou des organismes extérieurs (nationaux ou internationaux) elle établit aussi des rapports concernant le fonctionnement du système éducatif et les acquis des élèves.

⁴ Bien que sans rapport direct avec notre objet d'étude, nous évoquerons ces évaluations et leurs résultats dans la mesure où ils participent très largement au débat sur la baisse du niveau en lecture.

⁵ Nous présenterons aussi ces études pour les mêmes raisons que précédemment.

Figure 1 : les différents types d'évaluations en France en 2010



1. Les évaluations françaises

a) Les évaluations nationales dans le système éducatif : entre évaluations-bilan et évaluations-diagnostiques

En 1979, suite à une demande de la Direction des écoles, des collèges et des lycées, les premières évaluations-bilan pédagogiques nationales sont mises en place dans les écoles par la DEP. L'objectif de ce dispositif est de " *fournir à la Direction des Ecoles des moyens d'apprécier aussi objectivement que possible en vue de l'améliorer, l'efficacité de l'enseignement et de l'action éducative de l'école; c'est en identifiant avec précision les insuffisances qu'on peut déterminer les opérations à conduire pour y remédier, émission d'instructions, actions de formation continue, aménagement de formation initiale* " (Seibel et Couanau, 1980). Ainsi, en mesurant les performances des élèves et en identifiant leurs compétences et leurs connaissances à des moments clés de la scolarité, elles doivent permettre de « *connaître les résultats du fonctionnement pédagogique du système éducatif au regard des objectifs qui lui sont assignés* »⁶. Ces évaluations sont donc essentiellement destinées aux décideurs pour leur permettre d'améliorer le pilotage du système éducatif. Elles sont réalisées ponctuellement, sur des échantillons représentatifs d'élèves scolarisés dans plusieurs établissements appartenant à des académies différentes. Ce dispositif permet d'évaluer les apprentissages instrumentaux : langue française et mathématiques. Les premières évaluations sont effectuées en fin de CP en 1979, puis en fin de CE2 en 1981 et en fin de CM2 en 1983.

⁶ « Evaluation pédagogique dans les collèges », juin 1982, page de présentation cité dans « L'appréciation des compétences des élèves et des jeunes en lecture et en écriture et l'évolution de ces compétences dans le temps » de CÉARD M-T, RÉMOND M. et VARIER M., 2003.

Sur cette base, à partir de la fin des années 80, une culture de l'évaluation va se développer en France, entraînant une multiplication des dispositifs d'évaluation. En effet, face à une hétérogénéité croissante des élèves, les évaluations menées jusque-là ne permettaient pas aux enseignants de repérer les acquis et les lacunes de manière individuelle. De la sorte, plusieurs outils d'évaluation aux fonctions différentes (de bilan ou diagnostique) vont être créés.

Ainsi, suite à la loi d'orientation du 10 juillet 1989, réorganisant la scolarité en cycles⁷, le ministre de l'Education nationale, Lionel Jospin, met en place un nouveau dispositif d'évaluation (des « *évaluations de masse* ») davantage axé sur les élèves. Dorénavant, l'objectif est d'établir un diagnostic des acquis et des difficultés des élèves afin d'aider les enseignants, d'une part, et, d'autre part, de faire un bilan afin d'identifier les compétences à acquérir au cours de l'année. Contrairement aux précédentes, ces évaluations sont annuelles et passées par tous les élèves à l'entrée du CE2 et de la 6^{ème}. Les sessions se déroulent en début d'année, ce qui laisse une marge aux enseignants pour mieux organiser leur enseignement et éventuellement, mettre en place des moyens pour remédier aux difficultés que pourraient rencontrer certains élèves.

A ces évaluations de « masse », le dispositif d'évaluation-bilan mis en place en 1979 s'ajoutent des évaluations-bilan plus ponctuelles mesurant les compétences dans un domaine précis, à un moment clef de la scolarité, afin de confronter les performances des élèves face aux exigences du système éducatif.

Ainsi, en 1987, une enquête est menée en lecture, écriture et calcul (« *Lire, écrire, compter : les performances des élèves de CM2 à vingt ans d'intervalle 1987-2007* ») sur un échantillon d'élèves en fin de CM2 afin de vérifier dans quelle mesure les élèves ont acquis les savoirs et savoir-faire correspondant aux exigences du programme de sixième.

En 1997, une évaluation destinée à cerner les difficultés en lecture à l'entrée en sixième est proposée aux élèves (« *Étude spécifique sur les élèves en difficulté en lecture à l'entrée en sixième* »).

Répliquées dans le temps, ces évaluations permettent d'établir une évolution des acquis des élèves en lecture.

⁷ En France, de l'école maternelle à l'école élémentaire, la scolarité s'organise en trois cycles « pour lesquels sont définis des objectifs et des programmes nationaux de formation comportant une progression annuelle ainsi que des critères d'évaluation » (cf. L. JOSPIN, loi d'orientation du 10 juillet 1989, art.4): le cycle des apprentissages premiers (très petite section, petite section et moyenne section), le cycle des apprentissages fondamentaux (grande section, cours préparatoire et cours élémentaire première année) et le cycle des approfondissements (cours élémentaire deuxième année, cours moyen première année et cours moyen deuxième année).

En 2000, un cycle d'évaluations-bilan est mis en place en fin d'école primaire et de collège afin de mieux cerner les niveaux de connaissances et de compétences atteints par rapport aux programmes. Depuis, chaque année, des échantillons d'élèves passent des épreuves dans une discipline clef du cursus scolaire. Une évaluation portant sur la maîtrise du langage et de la langue française a été réalisée en 2003 en CM2⁸.

En 2007, suite à l'entrée en vigueur de la Loi Organique relative aux Lois de Finances (LOLF), le ministère de l'Education nationale établit un certain nombre d'indicateurs (à partir du socle commun) destinés à mesurer la proportion d'élèves maîtrisant les compétences de base en français et mathématiques afin de « *conduire tous les élèves à la maîtrise des compétences de base exigibles au terme de la scolarité primaire* ». Ainsi, sont menées chaque année, en plus des précédentes, des évaluations-bilan sur des échantillons représentatifs d'élèves de CM2.

Cette même année 2007, faisant suite à l'instauration d'un « socle commun de connaissances et de compétences⁹ », de nouveaux protocoles nationaux d'évaluation de « masse » sont mis en place à l'école en CE1 et CM2, les deux premiers paliers du socle (fin du cycle des apprentissages fondamentaux et fin d'école élémentaire). Dès lors, ces évaluations remplacent celles qui avaient lieu chaque année à la rentrée du CE2 et de la 6^{ème}.

En 2009, la fonction de ces évaluations change. L'idée n'est plus d'établir un diagnostic (repérer les élèves en difficulté à l'échelle de la classe) mais de dresser un bilan afin de « *voir comment chaque élève se situe par rapport aux objectifs définis dans les nouveaux programmes et de remédier, le cas échéant, aux difficultés constatées* »¹⁰.

Ainsi, en 2010, faisant suite au souci de mesurer l'efficacité du système éducatif, les dispositifs français d'évaluation des compétences des élèves sont principalement des outils permettant de dresser un bilan des acquis des élèves. Ayant lieu à des étapes clef de leur parcours scolaire et concernant les domaines fondamentaux, ils sont conçus en référence aux exigences des programmes scolaires. Néanmoins, depuis les années 2000, une banque d'outils

⁸ « *Compréhension à l'écrit et à l'oral des élèves en fin d'école primaire* », DEPP, 2003.

⁹ Article 9 de la loi du 23 avril 2005 : « la scolarité obligatoire doit au moins garantir à chaque élève les moyens nécessaires à l'acquisition d'un socle commun constitué d'un ensemble de connaissances et de compétences qu'il est indispensable de maîtriser pour accomplir avec succès sa scolarité, poursuivre sa formation, construire son avenir personnel et professionnel et réussir sa vie en société ».

¹⁰ Discours de X. Darcos du 21.07.2008.

diagnostiques regroupant un millier d'évaluations est mise à la disposition des enseignants sur internet afin qu'ils puissent ajuster leurs pratiques pédagogiques aux besoins des élèves.

En outre, il y a une dizaine d'années, dans le cadre de la réforme du service national, le gouvernement a mis en place pour tous les jeunes Français un « *parcours* citoyenneté ». Il se compose de trois étapes dont, entre autres, la participation obligatoire à une Journée Défense et Citoyenneté¹¹ au cours de laquelle des tests en lecture sont passés.

b) L'évaluation nationale hors du système éducatif : la Journée Défense et Citoyenneté

Depuis 1998, la participation à la Journée Défense et Citoyenneté (JDC) est obligatoire pour tous les jeunes citoyens français âgés de 16 à 18 ans¹². Les tests en lecture, conçus par la DEPP, ont un double objectif. D'abord, sur le plan national, ils permettent d'avoir un indicateur statistique quant au niveau de maîtrise de la lecture à l'issue de la scolarité obligatoire. Puis, sur le plan individuel, ils permettent de repérer les jeunes les plus en difficulté afin de les informer et de leur proposer de l'aide. Les tests évaluent deux dimensions fondamentales de la lecture : les mécanismes de base et la compréhension. En 2009, afin d'améliorer les procédures de standardisation et de réduire les contraintes logistiques, les conditions de passation de ces épreuves ont été modifiées. Désormais, les jeunes répondent aux questions, projetées sur un grand écran, à l'aide d'un boîtier électronique. Les réponses et les temps de réaction peuvent ainsi être enregistrés. Les scores sont calculés automatiquement et des profils de lecteurs sont générés.

En complément à ces évaluations françaises menées par l'Education nationale, d'autres évaluations permettent d'apprécier les compétences en lecture des élèves. En effet, depuis quelques années la nécessité de mesurer les performances du système éducatif donne lieu à des évaluations internationales qui tendent à se multiplier. Prises de plus en plus en considération, elles ont pour objectif de « *repérer les bonnes pratiques et les dispositifs innovants afin de faire converger les systèmes éducatifs européens, pour donner naissance à l'Europe de la connaissance* »¹³.

¹¹ Ancienne Journée d'Appel de Préparation à la Défense (JAPD) devenue la Journée Défense et Citoyenneté depuis 2010.

¹² Jusqu'en 2010, cette session s'appelait la « Journée d'Appel et de Préparation à la Défense (JAPD) » et tous les jeunes citoyens âgés entre 16 et 25 ans devaient y participer. En 2010, l'âge limite est passé à 18 ans. La participation à cette journée est considérée comme la 3ème étape du « parcours de citoyenneté ».

¹³ Cf. Discours du 14.11.08 lors de la conférence de presse de Xavier Darcos, Ministre de l'Education nationale en 2008.

2. Les évaluations internationales

Destinées à comparer les performances et les modes de fonctionnement des différents systèmes éducatifs, aussi bien pour les enfants du primaire que pour les élèves plus âgés, les évaluations internationales ne cessent de susciter de l'intérêt.¹⁴

En 1991, l'Association Internationale pour l'Evaluation du rendement scolaire (IEA)¹⁵ décide de mener une première enquête sur des élèves âgés de 9 et 14 ans en « *Reading Literacy* ». Ce concept est défini à l'époque comme « *la capacité à comprendre et à utiliser les formes du langage écrit requises par la société et/ou valorisées par l'individu* » (Elley W.B., 1992). Cette enquête a deux objectifs majeurs. D'une part, décrire les niveaux moyens de performances en « *reading literacy* » des élèves des 32 systèmes éducatifs participants¹⁶. D'autre part, créer des tests et des questionnaires internationaux valides pouvant être utilisés dans tous les pays pour mesurer la « *reading literacy* » permettant ainsi des comparaisons internationales.

En 2000, un Programme International pour le Suivi des Acquis des élèves (PISA)¹⁷ est mis en place par l'OCDE (Organisation de Coopération et de Développement Economique). C'est une évaluation internationale qui a lieu tous les trois ans dans les pays membres¹⁸ et quelques pays partenaires. Les premières collectes de données ont été effectuées en 2000, puis 2003, 2006, 2009 et 2015. La prochaine est prévue en 2012¹⁹. L'objectif est de mesurer « *l'acquisition des savoirs et savoir-faire essentiels à la vie quotidienne au terme de la scolarité obligatoire* ». Autrement dit, contrairement aux évaluations nationales qui mesurent les compétences attendues par les programmes scolaires, PISA a pour but de tester « *les aptitudes des élèves à appliquer les connaissances acquises à l'école aux situations de la vie réelle* ». Cette évaluation se fait sur un échantillon aléatoire d'élèves scolarisés dans des

¹⁴ Ces recherches qui conduisent à un classement des pays selon le degré de performance de leurs élèves sont particulièrement médiatisées, ce qui participe à la dramatisation du débat sur le niveau scolaire en général et plus particulièrement en lecture. Même si actuellement ce qui semble poser problème c'est le niveau en mathématiques,

¹⁵ « *International association for the Evaluation of educational Achievement* » (IEA)

¹⁶ Belgique, Botswana, Canada, Chypre, Danemark, Finlande, France, Allemagne (RFA et RDA), Grèce, Hong-Kong, Hongrie, Islande, Indonésie, Irlande, Italie, Pays-Bas, Nouvelle-Zélande, Nigéria, Norvège, Philippines, Portugal, Singapour, Slovénie, Espagne, Suède, Suisse, Thaïlande, Trinité et Tobago, Etats-Unis, Venezuela, Zimbabwe.

¹⁷ « *Programme for International Student Assessment* » (PISA)

¹⁸ 34 pays sont membres de l'OCDE: Allemagne, Australie, Autriche, Belgique, Canada, Chili, Corée, Danemark, Espagne, Estonie, Etats-Unis, Finlande, France, Grèce, Hongrie, Irlande, Islande, Israël, Italie, Japon, Luxembourg, Mexique, Norvège, Nouvelle-Zélande, Pays-Bas, Pologne, Portugal, République slovaque, République tchèque, Royaume-Uni, Slovénie, Suède, Suisse et Turquie.

¹⁹ La publication des résultats est prévue pour le 3.12.13.

établissements publics ou privés et dont l'âge varie entre 15 ans et 3 mois et 16 ans et 2 mois au début de l'évaluation. Les tests concernent trois domaines : la compréhension de l'écrit (« *reading literacy* »), la culture mathématique (« *mathematical literacy* ») et la culture scientifique (« *scientific literacy* »). Lors de chaque session, un domaine est privilégié par rapport aux autres et un temps plus long lui est accordé. Concernant la compréhension de l'écrit, celle-ci a été évaluée en 2000 et 2009. Dans le cadre de cette enquête, la « *reading literacy* » est définie comme étant « *la capacité de comprendre, d'utiliser et d'analyser des textes écrits afin de pouvoir réaliser ses objectifs, développer ses connaissances et son potentiel et jouer un rôle actif dans la société* ».

Le programme PIRLS (« *Progress in International Reading Literacy Study* »), élaboré par l'IEA en 2001, se rapproche de celui du PISA car il repose lui aussi sur un cadre conceptuel de « *reading literacy* ». Ce dispositif définit la lecture comme étant « *l'aptitude à comprendre et à utiliser les formes du langage écrit que requiert la société ou qui sont importants pour l'individu. Les jeunes lecteurs peuvent construire du sens à partir de textes très variés. Ils lisent pour apprendre, pour s'intégrer dans une société où la lecture joue un rôle essentiel et pour leur plaisir* ». Il émet l'hypothèse que « *la compétence en lecture est directement liée aux raisons qui incitent le lecteur à lire* »²⁰. Dirigée par le Centre international d'études de l'université de Boston (« *International Study Center at Boston College* ») cette étude a lieu tous les cinq ans. Les épreuves, réalisées sur des enfants de 9-10 ans (4^{ème} année de scolarité obligatoire), mesurent deux composantes : les compétences en lecture et les objectifs donnés à la lecture. Les compétences sont évaluées par des questions de compréhension de texte permettant de voir si les élèves sont capables de prélever des informations explicites, de faire des inférences directes, d'interpréter et d'assimiler les idées et informations, d'évaluer et examiner le contenu, la langue et les éléments textuels. Les objectifs de lecture sont, quant à eux, mesurés par des items qui évaluent, pour chaque élève, le degré de nécessité de lire pour accéder aux textes littéraires et pour acquérir et utiliser des informations de la vie courante.

²⁰ « *Les élèves de CM1, premiers résultats d'une évaluation internationale en lecture (PIRLS)* », Note d'Information 03-22, DEPP.

Tableau 1 : les différents types d'évaluation en lecture en France en 2010

	Evaluations de « masse »	Evaluations bilan	Evaluations diagnostiques	Journée Défense Citoyenneté	PISA	PIRLS
Fonction	Dresser un bilan	Dresser un bilan	Etablir un diagnostic	Dresser un bilan	Dresser un bilan	Dresser un bilan
Finalités	Finalité sociale	Finalité sociale	Finalité pédagogique et didactique	Finalité sociale	Finalité sociale	Finalité sociale
Population	Tous les élèves de 7 et 10 ans environ	Echantillons d'élèves	Individuel	Tous les jeunes citoyens français entre 16 et 18 ans	Echantillons d'élèves de 15 ans	Echantillons d'élèves de 9 ans environ
Moment de passation	En CE1 et CM2	En fin d'école primaire	En début d'année	Entre le recensement et l'âge de 18 ans	Au cours de la 15 ^{ème} année	En 4 ^{ème} année de scolarité obligatoire
Dimensions mesurées	Les acquis en français	Les compétences de base en lecture	Les compétences en lecture	Processus de lecture et la compréhension	La compréhension de l'écrit	La compréhension de l'écrit
En référence...	au programme	au programme	au programme	A la conception classique de la lecture	à la vie sociale	à la vie sociale
Possibilité de comparaison	Inter-élèves	Inter-élèves	Inter-élèves	Inter-élèves	Entre systèmes éducatifs	Entre systèmes éducatifs

Ainsi, depuis une trentaine d'années, la culture de l'évaluation se développe fortement et le nombre de dispositifs, mesurant entre autre les compétences en lecture, s'est multiplié (cf. tableau 1). Avant la loi d'orientation de 1989, des évaluations bilan étaient effectuées de façon ponctuelle sur des échantillons d'élèves afin de mesurer l'efficacité du système éducatif. Suite à cette loi, l'école a vu se développer des « évaluations de masse » annuelles et passées obligatoirement par tous les élèves à des niveaux clef du parcours scolaire. Ainsi, jusqu'en 2008, ces évaluations devaient permettre d'établir un diagnostic des acquis des élèves dans les apprentissages fondamentaux à des moments clef de leur parcours scolaire. À partir de 2009, leur objectif change : l'intérêt est de situer les acquis des élèves par rapport aux exigences des programmes nationaux. Parallèlement à ces évaluations nationales, des évaluations-bilan sont régulièrement menées sur des échantillons représentatifs afin de mesurer les compétences de base des élèves en fin d'école primaire. S'y ajoutent également les tests en compréhension de l'écrit réalisés lors de la JDC qui permettent d'avoir des indications quant au niveau de lecture des jeunes à l'issue de la scolarité obligatoire. Enfin, les années 2000 voient émerger des évaluations internationales destinées à comparer les performances des systèmes éducatifs de plusieurs pays. Contrairement aux évaluations de masse, celles-ci ne mesurent pas directement les compétences en lien avec les programmes. Elles tendent davantage à évaluer les capacités qu'ont les jeunes à exploiter leurs connaissances et leurs compétences acquises à l'école dans des situations de la vie réelle.

Complémentaires des évaluations-bilan, elles permettent alors, en comparaison avec d'autres systèmes éducatifs, de révéler les points forts et les points faibles des élèves.

II. Les résultats en lecture aux dernières évaluations

Les acquis et les difficultés des jeunes en lecture, le nombre de personnes jugées en situation d'illettrisme ou encore la mauvaise position qu'occupent les élèves français en compréhension de l'écrit au niveau international sont autant de sujets qui préoccupent et qui font débat dans notre société.

Tantôt mesurée en référence aux programmes (compétences minimales requises), tantôt par rapport à son exploitation (capacité d'utilisation), la lecture est évaluée par un grand nombre de dispositifs. Qu'ils soient nationaux ou internationaux, ils permettent tous d'avoir des indicateurs quant au niveau de maîtrise de la lecture à différents moments de la scolarité.
Où en est-on en ce qui concerne la lecture ces trois dernières années ?

1. La maîtrise de la lecture en référence aux programmes nationaux en fin d'école primaire.

Afin de dresser un bilan des acquis des élèves en français, une évaluation nationale a lieu actuellement aux deux moments clef de la scolarité que sont le CE1 (en mai) et le CM2 (en janvier). Les épreuves sont conçues en référence aux programmes. Ainsi, en français, les domaines évalués sont la lecture, l'écriture, le vocabulaire, la grammaire et l'orthographe. Pour analyser les performances des élèves, quatre groupes sont constitués en fonction du nombre de bonnes réponses données sur les 60 items proposés. Plus un élève comptabilise de bonnes réponses et plus il a d'acquis. Ainsi, en 2010, 44% des élèves ont des acquis solides en CE1 et 43% en CM2. Du côté des élèves les plus en difficulté, on considère qu'en fin de cycle 2, 17% des élèves de CE1 ont des acquis fragiles et 8% ont des acquis insuffisants. En dernière année d'école primaire, les taux sont quasiment les mêmes : 20% des élèves de CM2 ont des acquis fragiles et 7% ont des acquis insuffisants. Autrement dit, que ce soit à un niveau intermédiaire (CE1) ou en fin d'école primaire (CM2), environ un quart des élèves ont des acquis insuffisants en français, y compris en lecture, qui seront à « *consolider dans les mois à venir* ». Ceux qui sont le plus en difficulté (environ 7%) vont alors « *bénéficier d'une aide spécifique* ».

Cette même année, en 2010, pour la quatrième année consécutive, un échantillon d'élèves de CM2 a passé une évaluation de la maîtrise des compétences de base en fin d'école afin de voir dans quelle proportion ils maîtrisent le socle commun. À ce niveau, les compétences en français regroupent celles relevant de la lecture et de la maîtrise des outils de la langue. Les compétences liées à l'expression orale et écrite n'ont pas été retenues. Les compétences en français ont été évaluées à travers un QCM et, suite à l'analyse des résultats, un seuil d'exigence a été fixé en dessous duquel on considère que les élèves ne maîtrisent pas les compétences définies par le socle commun. Les résultats indiquent qu'en 2010, en fin de CM2, 12,4% des élèves ne maîtrisent pas les compétences de base en français. Plus précisément, ces élèves éprouvent des difficultés pour chercher et prélever des informations, mais aussi pour comprendre un texte court de genre littéraire ou documentaire. Ils ne maîtrisent pas totalement les correspondances grapho-phonologiques, ont des difficultés en conjugaison lorsqu'il s'agit d'identifier les principaux temps de l'indicatif et enfin, ont du mal à reconnaître les règles simples d'orthographe grammaticale et lexicale.

Il semble donc, qu'au niveau national, la proportion d'élèves ne maîtrisant pas les compétences de base, y inclus la lecture, varie d'un dispositif à l'autre. Alors qu'en 2010, aux évaluations nationales en français, environ un quart des élèves de CM2 n'ont pas les compétences requises par les programmes, ils sont 12,4% lors de l'évaluation-bilan. Ceux qui sont le plus en difficulté représentent 7% des élèves en fin d'école primaire. Les résultats détaillés par domaine de ces évaluations étant réservés aux établissements, il ne nous est alors pas possible de voir ce qu'il en est, plus précisément, dans le cas de la lecture. Les tests de la JDC vont nous permettre de voir plus exactement quelle est la proportion de jeunes en difficulté de lecture en fin de scolarité obligatoire.

2. Les compétences en lecture en fin de scolarité obligatoire

L'objectif des tests de lecture réalisés dans le cadre de la JDC est de repérer chez les faibles lecteurs trois grandes difficultés estimées majeures dans l'acte de lire : une mauvaise automatisation des mécanismes dans l'identification des mots, une compétence langagière insuffisante (c'est-à-dire peu de connaissances lexicales) et une pratique défailante de la compréhension de textes. En 2009, les résultats ont permis de dégager huit profils de lecteurs qui peuvent être regroupés deux à deux: ceux qui ont des difficultés sévères (profils 1 et 2),

ceux qui ont de très faibles capacités de lecture (profils 3 et 4), les lecteurs médiocres (profils 5a et 5b) et les lecteurs efficaces (profils 5c et 5d). Il apparaît qu'environ 80 % des jeunes âgés de 17 ans environ sont des bons lecteurs (profils 5c et 5d) tandis que 20% ne sont pas des lecteurs efficaces (profils 1 à 4 et profils 5a et 5b). Parmi ces derniers, presque 10% sont considérés comme des lecteurs médiocres. Ce sont des jeunes qui, malgré leurs difficultés lors de l'identification de certains mots et/ou leur pauvreté en connaissances lexicales, parviennent à une compréhension minimale du texte. Par ailleurs, 5% ont de faibles capacités en lecture. Ces jeunes ont un niveau lexical oral plutôt correct mais ne réussissent pas à comprendre les textes écrits. Enfin, ceux qui présentent de graves difficultés sont de l'ordre de 5%. Ces derniers présentent un important déficit en vocabulaire, ce qui ne leur permet pas d'arriver à la compréhension du texte. Parmi eux, une majorité (3%) n'a pas encore installé les mécanismes de base de traitement du langage écrit, ce qui les place en situation de « quasi-illettrisme ».

Alors que les résultats aux évaluations nationales indiquent qu'entre 12% et 20% des élèves ne maîtrisent pas les compétences de base en fin de primaire, les tests de la JDC en décomptent 15%. En ce qui concerne ceux qui sont le plus en difficulté, les proportions ne semblent pas beaucoup varier ce, quels que soient les acquis mesurés et le niveau de scolarité (en fin d'école ou en fin de scolarité) (entre 5% et 7%). Les compétences en lecture évaluées dans le cadre du PISA étant différentes des précédentes, il nous a semblé intéressant de voir ce qu'il en est pour les jeunes de 15 ans lorsqu'il s'agit d'exploiter les savoirs et savoir-faire en lecture acquis à l'école.

3. Les capacités à utiliser la langue écrite chez les jeunes de 15 ans

De manière générale, le score de la France lors de la dernière enquête du PISA en compréhension de l'écrit (2009), se situe légèrement au-dessus de la moyenne des pays de l'OCDE (496 points pour la France et 493 points pour les pays de l'OCDE). Pour une analyse plus fine des résultats, les élèves ont été répartis selon sept niveaux de compétences en fonction des scores obtenus. Rejoignant les constats précédents, il apparaît que 19,7% des jeunes Français âgés de 15 ans ont un faible niveau en compréhension de l'écrit dont 5,6% ont un niveau très faible voire médiocre pour 2,3% d'entre eux. Autrement dit, ils ne sont pas -ou très peu- capables de repérer une information explicite dans un texte court sur un sujet qui leur est familier. Pour cette catégorie, ces résultats placent donc la France à un niveau

inférieur à la moyenne des autres pays de l'OCDE (18,8% d'élèves avec un faible niveau). A l'autre bout, environ 10% des jeunes Français ont un bon niveau en compréhension de l'écrit, c'est-à-dire qu'ils arrivent à une compréhension fine du texte et à porter un jugement sur des documents dont le thème leur est peu familier. Par ailleurs, il s'avère que le taux de non-réponse est plus élevé en France lorsqu'il s'agit de répondre à des questions ouvertes (16% en France contre 11% pour les pays de l'OCDE) plutôt que fermées (7% en France contre 4% pour l'OCDE).

En **résumé**, les différents chiffres semblent donc confirmer le récent constat de la Cour des Comptes (2010) qui, au terme d'une enquête de trois ans, publie un rapport²¹ dans lequel elle dénonce l'échec de l'école dans sa mission pour « *donner à tous les élèves un socle commun de compétences et de connaissances* », notamment en ce qui concerne la maîtrise de la langue.

Toutefois, cet état des lieux nous permet de constater que la proportion d'élèves ou de jeunes en difficulté de lecture plus précisément, varie selon la dimension qui est évaluée mais aussi selon le degré de difficulté considéré.

En effet, si on considère la proportion d'élèves n'ayant pas les acquis exigés par les programmes en français en fin d'école primaire, ce taux varie entre 12% et 27%. En fin de scolarité obligatoire, ils sont 20% à ne pas maîtriser les mécanismes de base en jeu dans la lecture. Ce constat semble se confirmer avec l'enquête internationale PISA qui indique qu'en France environ 20% des jeunes âgés de 15 ans ne sont pas capables de comprendre, d'utiliser et d'analyser des textes écrits nécessaires à la vie quotidienne. Par ailleurs, la proportion d'élèves en grave difficulté semble être quasiment la même quelle que soit la dimension de la lecture évaluée et le niveau de scolarité (entre 5% et 7%).

Il nous est donc impossible de conclure simplement sur l'état actuel des compétences en lecture de nos élèves. Les résultats obtenus à ces différentes évaluations ne peuvent ou plutôt ne doivent pas être analysés d'une façon générale. En effet, comme nous l'avons vu, malgré la diversité des dispositifs d'évaluation, les objectifs mesurés ne sont pas toujours les mêmes, tout comme la population de référence. De la sorte, les résultats doivent être interprétés avec prudence et ne sauraient être généralisés.

²¹ « L'Education nationale face à l'objectif de la réussite de tous les élèves », mai 2010

Néanmoins, en dépit de ces différences inhérentes aux dispositifs d'évaluation, ces chiffres incitent à penser que le niveau en lecture en France fait problème et relancent la polémique sur la baisse du niveau de maîtrise de la langue écrite.

III. Une baisse du niveau en lecture ?

« *Le niveau baisse* », « *il y a de plus en plus d'illettrés* »²² sont le genre d'idées reçues largement énoncées dans la presse et les médias. En effet, il existe en France une longue tradition de déploration de la baisse du niveau et il est couramment admis que l'échec est massif en ce qui concerne la maîtrise de la langue et plus particulièrement en lecture²³. Précédemment, nous avons pu constater que le niveau en lecture en France fait problème. Pour autant, est-ce que ce constat nous permet de dire que le niveau de nos élèves est en baisse ? Seule une analyse de l'évolution des performances en lecture dans le temps nous permettra de répondre à cette question.

Malgré le nombre important de dispositifs permettant d'évaluer les performances en lecture, il reste difficile d'établir une évolution des performances dans le temps. En effet, les protocoles d'évaluation étant la plupart du temps différents d'une année à l'autre, les comparaisons avec les années antérieures sont difficilement réalisables. Toutefois, la variété des dispositifs nous permet de regarder ce qu'il en est des compétences en lecture à différents moments de la scolarité et sur des périodes plus ou moins longues.

1. Evolution des compétences en lecture à court terme

a) Evolution des compétences de base en lecture en fin d'école primaire

Rappelons qu'en 2009 le dispositif d'évaluation nationale a changé. De ce fait, toute comparaison rigoureuse avec les années antérieures est irréalisable. Il est toutefois possible de regarder les résultats obtenus pour 2009. Ainsi, alors que les résultats aux évaluations nationales de 2010 indiquent que 25% des élèves de CE1 ont des acquis fragiles voire

²² « *La lecture : idées reçues* », J. Fijalkow, 2003.

²³ Selon un sondage mené par la TNS-SOFRES en 2008, 60% des Français estiment que le niveau scolaire des élèves est moins bon qu'il y a 20 ans. Plus particulièrement, 78% des Français jugent que la maîtrise du français est insatisfaisante.

insuffisants, ils étaient 27% en 2009. Concernant les élèves de CM2, ils sont 27% en 2010 et 25% en 2009. Ces écarts minimes et sans doute non significatifs semblent indiquer une stabilité des résultats depuis 2009 : la proportion d'élèves ayant des acquis fragiles ou insuffisants en français, et par conséquent en lecture, ne semble pas avoir changé d'une année sur l'autre.

Il en est de même en ce qui concerne la maîtrise des compétences de base en français en CM2. La proportion d'élèves n'ayant pas les bases suffisantes en français est de 14% en 2007 et de 12% en 2010. Malgré une baisse de deux points, la différence paraît peu significative. Ainsi, en quatre ans la proportion d'élèves qui ne maîtrisent pas les compétences de base en français à l'issue de l'école primaire est restée la même.

Ce dernier chiffre rejoint le constat effectué lors de l'évaluation-bilan menée par la DEPP en 2003 sur la maîtrise du langage et de la langue française en fin d'école primaire. Les résultats de cette étude ont été analysés selon une échelle de performances en six niveaux (du niveau 0 au niveau 5), décrivant pour chacun d'eux les types de compétences qui sont maîtrisés. Ainsi en 2003, 15% des élèves de CM2 présentent des difficultés, voire de grandes difficultés pour 3% d'entre eux. Ils ne maîtrisent pas, ou très mal, les compétences en langue nécessaires, selon les programmes, pour le passage en 6e.

b) Evolution des compétences en lecture en fin de scolarité obligatoire

Les protocoles de l'évaluation en lecture passée lors de la JDC ayant été modifiés en 2009, nous ne pouvons pas davantage établir une comparaison avec les années précédentes. Néanmoins, si on regarde l'évolution des scores obtenus avant la mise en place de ces nouveaux tests, on constate qu'à l'issue de la scolarité obligatoire, la proportion des jeunes en difficulté de lecture qui était de 11% en 2004 est de 11,8% en 2008. Parmi eux, ceux qui sont en grave difficulté semblent être aussi plus nombreux aujourd'hui qu'il y a quatre ans (4,4% en 2004 contre 4,9% en 2008). La part des lecteurs médiocres a elle aussi augmenté. Alors qu'ils étaient 9,5% en 2004 ils sont 9,8% en 2008. Toutefois, cette légère hausse doit être relativisée compte tenu des erreurs de mesure en ce qui concerne la correction. En effet, les épreuves étaient alors corrigées directement après les passations dans les centres d'accueil. Or, une analyse d'un échantillon de 50 000 jeunes démontre que la qualité des corrections augmente d'une année sur l'autre. Il semblerait que les correcteurs aient tendance à

légèrement surestimer le niveau réel des performances des jeunes lors de l'épreuve des traitements complexes. Le nouveau dispositif qui recueille de manière informatisée les réponses permettra ainsi d'écarter ce problème de correction.

c) Evolution des capacités à utiliser la langue écrite chez les jeunes de 15-17 ans

A l'échelle internationale, les résultats de l'ensemble des évaluations semblent être, là aussi, stables dans le temps.

D'une manière générale, en ce qui concerne l'enquête PISA, la moyenne des scores en France en compréhension de l'écrit ne connaît pas d'évolution significative entre 2000 et 2009. Il y a neuf ans, le score moyen des jeunes Français âgés de 15 ans était de 505 contre 496 aujourd'hui. Toutefois, malgré cette stabilité générale des scores, il apparaît des variations internes inhérentes à des facteurs individuels et socioculturels. Les performances ne semblent pas évoluer de la même façon pour tout le monde (nous reviendrons plus précisément sur ce point dans la partie suivante). Les résultats de cette enquête font aussi apparaître que depuis 2000, le taux de non-réponse des Français, plus important pour les questions ouvertes, est toujours supérieur à celui de la moyenne des pays de l'OCDE.

L'étude menée par PIRLS, qui mesure la compétence en lecture des élèves en quatrième année de scolarité obligatoire (élèves de CM1 pour la France), montre une stabilité des résultats entre 2001 et 2006 pour la France, tout comme l'enquête précédente. Bien que nous constatons une baisse de 3 points entre ces deux dates (avec un score de 525 en 2001 et de 522 en 2006), cette différence ne semble pas statistiquement significative. Les résultats indiquent, plus précisément, qu'en lecture, le niveau des écoliers français se situe en dessous de la moyenne des pays de l'Union Européenne. En effet, sachant que pour l'ensemble de l'évaluation, la France obtient un score de 522 (score supérieur à la moyenne fixée à 500), elle se situe en dessous de la moyenne des pays de l'OCDE (535) et de l'Union européenne (536). Cette position semble être le résultat d'une surreprésentation des élèves français dans le quart des plus faibles (32 % contre 25 % au niveau européen) et d'une sous-représentation dans celui des plus forts (17 % contre 25 %). Sachant que les compétences en lecture sont évaluées par des QCM et des productions écrites et que, pour chaque compétence évaluée, une échelle a été construite en fixant une moyenne européenne à 500, il s'avère que, pour les élèves français, quelle que soit la compétence évaluée, leurs scores se situent toujours en dessous de

la moyenne. Ainsi, ils obtiennent 497 points lorsqu'il s'agit de « prélever », 483 points pour « inférer » et 480 points pour « interpréter et apprécier ». Cette dernière compétence semble être la plus difficile pour nos élèves. Là aussi, cette enquête appuie le constat du PISA concernant le phénomène des non-réponses qui est plus important en France lorsqu'il s'agit de rédiger (2% de non réponses aux QCM contre 10% aux questions ouvertes).

Il semble donc qu'à court terme le niveau en lecture soit resté stable dans le temps quelle que soit la dimension de la lecture évaluée, la population de référence et quel que soit le degré de difficulté considéré. Qu'en est-il à plus long terme ? Les protocoles d'évaluation changeant régulièrement il nous est impossible de voir une évolution sur le long terme. Mener une étude comparative dans le temps des acquis des élèves n'est pas chose facile. En effet, pour que la comparaison soit la plus juste possible, il faut être sûr, d'une part, que les instruments et les méthodes de recueil soient les mêmes au fil du temps et, d'autre part, il faut pouvoir disposer des résultats antérieurs. La rigueur méthodologique nécessaire à ce genre d'étude fait que le nombre de comparaisons dans le temps est limité en France. Néanmoins il existe quelques études comparatives.

2. Evolution des compétences en lecture à long terme : les études comparatives françaises

En 1997, suite à la demande de l'Observatoire national de la lecture, une étude spécifique sur les difficultés de lecture à l'entrée en 6^e a été conduite sur un échantillon d'élèves (« *Étude spécifique sur les élèves en difficulté en lecture à l'entrée en sixième* », DEP, 1999). À l'aide d'un test destiné à évaluer l'identification des mots, l'objectif est de mieux comprendre pourquoi certains élèves ne maîtrisent pas correctement les compétences mises en œuvre dans l'acte de lire à l'arrivée au collège. Dix ans plus tard, en 2007, cette épreuve a été reconduite afin de voir de quelle façon ces compétences ont évolué dans le temps. L'étude indique que la proportion des élèves en difficulté a augmenté, passant de 15% à 19%. Lors de cette évaluation, différentes compétences ont été testées : la capacité d'attention des élèves, le fonctionnement de la voie directe et indirecte de l'identification des mots, la sensibilité phonologique, la maîtrise de l'orthographe, la sensibilité lexicale, l'aptitude à utiliser le contexte et la morphologie pour identifier des mots et la compréhension

écrite. Les résultats ont été analysés selon sept indicateurs constituant chacun d'eux un groupe de lecteurs : la rapidité de traitement, les connaissances lexicales sur les mots rares ou fréquents, les connaissances phonologiques, morphologiques et orthographiques et la compréhension d'énoncés. La comparaison des résultats montre que les scores obtenus en connaissances phonologiques et morphologiques et lexicales sur mots fréquents sont stables depuis 1997. C'est aussi le cas en ce qui concerne la rapidité des élèves. En revanche, les élèves de 2007 obtiennent de moins bons résultats en vocabulaire, orthographe et compréhension. Ces derniers disposent d'un vocabulaire plus pauvre, d'une moins grande maîtrise de l'orthographe et d'une plus faible compréhension d'écrit. Par conséquent, il semble qu'en dix ans, ce sont davantage les compétences langagières des élèves qui ont diminué plutôt que la maîtrise des mécanismes de base en lecture.

Cette même année, la DEPP a repris une enquête précédemment menée en 1987 afin de mesurer l'évolution des acquis des élèves en fin de CM2 en lecture, calcul et orthographe à vingt ans d'intervalle (« *Lire, écrire, compter : les performances des élèves de CM2 à vingt ans d'intervalle 1987-2007* », DEPP, 2008). Dans le cas de la lecture et des mathématiques, il existe un point intermédiaire de comparaison en 1997. En effet, dix ans auparavant, les épreuves de lecture avaient déjà été reprises. En 2007, la DEPP, reproduisant une enquête précédemment menée en 1987 puis en 1997, constate une baisse significative des performances des élèves de CM2 en lecture à vingt ans d'intervalle. D'une manière plus précise, les résultats de cette étude montrent qu'en moyenne les scores obtenus en lecture sont stables entre 1987 et 1997 mais diminuent entre 1997 et 2007 (- 0,37 points). Précisons que cette baisse est plus marquée pour les élèves les plus faibles. Par ailleurs, il apparaît que le taux de non-réponses aux questions ouvertes ne cesse d'augmenter. Ces deux derniers constats ont déjà été pointés du doigt par les évaluations internationales PISA et PIRLS.

Parmi les 35 pays participants à l'enquête PIRLS en 2001, seize d'entre eux ont participé à l'étude sur la « Reading Literacy » en 1991. Parmi ces derniers, sept (dont la France) ont passé les deux évaluations avec des élèves d'âge moyen et de niveau scolaire équivalent. Il est donc possible, tout en restant prudent²⁴, d'observer les performances des ces pays d'une enquête sur l'autre. Comme pour PIRLS, la moyenne est fixée à 500. Rappelons que, lors de cette enquête, la France a obtenu un score moyen de 525 en 2001 et de 522 en

²⁴ Bien que la population testée soit sensiblement la même, que la méthodologie de l'évaluation soit conçue selon les mêmes principes et que les techniques de recueil et de traitement des données suivent la même méthodologie, les épreuves restent néanmoins différentes.

2006. En 1991, les élèves de la « Reading Literacy » ont obtenu quant à eux 531. Il semble donc qu'il y ait une légère baisse des performances. Néanmoins cette comparaison reste fragile pour les raisons énoncées ci-dessus.

En **conclusion**, il semble qu'à court terme, en France, le niveau en lecture des élèves soit resté stable dans le temps. En revanche, il semble qu'à moyen terme, les performances en lecture aient légèrement diminué et que les compétences aient changé. Il apparaît que les élèves d'aujourd'hui maîtrisent moins bien les compétences langagières (compréhension, interprétation et analyse de texte) qu'autrefois. L'intérêt des comparaisons internationales est de pouvoir comparer les performances et les compétences des élèves de plusieurs systèmes éducatifs. En dépit d'une stabilité des scores moyens de la France depuis les années 2000, les résultats font apparaître des variations, caractéristiques de notre pays, dans les performances des élèves. Ainsi, il apparaît que depuis quelques années, le taux de non-réponses aux questions ouvertes ne cesse d'augmenter. Il semble que nos élèves rencontrent davantage de difficulté lorsqu'il s'agit de rédiger une réponse. Par ailleurs, un autre fait important, et non des moindres, est l'existence d'inégalités dues à des facteurs individuels et socioculturels dans les performances des élèves et leur évolution. En effet, il semble que, lorsque le niveau baisse, il ne baisse pas pour tous de la même façon. Il y aurait de grandes inégalités en ce qui concerne les performances. C'est ce que nous proposons de voir maintenant.

IV. Des inégalités dans les performances en lecture

Outre une légère baisse du niveau en lecture, il semble que les performances en lecture n'évoluent pas de la même manière pour tout le monde.

Au-delà du classement des performances des pays, les études internationales permettent de comparer les compétences des élèves de plusieurs pays en mettant en relation les performances avec des facteurs socio-économiques (caractéristiques socio-économiques des élèves et des pays) et politico-éducatifs (politiques éducatives et pratiques pédagogiques). Les résultats à ces évaluations permettent de mettre en exergue une différence des compétences en

lecture selon des facteurs individuels et/ou socio-économiques, tantôt communes à tous les pays, tantôt propres à la France.

1. Une inégalité commune à tous les pays : le genre

En effet, en 2010, en France, 10% des filles ne maîtrisent pas les compétences de base en français en fin d'école primaire contre 15% des garçons.

A l'issue de la scolarité obligatoire, le constat est le même : 12,6% de garçons rencontrent des difficultés en lecture lors des tests de la JDC contre 8,6% de filles. Ils réussissent moins bien les épreuves de compréhension et présentent plus souvent un déficit des mécanismes de base de traitement du langage écrit. Par contre, au niveau lexical, leurs performances sont comparables.

En 2007, 23% des garçons ont des difficultés de lecture à l'entrée en 6^{ème} contre 14,9% des filles. Cet écart de performances selon le genre est aussi valable les années précédentes. En effet, dix ans auparavant, le constat était le même: 19% des garçons avaient des difficultés de lecture à l'entrée de la 6^{ème} contre 10,6% des filles.

L'étude comparative « *Lire, écrire, compter* » à vingt ans d'intervalle confirme ce constat. Que ce soit en 1987 ou en 2007, les filles réussissent mieux que les garçons en lecture.

Même au niveau international, les filles sont meilleures.

En effet, lors de l'évaluation réalisée dans le cadre du PISA, les filles obtiennent en France un score de 515 contre 475 pour les garçons. De plus, il apparaît que les performances des garçons entre 2000 et 2009 sont en baisse alors que celles des filles restent stables.

L'enquête de PIRLS confirme ce constat en indiquant un écart de 11 points en faveur des filles dans tous les pays, sauf le Luxembourg et l'Espagne. En revanche, contrairement aux résultats de PISA, cet écart est le même en comparaison avec la session 2001.

Au-delà du genre, des différences dans les performances, davantage propres à notre pays, apparaissent selon le milieu social ou le type d'établissement fréquenté.

2. Des inégalités caractéristiques de notre pays

a) Des inégalités selon le milieu social

L'enquête de la DEPP indiquant une baisse moyenne des scores en lecture depuis 1987 précise que la dispersion des résultats est plus élevée en 2007 qu'en 1987 ou 1997. Cet accroissement des écarts est dû au fait que la baisse observée est plus forte pour les élèves issus de classes populaires (- 0,8 points) que pour ceux issus de classes moyennes ou élevées (- 0,1 points).

Si on regarde du côté des retards à l'entrée en sixième, on constate qu'en 2009, les enfants d'inactifs sont plus nombreux à être en retard que les enfants de cadres (31% contre 4%).

A l'échelle internationale, PISA indique que la France est l'un des pays où le milieu social joue le plus dans la détermination du niveau scolaire. Sur les 30 pays participants, c'est celui où les différences de performances entre les élèves issus d'un milieu social favorisé et ceux d'un milieu défavorisé sont les plus importantes. En effet, en compréhension de l'écrit, les élèves dont le statut économique, social et culturel est faible obtiennent en moyenne un score de 432 contre 544 pour ceux dont le statut économique, social et culturel est plus élevé. De plus, en comparaison avec les précédents cycles d'évaluation, il apparaît que cet écart ne cesse de s'accroître.

b) Des inégalités selon le type d'établissement fréquenté

De façon générale, selon les chiffres de l'Education nationale, à la rentrée 2008, 12% des élèves venant d'une école privée avaient au moins un an de retard à l'entrée en 6^{ème} contre 16% pour ceux issus d'une école publique et 30% pour ceux sortant d'une école située en RAR (Réseau Ambition Réussite).

En 2010, l'évaluation de la maîtrise des compétences de base en français fait ressortir qu'en fin de CM2, 12% des élèves scolarisés dans un établissement public hors éducation prioritaire ne maîtrisent pas les compétences exigées par les programmes. Ils sont en moyenne un quart dans les établissements RAR et RRS (Réseaux de Réussite Scolaire).

L'enquête de PIRLS confirme ce fait puisqu'elle indique une différence entre les scores en défaveur des élèves du public ZEP (478) par rapport à ceux du public hors ZEP (525) et du privé (536).

Précisons que, même si nous avons abordé de façon distincte les inégalités de performances selon le milieu social et le type d'établissement fréquenté, ces deux facteurs sont liés. En effet, d'après les statistiques nationales (« *Repères et références statistiques* », 2008), les collégiens scolarisés dans des RAR et des RRS sont essentiellement d'origine sociale défavorisée. Ainsi, environ 74% des élèves en RAR ont des parents ouvriers ou inactifs, contre 58% en RRS et 35% en établissement classés hors Education Prioritaire (EP).

Outre le fait que ces résultats confirment le fait d'un déterminisme social bien connu depuis les travaux de Bourdieu, les dernières études françaises et internationales font émerger une nouvelle dimension dans les inégalités. En effet, il semble se dessiner depuis quelques années un accroissement des écarts entre les élèves les plus forts et les élèves les plus faibles.

c) Des inégalités entre les plus faibles et les meilleurs

L'enquête menée par la DEPP sur les compétences en lecture à vingt ans d'intervalle indique que la baisse constatée du niveau en lecture est davantage marquée pour les élèves les plus faibles : en 2007, 21% d'élèves se situent au niveau de compétence des 10% des élèves les plus en difficulté en 1987. A l'autre bout, 8% des élèves en 2007 ont un niveau supérieur à celui des 10% les plus forts en 1987. Autrement dit, il semble qu'en vingt ans, la proportion d'élèves les plus faibles ait augmenté alors que, à l'inverse, celle des plus forts a diminué.

PISA confirme ce constat en précisant que la France fait partie, avec Israël, la Belgique et l'Autriche, des pays où l'écart est le plus fort entre les meilleurs élèves et les plus faibles. En lecture, la part des élèves les plus faibles est supérieure à la moyenne de l'OCDE : 7,9% des jeunes Français âgés de 15 ans ont un très faible niveau en lecture contre 5,7% en général. Aux niveaux les plus élevés, 9,6% ont un très bon niveau en France contre 8% pour l'OCDE. De plus, il apparaît que la proportion de jeunes ayant un faible niveau de compétence est en constante augmentation alors que celle des jeunes ayant un niveau plus élevé reste stable. En effet, les jeunes les plus faibles sont quasiment deux fois plus nombreux aujourd'hui (7.9%)

qu'hier (4.2% en 2000), alors que la proportion des plus forts n'a pas sensiblement évolué (elle était de 8,5% en 2000 contre 9,6% en 2009). Il semble donc, qu'en dépit d'une stabilité générale des scores moyens en France, la proportion de jeunes en difficulté de lecture s'est accrue depuis 2000.

En **résumé**, outre la mesure de l'évolution des acquis des élèves en lecture, l'examen des résultats obtenus aux différents types d'évaluation permet de mettre en avant le fait qu'il existe des inégalités de performances inhérentes aux populations testées²⁵. Alors que certaines sont observables en France comme à l'étranger, d'autres semblent être davantage caractéristiques du système éducatif français. Ainsi, que ce soit en France ou ailleurs, les performances et les compétences en lecture varient selon le genre : les filles rencontrent moins de difficulté que les garçons. En revanche, les enquêtes internationales font ressortir deux caractéristiques propres à notre pays. D'abord, la France semble être un des pays dans lequel le milieu social influence le plus les performances : les élèves issus d'un milieu social défavorisé obtiennent de plus mauvais résultats que ceux issus d'un milieu plus favorisé. De même, la France semble être l'un des pays où l'écart entre les élèves les plus faibles et les plus forts est le plus important. De plus, cet écart ne cesse de s'accroître.

Ce constat rejoint celui déjà émis quelques années auparavant, en 1989, par les sociologues Baudelot et Establet qui, au terme d'un examen très large des données²⁶, indiquaient que le niveau ait monté mais pas pour tous, ni partout également. Ils concluaient en effet que l'amélioration des performances scolaires, liée à l'élévation du niveau d'instruction, n'ait concerné que l'élite : *« L'écart se creuse entre un peloton de tête plus étoffé qu'hier, plus performant que jamais ; et les autres, moins nombreux qu'hier, mais plus démunis que jamais »*.

²⁵ L'école semble donc échouer dans sa mission pour garantir « à tous les élèves l'apprentissage et la maîtrise de la langue française » et contribuer à « l'égalité des chances » (art. L111-1 du code de l'éducation).

²⁶ Résultats des conscrits lors des tests passés au moment de leur incorporation dans l'armée tests mesurant des aptitudes intellectuelles telles que le raisonnement, la logique ou encore le vocabulaire dans « *Le niveau monte, réfutation d'une vieille idée concernant la prétendue décadence de nos écoles* », Baudelot et Establet, 1989.

CONCLUSION

Alors que, d'un côté, on ne cesse de prôner son importance, on déplore, de l'autre, un nombre croissant de mauvais lecteurs et on considère que le niveau en lecture a baissé. La volonté de quantifier ces élèves a donné naissance à de nombreux dispositifs, mis en place dans les années 1990, pour évaluer les performances et les compétences en lecture des élèves.

L'examen des résultats issus des différentes évaluations semble indiquer que 12% à 25% des jeunes éprouvent des difficultés en lecture en fin d'école primaire ou en fin de parcours obligatoire, dont 5% rencontrent de sérieuses difficultés. Autrement dit, un quart des élèves ne maîtriseraient pas les compétences de base dont la lecture. Le niveau en lecture semble donc insuffisant et l'étude internationale PIRLS confirme ce constat en précisant qu'en France, les élèves de CM1 sont surreprésentés dans les niveaux les plus faibles et sous-représentés dans les niveaux les plus hauts. De tels chiffres relancent donc le débat sur la baisse du niveau. Pour autant, ces résultats ne semblent pas nouveaux. En effet, il semble qu'à court terme, les performances des élèves dans ce domaine soient restées stables dans le temps. En revanche, si on regarde les résultats sur un moyen terme (à vingt ans de distance) les performances en lecture semblent avoir légèrement diminué. Par ailleurs, cette baisse ne serait pas la même pour tous les élèves. Des facteurs liés aux caractéristiques individuelles et/ou sociales des élèves semblent avoir un impact sur les performances en lecture. En effet, il apparaît que, d'une manière générale, les filles réussissent mieux que les garçons, quel que soit l'âge, le pays ou la période considérée. Par ailleurs, les études internationales indiquent que la France est l'un des pays où le niveau social influe le plus sur les performances en lecture. Les comparaisons internationales permettent aussi de mettre en avant un fait plutôt inquiétant concernant les élèves Français : il s'avère que la France serait l'un des pays dans lequel l'écart entre les élèves les plus faibles et les élèves les plus forts est le plus important. Fait d'autant plus inquiétant que cet écart serait en constante augmentation.

En conséquence, l'école semble donc échouer dans sa mission pour garantir « à tous les élèves l'apprentissage et la maîtrise de la langue française »²⁷ ainsi que dans sa contribution à « l'égalité des chances »²⁸.

²⁷ Cf. art. L111-1 du code de l'éducation

²⁸ idem

Alors, pourquoi dans une société telle que la nôtre, où quasiment toute la population est alphabétisée et dans laquelle le niveau de formation ne cesse de croître²⁹, le niveau en lecture semble-t-il diminuer ? Que mesure-t-on exactement ?

Bien que les résultats à ces différents types d'évaluation semblent convergents, nous devons cependant être prudents quant à l'interprétation de ces données. En effet, la grande diversité des outils d'évaluation ne semble pas permettre d'informer de manière claire « *les citoyens et les usagers de l'état du système éducatif* ». Elle ne semble pas non plus permettre « *de donner aux responsables et aux acteurs les moyens d'une régulation et d'un pilotage* »³⁰.

Ces différents dispositifs ne peuvent pas, selon nous, apporter une réponse générale sur les compétences des élèves en lecture et sur leurs évolutions à travers le temps. Pourquoi ? Tout d'abord, parce que les méthodologies de traitement des données sont différentes pour chacun d'eux. Ensuite, parce que les compétences testées sont différentes d'un protocole à l'autre, même s'ils prétendent tous mesurer la lecture. Enfin, les objectifs visés ne sont pas les mêmes selon qu'il s'agit des évaluations nationales ou internationales.

Alors que les évaluations nationales ont pour objectif de mesurer les acquis des élèves en lecture en référence aux programmes, les évaluations internationales portent quant à elles sur la capacité des élèves à exploiter leurs connaissances et leurs compétences à l'écrit dans des situations de la vie réelle. Ainsi, ces évaluations ne s'intéressent pas au fonctionnement du système éducatif mais plutôt à ce qu'il produit. Par ailleurs, la dépendance à l'égard des programmes nationaux des évaluations françaises rend difficile toute comparaison temporelle puisque ces programmes changent régulièrement, tout comme les outils d'évaluation.

En ce qui concerne le mode de questionnement, les items des évaluations françaises se présentent la plupart du temps sous forme de QCM. Les épreuves internationales montrent que lorsqu'il s'agit de rédiger une réponse, les élèves français préfèrent s'abstenir. Cette différence méthodologique peut produire alors une variation des performances des élèves, qui pourrait expliquer l'augmentation croissante des taux de non réponses à ces évaluations, ce qui influence donc le niveau en lecture de nos élèves.

²⁹ Le nombre d'élèves sans qualification et sans diplôme, c'est-à-dire sans un minimum d'acquis et de compétences notamment dans le domaine de la maîtrise de la langue est en baisse. Ainsi, alors qu'en 1965, 35% des élèves sortant sont sans qualification (élèves ayant un niveau inférieur au CAP), ils sont 22% en 1975, 11% en 1989, 8% en 1996 et représentent 6% en 2008. Il en est de même lorsqu'on observe la proportion de jeunes âgés de 20 à 24 ans qui déclarent n'avoir aucun diplôme ou un seulement un brevet. En effet, en 1991, ce taux s'élève à 31%. Il passe à 23% en 1996, puis 18% en 2001 et 17% en 2009.

³⁰ Rapport du Haut Conseil de l'Ecole. (2007). « L'école primaire, bilan 2007 ».

D'une manière générale, bien que les protocoles (nationaux ou internationaux) soient des épreuves standardisées, que les méthodologies de recueil et de traitement soient particulièrement précises, les conditions de passation peuvent être différentes car ce sont les professeurs qui font passer les épreuves.

Enfin, ces différents dispositifs tendent à mesurer les compétences de base en lecture. Or, nous avons constaté que les résultats sont différents selon les compétences mises en jeu dans l'acte de lire. C'est ainsi que la proportion d'élèves ne maîtrisant pas les compétences de base en lecture en fin d'école primaire s'élève à 25% lors des évaluations nationales, alors que pour d'autres évaluations (évaluations-bilan) elle tombe à 12%. Le taux d'élèves en difficulté de lecture semble donc être différent selon la définition que l'on adopte du savoir lire. Ainsi, de quoi parle-t-on lorsqu'on dit qu'un élève sait lire ?

Tout comme sa finalité, l'apprentissage de la lecture et l'acte même de lire n'ont pas toujours été définis de la même façon. Lire et apprendre à lire sont envisagés de manière différente selon l'époque à laquelle on se situe. On n'enseigne pas à lire de la même façon sous l'Ancien Régime et sous la IIIe République. Egalement, le savoir-lire au XVIIe siècle ne se définit pas de la même façon qu'au XXIe siècle. Pendant longtemps, on s'est surtout préoccupé de la finalité de la lecture et de son apprentissage. De nouvelles disciplines dans les années 1970 ont fait de l'acte de lire leur objet de recherche et les études se sont tournées vers les processus qui sous-tendent la lecture.

Chapitre 2



L'APPRENTISSAGE DE LA LECTURE

Plan du chapitre

INTRODUCTION	51
I. Les premiers modèles de lecture	51
1. Un exemple de modèle descendant : Le modèle de Goodman (1967)	52
2. Un exemple de modèle ascendant : Le modèle de Gough (1972)	53
3. Un exemple de modèle interactif : <i>Le modèle de Rumelhart (1977)</i>	55
II. Les mécanismes nécessaires à la lecture	56
1. L'identification des mots écrits	56
a) Le lexique mental	57
b) L'accès au lexique par la voie grapho-sémantique	57
c) L'accès au lexique par la voie grapho-phonétique	58
d) L'accès au lexique par les deux voies	59
2. La compréhension	59
III. L'apprentissage de la lecture	60
1. Les modèles d'identification des mots écrits	60
a) Les modèles développementaux	61
b) Les modèles par analogie	64
c) Les modèles connexionnistes	64
2. Les modèles de compréhension : l'exemple du modèle de Giasson (1990)	65
IV. Les difficultés en lecture	66
1. L'approche organique	66
2. L'approche psychanalytique	67
3. L'approche pédagogique	67
4. L'approche psycholinguistique	68
5. L'approche socio-constructiviste	69
6. L'approche sociologique	69
V. Caractéristiques du français	71
1. Les relations graphèmes-phonèmes	71
2. La structure syllabique	72
CONCLUSION	74

INTRODUCTION

L'apprentissage de la lecture est un processus complexe dans le parcours scolaire d'un élève. En effet, lors du passage en cycle 2, il va devoir comprendre que la langue écrite est une représentation de la langue orale, c'est-à-dire qu'elle est porteuse de son et de sens et va devoir faire appel à un certain nombre de capacités (attention, mémoire), de savoirs (connaissance des lettres, des sons) et de savoir-faire (correspondances grapho-phonémiques).

Comme nous l'avons vu précédemment, la lecture devient dans les années 70 une priorité du fait de l'échec scolaire et l'illettrisme. Cette même période voit émerger de nouvelles disciplines (psychologie cognitive, psycholinguistique et neuropsychologie) qui vont s'intéresser plus particulièrement à la lecture et notamment aux traitements de l'information. Le nombre de recherches visant à étudier les processus cognitifs qui soutiennent l'acte de lire va se développer et de nombreux modèles rendant compte du fonctionnement normal chez le lecteur débutant ou expert vont être proposés.

Dans le prolongement des travaux anglo-saxons des décennies antérieures, de très nombreuses recherches relevant des sciences cognitives apportent en effet une contribution majeure à la compréhension des mécanismes impliqués dans l'acquisition normale de la lecture. Ces recherches fournissent alors les connaissances et les cadres théoriques permettant de construire des outils d'évaluation davantage centrés sur l'analyse des processus que sur les performances.

I. Les premiers modèles de lecture

L'émergence de la psychologie cognitive dans les années 60 entraîne un développement des recherches en lecture. Durant les années 70, plusieurs tentatives de modélisation de la lecture ont eu lieu afin de décrire le fonctionnement cognitif du système en précisant les étapes de traitement de l'information (diffusion de l'information sérielle ou en parallèle, direction des flux d'information descendants ou ascendants).

A l'heure actuelle, on distingue trois modèles de lecture courants : les modèles descendants (ou « top-down »), ascendants (ou « bottom-up ») et interactifs. Les premiers sont dits descendants car le traitement de l'information s'effectue des niveaux cognitifs supérieurs vers les niveaux inférieurs. Autrement dit, la compréhension précède ou guide

l'identification des mots écrits. A l'inverse, dans les modèles ascendants le traitement de l'information s'effectue de bas en haut c'est-à-dire que les processus d'identification précèdent la compréhension. Les modèles interactifs, quant à eux, postulent que l'information est traitée par des processus descendants ou ascendants en fonction du degré d'information du contexte. Ainsi, le lecteur aurait recours aux procédures top-down lorsque l'information contextuelle serait importante.

Pour illustrer nos propos nous avons choisi parmi les modèles de lecture les plus connus, le modèle descendant de Goodman (1967), le modèle ascendant de Gough (1972) et le modèle interactif de Rumelhart (1977).

1. Un exemple de modèle descendant : Le modèle de Goodman (1967)

Le modèle de Goodman envisage l'acte lexique comme une succession d'opérations qui va de la perception visuelle jusqu'au décodage du message. L'objectif de Goodman est de montrer l'importance de la construction du sens dans la mise en correspondance des graphèmes et des phonèmes. Selon lui, le mécanisme de base de la lecture est la recherche du sens à l'aide du contexte. C'est seulement dans un second temps que l'information visuelle viendrait confirmer les hypothèses du lecteur. Pour lui, la lecture est « un jeu de devinette psycholinguistique », c'est-à-dire qu'elle est faite d'anticipations. L'identification des mots ne sert qu'à la vérification des hypothèses du lecteur. Autrement dit, le bon lecteur serait celui qui, grâce à des hypothèses fondées sur le contexte, sélectionne les traits visuels les plus pertinents pour faire du sens. L'acte de lire est donc un raisonnement, une mobilisation des connaissances, des prédictions sémantiques, une utilisation du contexte et une formulation d'hypothèses.

Goodman dans son modèle propose une succession d'étapes mettant en jeu des processus visuels, perceptifs, syntaxiques et sémantiques. Ce modèle comporte dix étapes. La première opération débute par un balayage visuel du matériel. Ensuite, le lecteur va fixer son regard sur certains éléments. A la suite de ça, il va sélectionner des indices graphiques qui vont lui permettre de se constituer une image perceptive du texte. Le résultat de la perception dépend donc des indices perçus et de ce qu'il s'attend à voir. Lors de l'étape d'après, le lecteur va chercher des indices syntaxiques, sémantiques et phonologiques qui vont lui

permettre d'enrichir cette image perceptive. C'est à ce moment que le lecteur va deviner ou essayer de faire un choix cohérent avec les indices graphiques prélevés. S'il y parvient, il va stocker son choix dans une mémoire à moyen terme. S'il n'y arrive pas, alors il réessaie et en dernier recours revient plus haut dans le texte. Une fois que le choix est fait, il va en vérifier la cohérence sémantique et grammaticale avec le contexte. Si sa réponse est cohérente alors elle va être stockée dans la mémoire à long terme. Si son choix n'est pas en cohérence syntaxique ou sémantique, alors il revient en arrière pour comprendre pourquoi. S'il trouve, il repart de ce point et ré-effectue les étapes précédentes. S'il ne trouve pas, il avance un peu dans sa lecture pour essayer de résoudre l'anomalie.

Goodman, dans son modèle indique que les bons lecteurs commettent des erreurs en cohérence avec le contexte alors que les apprentis lecteurs font plutôt des erreurs de confusion visuelle. Cette hypothèse va à l'encontre de celle émise par Stanovich (1981) qui pense qu'au contraire, le recours au contexte serait une stratégie compensatoire des mauvais lecteurs.

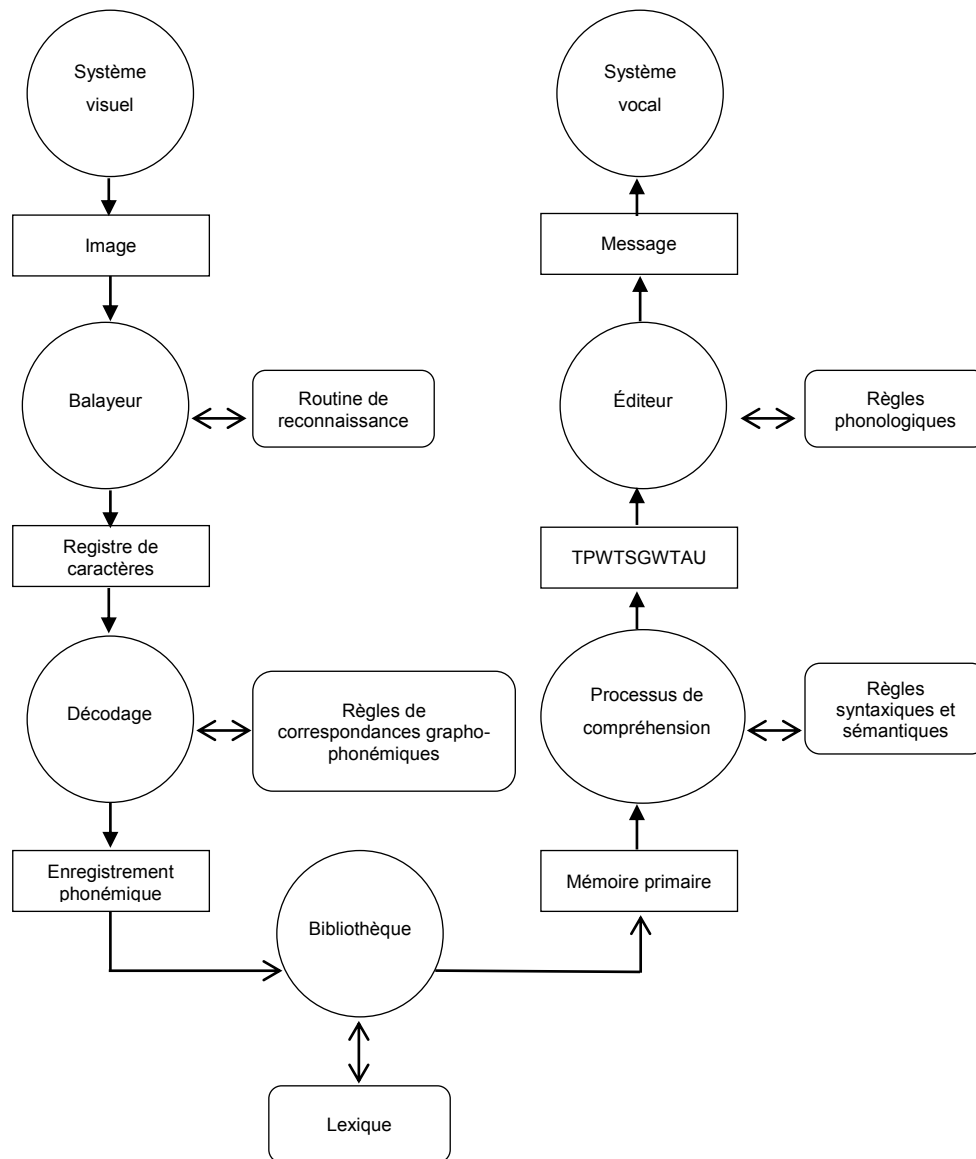
Remarquons que, bien que les flux d'informations s'effectuent des niveaux supérieurs vers les niveaux inférieurs, c'est-à-dire que la compréhension guide l'identification des mots, ce modèle ne peut se qualifier de strictement « top-down ». C'est davantage un modèle interactif dans le sens où il existe toujours des informations qui partent des niveaux inférieurs.

2. Un exemple de modèle ascendant : Le modèle de Gough (1972)

A l'inverse de Goodman, le modèle que propose Gough est de type ascendant (ou *bottom-up*) en suggérant que les flux d'information partent des niveaux perceptifs inférieurs vers les niveaux supérieurs. Autrement dit, le traitement de l'information passe d'abord par une conversion grapho-phonémique pour accéder au sens.

Le modèle de Gough se propose de retracer les événements qui se produisent durant une seconde de lecture à voix haute. Il décrit une série d'étapes qui se succèdent dans un ordre fixe (cf. figure 2).

Figure 2 : modèle de Gough (1972)



Source : The process of Reading (1983). D.C. Mitchell, p.132

D'abord, le signal visuel est stocké dans la mémoire iconique. Ensuite, les lettres sont identifiées par un balayage de la mémoire iconique. Une fois cette étape effectuée, le lecteur va transformer les codes lettres en code phonologique (Gough fait l'hypothèse de la médiation phonologique) pour identifier les mots. Cet encodage s'effectue à l'aide des règles de correspondances grapho-phonémiques stockées dans sa mémoire à long terme. Lorsqu'un mot de son lexique interne correspond à la chaîne de code phonémique, il va le stocker dans sa mémoire de travail (bibliothèque) qui va être utilisée pour comprendre la phrase. Il va procéder de la sorte jusqu'à ce que l'information soit suffisante pour être traitée

sémantiquement en fonction des bases syntaxiques et sémantiques. On est donc au niveau de la compréhension de la phrase. Le résultat obtenu est enregistré dans une mémoire appelée TPWSGWTAU : The Place Where the Sentence Go When They Are Understood (littéralement : l'endroit où va la phrase quand elle est comprise) pour être éditée en codes phonologiques sous forme d'un message qui va être oralisé.

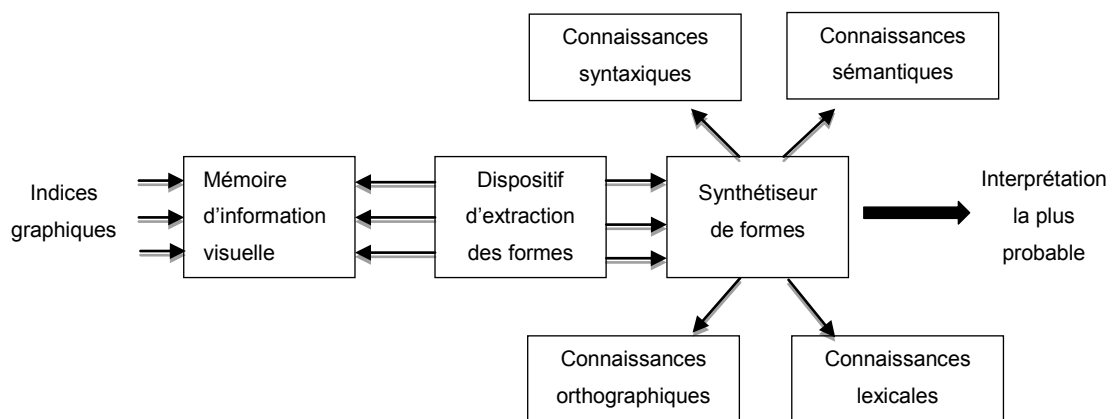
Précisons que, contrairement au modèle de Goodman, le modèle de Gough, et de façon plus générale les modèles de type *bottom-up*, ne tiennent pas compte du rôle du contexte dans l'acte de lire.

3. Un exemple de modèle interactif : Le modèle de Rumelhart (1977)

Rumelhart se distingue de Goodman et de Gough dans le sens où, selon lui, il existe une interaction entre les flux de bas niveau et de haut niveau.

Le modèle de Rumelhart suppose trois étapes principales de traitement de l'information (cf. figure 3).

Figure 3 : modèle de Rumelhart (1977)



Source : The process of Reading (1983). D.C. Mitchell, p.135

La première étape est celle de l'identification des lettres. Durant la seconde étape (appelée « synthétiseur de forme »), le lecteur, à partir de l'information visuelle et de ses connaissances, émet des hypothèses. Pour finir, l'information passe par ce qu'il appelle le « centre des messages » qui va lui permettre de valider ses hypothèses en fonction des informations : les traits des lettres, les lettres, les groupes de lettres, les informations lexicales,

syntaxiques et sémantiques. C'est en ce sens que ce modèle est interactif car durant l'acte lexique des hypothèses de bas et de haut niveau sont formulées.

En **résumé**, on remarque que l'acte lexique est essentiellement conçu de manière modulaire, c'est-à-dire que l'information est traitée de manière exhaustive par un module avant de passer à un autre. Alors que pour certains, l'acte lexique repose sur un traitement de l'information descendant, c'est-à-dire que la compréhension guide l'identification (Goodman, 1967), pour d'autres, le traitement de l'information en lecture est ascendant, c'est-à-dire que l'identification du mot précède la compréhension (Gough, 1972). Une troisième conception de l'acte lexique consiste à penser que le traitement de l'information ne s'effectuerait pas que dans une seule direction (du bas vers le haut ou du haut vers le bas). Au contraire, l'information serait traitée à la fois par des flux descendants et ascendants (Rumelhart, 1977).

Bien que ces différentes tentatives de modélisations de l'acte lexique se différencient sur plusieurs points (direction des flux d'informations, organisation du système), elles s'accordent toutefois sur le fait que la lecture suppose, d'une part une identification/reconnaissance des mots et, d'autre part, un accès à la signification.

En effet, l'acte de lire implique deux composantes : la reconnaissance des mots et la compréhension. Gough & Tunmer (1986) proposent la célèbre équation : $L=R*C$ où L est la lecture, R la reconnaissance des mots et C la compréhension. Cette formule signifie que l'acte de lire nécessite ces deux dimensions, mécanismes de base de l'acte de lire.

Ainsi, maintenant que nous avons une idée de la façon dont le système cognitif fonctionne lors de la lecture, nous allons nous intéresser plus précisément aux mécanismes de base impliqués dans l'acte de lire : l'identification des mots écrits et la compréhension.

II. Les mécanismes nécessaires à la lecture

1. L'identification des mots écrits

Afin d'accéder au sens, le lecteur qu'il soit débutant ou expert doit avoir auparavant identifié/reconnu le mot. Cette étape est essentielle pour arriver à la compréhension. Nous emploierons ici de façon distincte ces deux termes. En effet, pour le lecteur expert, nous

utiliserons plutôt le terme de « reconnaissance » puisque pour reconnaître un mot il faut déjà l'avoir vu auparavant. Ainsi pour lui, la reconnaissance d'un mot consiste, à partir de ses caractéristiques (visuelles et/ou graphiques), à le retrouver en mémoire dans son lexique mental. En revanche, nous parlerons d'identification pour le lecteur débutant qui entre juste dans le monde de l'écrit.

a) Le lexique mental

La notion de « lexique mental » (ou lexique interne) désigne une composante du système de traitement de l'information qui contient les connaissances du lecteur à propos des mots. Elle peut être imaginée comme une sorte de dictionnaire interne dans lequel sont stockés tous les mots connus auxquels sont associées leurs propres particularités orthographiques, phonologiques, morphologiques, syntaxiques et sémantiques. Ainsi, reconnaître un mot revient à chercher sa représentation lexicale dans son lexique interne. Le lexique mental est propre à chaque individu car il se construit en fonction des expériences de lecture et par conséquent, la liste de mots se complète sans cesse.

Le modèle proposé par Forster (1976) nous permet de mieux comprendre le fonctionnement du lexique interne. Dans son modèle, il distingue le lexique à proprement parler (appelé le fichier central), des fichiers d'accès au lexique. Ces derniers peuvent être de trois types : orthographique, sémantique ou syntaxique. Chaque fichier contient les caractéristiques du stimulus associées au mot qui sont transcrites en code orthographique, sémantique et syntaxique. Les mots du lexique interne seraient donc classés dans ces fichiers en fonction de leurs propriétés formelles et sont classés par ordre de fréquence décroissante, ce qui implique que la recherche de l'entrée lexicale est d'autant plus lente que le mot est rare. Ainsi, lorsqu'un mot est lu, ses caractéristiques vont être extraites et encodées. Ensuite, une recherche active est effectuée dans le fichier orthographique pour trouver une correspondance entre les deux codes (celui extrait du stimulus et celui qui se trouve dans le fichier d'accès).

b) L'accès au lexique par la voie grapho-sémantique

C'est en 1969 que Morton effectue une première modélisation de l'accès au lexique à l'aide du paradigme de la décision lexicale. C'est le modèle des logogènes. Selon lui, dans le

lexique mental, chaque mot correspond à un logogène, sorte de récepteur d'informations qui dispose d'un seuil d'activation, de repos et de reconnaissance. Chaque fois que des informations compatibles avec le mot sont présentées, le niveau d'activation du logogène s'accroît et lorsque le seuil est atteint, le mot est identifié. Dans ce modèle, l'activation est due à une information issue de l'analyse perceptive du stimulus physique ou du contexte. L'accès au lexique par la voie grapho-sémantique (voie directe) signifie alors que le mot est identifié à l'aide d'un adressage direct avec le lexique mental grâce à ses caractéristiques orthographiques sans passer par la médiation phonologique, c'est-à-dire la mise en correspondance des graphèmes et des phonèmes.

Dans la lignée de Morton, Rumelhart et Mc Clelland (1981) proposent un modèle d'accès au lexique ayant une structure davantage connexionniste. Leur modèle d'activation interactive suppose une organisation en trois niveaux connectés entre eux : les traits, les lettres et les mots. Par exemple, chaque lettre est connectée aux traits et aux mots. Ainsi, la lettre b en position initiale peut être reliée au mot bateau. La présentation d'un mot entraîne alors la diffusion de l'information vers les connexions des autres niveaux pour aboutir à l'identification du mot et à sa signification. Ce modèle, qui a permis des simulations par ordinateur, a démontré un effet du voisinage orthographique¹ : un mot est moins rapidement identifié s'il possède au moins un voisin orthographique plus fréquent que lui (Grainger et Segui, 1990).

c) L'accès au lexique par la voie grapho-phonétique

Contrairement au modèle de Morton, Gough (1972) dans sa modélisation de l'acte lexique, émet l'hypothèse d'une médiation phonologique. En effet, d'après lui, l'identification d'un mot consiste à transformer les codes des lettres (graphèmes) en codes phonologiques.

L'indépendance de ces deux voies a été remise progressivement en question et l'on suppose maintenant que l'identification des mots chez le lecteur expert passe par une activation parallèle de plusieurs représentations et que chaque voie peut influencer sur l'autre.

¹ Deux mots sont des voisins orthographiques lorsqu'ils contiennent les mêmes lettres sauf une. Par exemple foie et fois.

d) L'accès au lexique par les deux voies

Le modèle à double voie de Coltheart (*Dual route model*, 1978) postule un accès au lexique associant ces deux voies. La première, la voie directe (voie grapho-sémantique), permet d'identifier un mot par la technique de l'adressage. La seconde, la voie indirecte (voie grapho-phonétique), permet l'identification par la technique de l'assemblage. Ce sont les caractéristiques du stimulus écrit qui vont déterminer l'interaction des deux voies.

Ainsi, ici aussi la question des voies d'accès au lexique n'est pas tranchée : alors que certains postulent que l'accès au lexique s'effectue par la voie grapho-phonétique, d'autres à l'inverse pensent que l'accès au lexique se fait par la voie grapho-sémantique. Quoiqu'il en soit, bien qu'il existe des divergences à ce sujet, la finalité de la lecture reste la compréhension.

2. La compréhension

Les recherches sur la compréhension des textes écrits montrent que l'élaboration de la signification globale d'un texte résulte d'une interaction complexe entre les caractéristiques du lecteur, qu'il s'agisse de son projet de lecture, des connaissances dont il dispose préalablement à la lecture ou des capacités cognitives qu'il est capable de mettre en œuvre aux différentes phases de l'activité (Fayol, 1992 ; 1997), du texte (les mots, les phrases, les paragraphes, la ponctuation, etc.) et du contexte de la situation de lecture.

En **résumé**, l'acte de lire fait appel à deux mécanismes de base : d'une part, l'identification de mot et d'autre part, la compréhension. L'identification du mot consiste à chercher sa représentation lexicale dans son lexique interne. L'accès au lexique peut s'effectuer soit par la voie grapho-phonétique soit par la voie grapho-sémantique.

Ainsi, de la même façon qu'il existe plusieurs conceptions de l'acte lexique, il existe là aussi plusieurs points de vues en ce qui concerne les voies d'accès au lexique. Ces différentes conceptions ont par la suite amené les chercheurs à formuler différents modèles d'apprentissage de la lecture.

III. L'apprentissage de la lecture

Comme nous l'avons vu précédemment, depuis les années 80, la lecture est davantage conçue de façon modulaire. L'information est traitée de manière exhaustive par un module avant d'être traitée par le module suivant. La reconnaissance d'un mot se fait par des voies d'accès à un lexique interne (module lexical) : la voie grapho-phonétique (ou indirecte) et la voie grapho-sémantique (ou directe). Basés sur ces voies d'accès, plusieurs chercheurs ont ainsi proposé des modèles d'apprentissage de la lecture permettant de rendre compte de l'identification des mots écrits.

1. Les modèles d'identification des mots écrits

Le plus ancien, celui de Marsh, Friedman, Welsh et Desberg (1981), est issu de la psychologie génétique. Ces chercheurs, proposent une entrée dans l'écrit, basée sur la théorie constructiviste, en quatre étapes qui correspondent plus ou moins aux stades de développement de Piaget (1956).

Lors d'une première étape, l'enfant utiliserait une stratégie de devinette linguistique (*linguistic guessing*) qui lui permettrait de reconnaître de façon globale quelques mots familiers qu'il a appris par cœur (*rote learning of whole word*). Cette phase basée sur la perception correspond plus ou moins au stade sensori-moteur de Piaget. La seconde phase est celle de la discrimination (*discrimination net guessing*). L'enfant va deviner par « coups de filet » les mots à partir d'indices linguistiques et visuels, par comparaison avec des mots connus (stade préopératoire de Piaget). Durant la troisième étape, il va mettre en œuvre une stratégie de décodage séquentiel (*sequential decoding*) en traitant des ensembles ordonnés de lettres à travers une mise en correspondance des graphèmes et des phonèmes. A ce stade, l'enfant ne peut lire que les mots réguliers c'est-à-dire des mots dont la prononciation correcte peut être dérivée de l'orthographe sur la base d'une connaissance des correspondances les plus fréquentes entre les graphèmes et les phonèmes (ex. « maison »). Les enfants passeraient du stade 2 au stade 3 lorsqu'ils auraient trop de mots à reconnaître. Ils adopteraient ainsi une stratégie plus fiable et plus efficace. Enfin, la dernière étape, qui concerne le lecteur expert, est caractérisée par une utilisation non seulement des règles de correspondance entre sons et lettres plus complexes, mais aussi par l'utilisation de règles orthographiques (ex. « femme »).

Ici, le décodage est davantage hiérarchique (*hierarchical decoding*). Ainsi la maîtrise de l'acte lexique se résume seulement à une maîtrise des correspondances grapho-phonétiques. Les voies d'accès au lexique par voie grapho-sémantique sont totalement ignorées : dans ce type de modèle, un bon lecteur est donc un bon décodeur.

a) Les modèles développementaux

Issus des modèles à double voie, les modèles développementaux (Frith, 1985, Harris & Coltheart, 1986, Morton, 1989) ont pour objectif de rendre compte, en termes de développement, des différentes étapes par lesquelles passe l'enfant dans la reconnaissance des mots. L'apprentissage de la lecture est alors considéré comme une suite d'étapes (ou stades) dont l'ordre de succession est strict et identique pour tous les élèves. Aucune étape ne peut être sautée. Le passage d'un stade à un autre s'effectue sous l'effet de l'instruction et l'élève ne peut changer de stade que si les compétences du stade antérieur sont maîtrisées. Chaque stade suppose donc un fonctionnement spécifique et le traitement des mots diffère d'un stade à l'autre.

Ces modèles considèrent donc que l'apprentissage de la lecture se déroule de façon discontinue car la progression passe par des étapes bien distinctes contrairement à un apprentissage continu où l'évolution se fait de manière progressive. Dans le cas d'une évolution discontinue, chaque étape amène à un changement qualitatif en termes de comportements ou de stratégies alors que dans le cas d'une évolution continue, les changements sont quantitatifs (par exemple, les acquisitions du vocabulaire). Ainsi, dans les modèles développementaux, les processus cognitifs en jeu, ici dans le cas de la lecture les processus grapho-phonétiques et grapho-sémantiques, opèrent de façon indépendante.

Dans le cadre de l'apprentissage de la lecture, les différents types de modèles définissent la plupart du temps la mise en place des deux procédures d'accès au lexique en trois ou quatre étapes. Nous nous intéresserons ici plus particulièrement au modèle de Frith (1985, 1986), le plus connu et le plus cité de tous.

Le modèle d'apprentissage de la lecture de Frith s'inspire du modèle proposé par Marsh & al. (1981) mais avec trois étapes qualifiées par leurs procédures : les procédures logographique, alphabétique et orthographique.

La première étape est l'équivalent du premier stade (i.e. la devinette linguistique) du modèle de Mars & *al.* (1981). Durant cette étape, l'enfant a recours à une **procédure logographique** pour lire les mots, c'est-à-dire qu'il va reconnaître un petit nombre de mots qu'il connaît par cœur d'après leur forme visuelle. Autrement dit, l'enfant va reconnaître un mot à partir d'indices graphiques visuels saillants tels que sa longueur, sa couleur ou encore la forme de certaines lettres qui le composent. Autrement dit, pour reconnaître un mot, l'enfant va utiliser des indices contextuels et des caractéristiques du mot. Le mot est traité comme un objet visuel et non linguistique. A ce stade, l'ordre des lettres n'est pas considéré comme un indice pertinent dans la reconnaissance du mot. Ainsi, à ce stade, un enfant peut reconnaître le logo « Esso » seulement à partir de sa forme visuelle. Ce stade permet à l'enfant de développer un vocabulaire visuel et lorsqu'il rencontre trop de mots et ne peut plus les mémoriser, il passe au stade suivant.

Au cours du second stade, l'enfant va avoir recours à une **procédure alphabétique**. Si on se réfère au modèle de Marsh & *al.* (1981), le stade alphabétique correspond aux stades 3 et 4 de leur modèle. A ce niveau, l'enfant utilise ses connaissances des correspondances entre les graphèmes et phonèmes pour lire un mot. Contrairement au stade précédent, l'ordre des lettres et leur valeur phonique jouent un rôle crucial dans la reconnaissance du mot. L'enfant va alors effectuer un décodage séquentiel, graphème par graphème sans tenir compte des lettres environnantes. Il va appliquer les règles simples de conversion grapho-phonémique lui permettant ainsi de lire des mots longs ou rares. En revanche, à ce niveau, le traitement de graphèmes complexes à prononciation irrégulière (monsieur, femme, ...) et contextuels (vague vs sage, poison vs pinson) vont l'emmenner à commettre des erreurs. Lorsque l'enfant rencontre des mots irréguliers et n'arrive pas à faire la conversion, il passe au stade suivant.

Le troisième et dernier stade est caractérisé par une **procédure orthographique**. L'enfant va traiter les mots à partir de la voie grapho-sémantique. Ainsi, les mots sont envisagés comme des unités orthographiques et l'enfant ne passe plus par les correspondances grapho-phonémiques pour lire un mot. Autrement, dit, l'enfant va directement apparier les caractéristiques orthographiques du mot à sa représentation lexicale dans son lexique mental. A ce stade, Frith indique que l'unité de traitement n'est plus le graphème mais le morphème (par exemple pour le mot « printemps », l'enfant va reconnaître les morphèmes « prin » et « temps ») et que l'enfant peut aussi reconnaître un mot par analogie avec un autre.

Parmi les modèles développementaux (Harris & Coltheart, 1986, Morton 1989, Ehri & Wilce, 1985), certains tendent à remettre en cause l'existence et la nature de la procédure logographique. Ainsi, d'après une recherche menée par Stuart et Coltheart (1988), il apparaît que les élèves qui apprennent à lire avec déjà des connaissances phonologiques les utilisent dès le départ contrairement à ceux qui n'en ont pas. Par conséquent, la procédure logographique n'existerait pas pour cette catégorie d'élève. Ehri (1985) a montré qu'au début de l'apprentissage de la lecture, les enfants se servent davantage d'indices phonologiques que visuels. Par conséquent, si cette procédure existe, elle ne concerne pas pour autant tous les enfants. L'existence d'une procédure orthographique indépendante des deux autres procédures n'a pas non plus été réellement prouvée. Les résultats des recherches montrent que les enfants utiliseraient en même temps la procédure orthographique et alphabétique.

Alors que certains remettent en cause les procédures logographique et orthographique, l'ensemble des recherches menées sur l'apprentissage de la lecture s'accordent toutefois sur le rôle essentiel que joue la procédure alphabétique (ou médiation phonologique) dans la dynamique de l'apprentissage de la lecture. En effet, au début de l'apprentissage de la lecture, les enfants utiliseraient prioritairement la voie indirecte pour identifier les mots.

Par ailleurs, bien que ces modèles rendent compte des modifications de traitement en fonction des étapes d'acquisition, ils restent toutefois trop généraux et fournissent peu d'informations sur la dynamique de l'apprentissage. En effet, on ne sait pas réellement qu'est-ce qui motive le passage d'un stade à un autre. Les conditions de passage de l'utilisation de la procédure logographique à la procédure alphabétique puis orthographique ne sont pas clairement définies.

De plus, la notion de successivité des étapes a été remise en cause pour envisager la lecture non plus de façon séquentielle mais plutôt de manière interactive dans lesquelles les deux voies d'accès au lexique coexistent et où les processus de traitement interagissent. Les différentes modalités de traitement correspondraient à des procédures disponibles pour le lecteur.

De cette façon, Seymour (1990) propose de dépasser la perspective développementale en stades, en présentant un modèle à « double fondation » dans lequel le lecteur aurait recours simultanément aux procédures logographiques et alphabétiques pour construire son lexique orthographique. Ce modèle comporte cinq processeurs : logographique, alphabétique, la conscience linguistique, orthographique et morphographique. Selon ce modèle, l'enfant

accède au lexique par la mise en place concomitante d'un processeur visuel et d'un processeur d'assemblage phonologique.

Perfetti (1992), quant à lui, défend la coexistence d'un lexique fonctionnel (construit au cours de l'apprentissage explicite de la lecture) et d'un lexique autonome. Son modèle tente de rendre compte des différentes étapes par lesquelles l'enfant parvient à passer du premier au second.

Bien que les modèles développementaux considèrent les sources d'indices grapho-phonétiques et grapho-sémantiques, ils restent centrés sur l'identification du mot et ignorent totalement les facteurs syntaxico-sémantiques, la compréhension est renvoyée à un processus externe.

b) Les modèles par analogie

De nouveaux types de modèles mettant en cause la nature non lexicale de la procédure indirecte défendent le rôle des analogies dans la lecture (Goswami et Bryant, 1990 ; Wimmer et *al.*, 1994). Ces modèles de lecture insistent fortement sur l'ancrage de nouvelles connaissances sur des connaissances antérieures et plus spécifiquement sur les liens entre les connaissances de niveau phonologique et orthographiques. Cependant, cette dynamique continue est loin d'être corroborée par les résultats des recherches. En effet, selon la langue dans laquelle les travaux ont été effectués et même dans les études portant sur la même langue, les résultats diffèrent.

c) Les modèles connexionnistes

En rupture avec les modèles cognitivistes classiques, se développent depuis une quinzaine d'années des modèles connexionnistes, implantés sur ordinateur, qui reposent sur un fonctionnement en réseau dans lequel toutes les connaissances sont reliées. On ne parle plus de lexique mental. Ainsi, reconnaître un mot, ce n'est donc pas le retrouver en mémoire, mais plutôt recouvrer un certain état d'activation des unités qui, dans le système cognitif sont concernées par le traitement de l'information lexicale. Le plus connu reste celui de

Seidenberg et Mc Clelland (1989) dans lequel le réseau comporte des unités orthographiques d'entrée et des unités phonologiques de sortie.

Bien que ces modèles permettent de formuler des hypothèses précises sur des unités orthographiques liées à des représentations phonologiques, ils ne simulent pas l'existence, avant l'apprentissage de la lecture, du lexique mental dans sa forme phonologique et ses relations avec le système sémantique.

L'ensemble de ces modèles d'apprentissage de la lecture, qu'ils soient développementaux, interactifs ou connexionnistes ne concernent que l'identification du mot écrit. Les processus de compréhensions ne sont pas intégrés. C'est pourquoi, nous allons voir maintenant un exemple de modèle de compréhension, celui de Giasson.

2. Les modèles de compréhension : l'exemple du modèle de Giasson (1990)

Le modèle de Giasson (1990) articule trois grandes composantes qui entrent en jeu lors de la lecture : le lecteur, le texte et le contexte.

La composante « lecteur » comprend les structures et les processus qu'il met en œuvre. D'une façon générale, les structures font référence au lecteur lui-même. Elles représentent les structures cognitives et affectives. Les structures cognitives regroupent les connaissances de la langue orale (phonologiques, syntaxiques, sémantiques et pragmatiques) et du monde. Les structures affectives quant à elles sont relatives aux intentions du lecteur, à ses attitudes. En revanche, les processus font référence aux habiletés du lecteur : des micro-processus (identification des mots) aux macro-processus (compréhension globale du texte), des processus d'intégration aux processus d'élaboration (inférences) ou encore méta-cognitifs.

La composante « texte » concerne le matériel utilisé qui peut être défini selon quatre critères : l'intention de l'auteur, le genre littéraire, la structure et le contenu.

La composante « contexte » représente les conditions dans lesquelles se trouve le lecteur lorsqu'il va lire un texte. Ce contexte va influencer sur la lecture du texte et sa compréhension. Giasson distingue trois types de contexte : psychologique, social et physique. Le contexte psychologique est inhérent au lecteur (son intérêt pour lire le texte, sa motivation,

son intention de lecture). Cette dimension est importante dans le sens où la compréhension dépend de la façon dont le lecteur va aborder le texte. Le contexte social concerne l'environnement dans lequel se trouve le lecteur durant la lecture (présence d'un enseignant, de camarades, lecture individuelle, collective). Le contexte physique est relatif aux conditions matérielles dans lesquelles le lecteur se trouve lors de la lecture (bruit, temps de lecture, température, ...).

Ainsi, à partir des processus méta-cognitifs, les micro-processus vont entrer en jeu pour comprendre le sens de la phrase. Les processus d'intégration vont se charger de faire le lien entre les phrases et les macro-processus vont permettre au lecteur d'avoir une compréhension globale du texte.

IV. Les difficultés en lecture

Depuis l'émergence des premiers troubles de la lecture, de nombreux chercheurs tentent de décrire, d'expliquer et de comprendre ces difficultés de lecture. Ainsi, les causes des troubles en lecture s'expliquent de manière différente selon le courant de pensée dans lesquels ils s'inscrivent. Fijalkow (1986, 1998), après une revue des différents travaux en la matière, indique que les difficultés en lecture peuvent s'expliquer selon une perspective organiciste, psychanalytique, scolaire, psycho-linguistique et socio-constructiviste. Nous rajoutons à cette liste des facteurs sociologiques.

1. L'approche organique

Les premières difficultés de lecture émanent du corps médical vers la fin du XIXe siècle (aux alentours de 1890). Ce courant organiciste d'origine anglo-saxonne est soutenu par des neurologues et neuropsychologues. L'apprentissage de la lecture étant défini à travers le développement cognitif de l'enfant, on postule alors que les difficultés de lecture seraient d'origine biologique, médicale, résultant d'un dysfonctionnement du système nerveux central. Ainsi, un enfant qui éprouve des difficultés de lecture est celui qui souffre d'un dysfonctionnement cérébral (Orton, 1937). Ces auteurs emploieront alors le terme de

« dyslexie » pour qualifier des enfants qui ont des difficultés à lire. Leurs erreurs sont de nature pathologique et organique et on s'intéresse alors aux erreurs dites « de surface » telles que les inversions de lettres (confusion entre des lettres qui se ressemblent b, d, p, q), les inversions de sons (m, n) et les interversions (fil lu lif, lac lu cal).

Bien que ces études remontent à plus d'un demi-siècle, leur impact dans l'analyse des difficultés orales reste encore présent. « [...] elles constituent aujourd'hui encore, soixante ans plus tard, le fond commun de la pensée courante à propos des difficultés d'apprentissage de la lecture » (Fijalkow J., Pasa L. et Ragano S., 1998).

2. L'approche psychanalytique

Dans ce courant, on suppose que les difficultés de lecture ne sont pas liées au développement cérébral de l'enfant mais plutôt à son développement personnel, à son histoire. Cette perspective s'inscrit donc dans le prolongement des travaux de Freud. Ainsi, les difficultés en lecture seraient, dans ce courant, propres à chacun.

Ainsi, selon ces deux premières approches (organique et psychanalytique), les difficultés rencontrées en lecture sont inhérentes à l'enfant, qu'elles soient innées (biologiques) ou acquises (à travers son histoire).

3. L'approche pédagogique

Inspirée de l'analyse des courants explicatifs d'origine médicale, cette perspective tente d'expliquer les erreurs dans un cadre plus scolaire. Ainsi, on ne considère plus seulement les erreurs dites de surface mais on va envisager tous les types d'erreurs. Par conséquent, outre les inversions et les interversions, on trouve aussi des erreurs de substitution (mal / bal), d'omission (grand lu gand), d'insertion de répétition ou encore des erreurs de prononciation. « L'erreur est ici considérée d'une autre nature que la réponse exacte et résultant d'une insuffisance de l'élève qu'il revient au maître de corriger » (Fijalkow, 1998).

Ainsi, dans cette approche, les difficultés en lecture sont reliées à l'enseignement. On estime que l'enfant fait comme on lui apprend. Il se pose alors la question du comment fait-on et on va chercher quelle est la meilleure façon d'enseigner la lecture.

Cette approche nous amène directement au fameux débat concernant les méthodes d'enseignement de la lecture qui divise formateurs et praticiens. En effet, tandis que certains prônent la méthode syllabique (le b-a/ba, c'est-à-dire l'apprentissage des correspondances grapho-phonémiques), d'autres sont partisans de la méthode globale qui se centre davantage sur le sens. Ces dernières années, une troisième méthode est apparue, la méthode mixte qui consiste à faire progresser de front l'accès au sens et l'apprentissage des correspondances lettres/sons.

Ainsi, alors que pour certains courants de pensée, les difficultés de lecture sont de nature médicale, elles sont pour d'autres de nature pédagogique. Une autre approche voit le jour au cours des années 1970 qui va permettre de passer d'une analyse normative des erreurs à une analyse explicative. Il s'agit de l'approche psycholinguistique.

4. L'approche psycholinguistique

Le psycholinguiste Goodman (1965) se propose alors de remplacer le terme erreur, dont la connotation est trop négative, par *miscue*, c'est-à-dire « méprise » car, selon lui, la lecture repose sur une prise d'indices et l'erreur est par conséquent le résultat d'une mauvaise prise d'indices. Ce courant postule que les erreurs ne sont pas l'effet du hasard. Ainsi « *expliquer ces réponses inattendues constitue donc un moyen commode d'expliquer les réponses attendues, et plus généralement, d'expliquer ce que c'est que lire* » (Fijalkow J., Pasa L. et Ragano S., 1998). Par exemple, ces auteurs vont se demander si l'erreur commise est acceptable d'un point de vue grapho-phonétique, sémantique ou syntaxique. Dans cette approche, la question n'est ni de savoir comment est l'enfant, ni comment on doit faire mais plutôt, comment l'enfant fait.

D'autres recherches, basées sur la psycholinguistique, sont centrées sur le mot et prennent en compte ses variables. Ainsi, certains chercheurs ont tenté de voir si les erreurs varient selon la longueur ou la fréquence d'un mot. D'autres, centrés sur la syllabe, essaient de voir si les erreurs varient selon la structure syllabique d'un mot. Enfin, d'autres encore essaient de voir si les erreurs varient selon la structure grapho-phonétique d'un mot.

Jusqu'à présent, les difficultés rencontrées lors de l'apprentissage de la lecture sont envisagées soit du côté de l'enfant soit du côté des enseignants. Une autre approche consiste à considérer l'enfant certes, mais dans son contexte. C'est l'approche socio-constructiviste.

5. L'approche socio-constructiviste

L'origine de ce courant se situe dans le courant néo-piagétien qui introduit l'interaction sociale dans le développement cognitif. Vygotsky (1997) exprime l'idée d'une conception interactive du développement et va introduire le concept de la ZPD (« Zone Proximale de Développement »). Selon cette conception, le développement d'un enfant s'effectue à l'aide d'autrui. Le contexte social et les expériences vécues jouent par conséquent un rôle primordial dans l'apprentissage. Ainsi, dans cette perspective, les apprentissages en milieu scolaire seraient le résultat d'une interaction entre les élèves, l'objet de l'enseignement (ici, la langue écrite) et le contexte social (la classe).

6. L'approche sociologique

Cette approche consiste à penser que les difficultés en lecture seraient relatives à des facteurs sociologiques tels que le genre, le milieu social, etc.

En effet, filles et garçons ont des modes différents de socialisation (Duru-Bellat, 1997). Les attentes des parents ne sont pas les mêmes pour un garçon que pour une fille. Cette différence de pratiques éducatives amène progressivement l'enfant à intérioriser et à se conformer aux attentes en matière de rôle social lié à la norme. Les enfants construisent ainsi leur identité à travers ce que véhiculent les adultes.

Les recherches menées sur l'influence du milieu social montrent clairement que les enfants de milieu défavorisé éprouvent plus de difficultés de lecture que ceux issus de milieu favorisé. Empruntant le terme « habitus » à la sociologie, les chercheurs en psychologie sociale vont parler d'habitus lectural considérant que la lecture, envisagée comme un objet social, fait partie du capital culturel de chacun (Ecalte, 2002). L'habitus lectural se développe à un niveau micro-sociologique, c'est-à-dire au sein de la famille et à un niveau macro-sociologique, c'est-à-dire le contexte socio-politique. Ainsi, lorsque que l'habitus lectural est réduit, les conditions d'accès à la langue sont plus difficiles.

Un autre concept émanant de la psychologie sociale est celui des représentations sociales. Selon Jodelet (1994), les représentations sociales sont « *une forme de connaissance socialement élaborée et partagée, ayant une visée pratique et concourant à la construction d'une réalité commune à un ensemble social ou culturel* ». La construction de ces représentations s'effectue à travers les interactions. Ainsi, lors de l'apprentissage de la lecture, l'enfant élabore un ensemble de représentations à propos de cet objet. Carayon (1991)² lors d'une étude sur les représentations sociales de lecture chez les enfants de 6 à 11 ans montre que la lecture, à cet âge-là peut être associée à différentes représentations : la lecture plaisir, la lecture comme un moyen d'accès à la langue écrite, la lecture comme un travail scolaire, etc. En 1989, Chauveau et Rogovas-Chauveau constatent lors d'une étude sur les raisons d'apprendre à lire et l'utilité de la lecture, que les enfants qui se représentent la lecture comme une activité scolaire ou qui considèrent qu'apprendre à lire c'est pour savoir lire, éprouvent des difficultés d'apprentissage de la lecture l'année suivante.

En **résumé**, les difficultés en lecture peuvent être analysées sous différents angles d'approche. D'abord, on peut considérer que ces difficultés sont en réalité une pathologie, ce qui revient à penser que les difficultés en lecture sont d'origine médicale et par conséquent, inhérentes à l'enfant. Une autre explication, tend à penser que l'enfant apprend comme on lui enseigne. De ce fait, les pédagogues vont alors réfléchir sur la façon dont il faut procéder pour corriger les insuffisances de l'élève. Une troisième approche, moins descriptive va alors chercher à comprendre pourquoi l'enfant fait des erreurs, qu'est-ce qui lors de sa lecture, lui a posé des difficultés. Enfin, une dernière dimension consiste à penser que l'enfant est un être social et, par conséquent, que l'apprentissage s'effectue en interaction avec les pairs et les adultes. Selon cette approche, le contexte joue alors un rôle essentiel dans l'apprentissage de la lecture.

² Cité par Ecalte et Magnan dans *L'apprentissage de la lecture. Fonctionnement et développement cognitifs* (2002).

V. Caractéristiques du français

D'une manière générale, la plupart des recherches portant sur l'apprentissage de la lecture sont anglophones. Or, bien que l'anglais soit une langue alphabétique comme le français, cette dernière se différencie sur plusieurs points.

1. Les relations graphèmes-phonèmes

En effet, bien que ces deux langues utilisent le même système d'écriture, elles possèdent chacune leur propre orthographe et leur propre convention graphophonologique. Ainsi, alors que pour la première langue à chaque graphème correspond un phonème, celle de la seconde est plutôt opaque. En effet, le français n'est pas biunivoque : un graphème peut correspondre à plusieurs sons et un même phonème peut être transcrit de plusieurs façons.

Ainsi, Gak (1976) distingue six types de distorsions indiquant un désaccord entre le signe et le son :

- une lettre simple peut correspondre à un groupe de phonèmes, qui sous certaines conditions s'écrit par deux lettres consécutives. Par exemple, la lettre *x* dans les mots « *taxe* » et « *tocsin* ».
- un groupe de lettres peut correspondre à un seul phonème. Par exemple, *on*, *in*, *au*, *ph*, *gn*, *eau*, ...etc. C'est ce que l'on appelle les graphèmes composés.
- Une même lettre peut correspondre à deux phonèmes différents. En effet, en français, la prononciation de certains graphèmes est régie par des lois contextuelles, c'est-à-dire que leur prononciation dépend de l'endroit où elle se situe dans le mot. Par exemple, la lettre *g* va se lire différemment dans « *garçon* » et « *gitan* ».
- Un même phonème peut se retranscrire par des graphèmes différents. Par exemple le son [ɛ̃] peut s'écrire *in* (comme dans *vin*), *ein* (comme dans *sein*), *ain* (comme dans *étain*), *aim* (comme dans *daim*).
- Certaines lettres ne se prononcent pas, elles sont muettes. C'est le cas par exemple, du *h* muet (comme dans *haricot*), du *e* muet (comme dans *partie*) et des consonnes finales muettes (comme dans « *nid* » ou par exemple le morphogramme *s* qui indique le pluriel). Toutefois, il existe certains cas où ces consonnes finales muettes doivent se prononcer. Ainsi, on dira les_hommes ([ləzɔ̃mɛ]) et les / haricots ([ləʁiko]). C'est

ce qu'on appelle le phénomène de liaison. De la même façon, on dira un grand / lit ([gʁɑ̃li]) et un grand_écran ([gʁɑ̃tekrɑ̃]) ou les_yeux ([lezjø]) et les / yaourts ([lejaurt]).

- Enfin, certaines lettres n'expriment qu'une partie des sons prononcés. C'est le cas des abréviations et des sigles. Par exemple le *M.* pour dire monsieur.

Ainsi, si on considère que l'unité de traitement est le graphème, alors les correspondances graphèmes-phonèmes sont la plupart du temps régulières et prédictibles : « eau » se lira toujours [o]. Toutefois, compte tenu des particularités grapho-phonémiques du français énoncées précédemment, l'enfant peut rencontrer des difficultés au cours de l'apprentissage de la lecture lors de la mise en place de la procédure par médiation phonologique.

2. La structure syllabique

Bien que la structure infrasyllabique du français présente une certaine irrégularité, la structure syllabique, quant à elle, est plutôt régulière. En effet, dans la langue française, la structure de base, formée d'une consonne et d'une voyelle (CV), est largement représentée puisqu'elle recouvre 54,9% des syllabes (c'est le cas par exemple de la première syllabe dans le mot *petit*). Viennent ensuite les syllabes du type CVC (comme dans le mot *parmi*) qui représentent 17,1%, les syllabes CCV (comme dans le mot *printemps*) qui sont de 14,2%. Les syllabes du type VC (comme dans le mot *argentée*) sont peu nombreuses puisqu'elles représentent 1,9 %. Enfin, la catégorie « autres » représente 11,9% des syllabes. Cette catégorie comprend les syllabes composées de semi-voyelles (SV) comme par exemple le [wa] dans le monosyllabique « bois » qui est du type CSV.

Ainsi, en français, 73% des syllabes se terminent par une voyelle alors qu'en anglais 60% des syllabes se terminent par une consonne. Par conséquent, contrairement à l'anglais où la majorité des syllabes sont fermées, en français, elles sont essentiellement ouvertes³, ce qui peut entraîner lors de l'apprentissage de la lecture des erreurs de liaison et d'éliision.

³ Delattre (1965)

En **résumé**, le français se caractérise par :

- La complexité des structures graphémiques : un graphème est souvent représenté par un digramme (deux lettres) ou trigramme (trois lettres).
- La régularité des correspondances graphèmes-phonèmes alors que les correspondances phonèmes-graphèmes le sont moins.
- La structure syllabique dominée par les syllabes ouvertes.

Ainsi, ces caractéristiques font que l'apprentissage de la lecture en français, peuvent à un moment donné, plus précisément lors de la médiation phonologique, amener l'enfant à rencontrer des difficultés dans le traitement de certains mots.

CONCLUSION

L'apprentissage de la lecture fait l'objet depuis plusieurs années d'un grand nombre de recherches qui font émerger différents types de conceptions. De la conception de l'acte lexique, à l'apprentissage de la lecture et aux difficultés rencontrées lors de cet apprentissage, les explications sont divergentes.

Pour certains, l'acte de lire part des processus de haut niveau vers les processus de bas niveau. Pour eux, la compréhension va guider l'identification des mots (Goodman, 1967). A l'inverse, pour d'autres, lire est un traitement de l'information des processus de bas niveau aux processus de haut niveau. L'identification précède donc la compréhension (Gough, 1972). A ces deux conceptions extrêmes de l'acte lexique, s'ajoute une troisième, plus nuancée, qui intègre ces deux types de processus (Rumelhart, 1977). Malgré ces conceptions différentes de l'acte lexique, ces chercheurs se rejoignent sur les mécanismes de base de la lecture que sont l'identification des mots et la compréhension.

Identifier un mot c'est, à partir de ses caractéristiques (phonologiques, sémantiques, syntaxiques, orthographiques...), chercher dans son lexique mental sa représentation lexicale.

Les modèles d'accès au lexique sont là aussi nombreux et on postule que l'accès au lexique peut s'effectuer par la voie grapho-phonétique ou grapho sémantique. La première, de nature non lexicale, est dite indirecte car l'identification du mot repose sur une mise en correspondance des graphèmes et des phonèmes. C'est la procédure par médiation phonologique. La seconde est directe dans le sens où le mot est identifié par un adressage direct avec le lexique mental à partir de ses caractéristiques orthographiques. L'évolution des recherches a progressivement amené à remettre en cause l'indépendance de ces deux voies pour supposer que l'accès au lexique peut se faire tantôt par l'une ou l'autre voie en fonction des caractéristiques du stimulus écrit. Ces modèles à double voies vont alors être à l'origine des premiers modèles de l'apprentissage de la lecture. Dans une dynamique développementale, ces modèles supposent généralement trois étapes qui permettent la mise en place de ces deux voies : logographique (basée sur des indices visuels), alphabétique (médiation phonologique) et orthographique (procédure grapho-sémantique). D'autres modèles, ont été proposés afin de rendre compte de l'interaction des processus et de la façon dont les connaissances sont connectées entre elles. Bien que les modèles développementaux aient été remis en cause à maintes reprises, ils restent cependant les modèles dominants les

plus reconnus de tous. Certes, l'intérêt est qu'ils ont permis de mettre en évidence le rôle essentiel de la procédure par médiation phonologique dans l'apprentissage de la lecture. Mais, ils restent centrés sur l'identification du mot et ignorent totalement les facteurs syntaxico-sémantiques. L'acte de lire se résume, ici, au décodage, la construction de sens n'intervenant que plus tard dans l'apprentissage. De la même façon, selon cette conception, le contexte ne joue aucun rôle dans l'apprentissage de la lecture. D'autre part, le caractère discontinu de ces modèles signifie que tous les enfants passent par les mêmes stades dans un ordre bien défini. Les modèles développementaux ne semblent donc s'intéresser qu'aux processus cognitifs en jeu lors de l'identification des mots. Ainsi, lorsqu'il s'agit de décrire les difficultés rencontrées en lecture, on essaiera d'expliquer que c'est un trouble d'origine biologique, médical et on emploiera le terme de dyslexie pour parler des enfants qui rencontrent des difficultés lors de l'apprentissage de la lecture. L'analyse des erreurs va alors permettre d'effectuer une catégorisation type relative à ceux qui éprouvent des difficultés. L'approche psycholinguistique va sortir de ce cadre descriptif en essayant d'expliquer non pas comment est l'enfant, mais plutôt comment il fait car pour les tenants de cette approche, l'erreur résulterait d'une mauvaise prise d'indices (dans le mot, le texte, la phrase...). Autrement dit, on cherche alors à comprendre pourquoi l'enfant a commis ce type d'erreur. Un autre type d'explication, d'ordre plus sociologique va considérer que ces difficultés peuvent être liées à des facteurs tels que le genre ou le milieu social d'origine. En effet, pour les socio-constructivistes, le contexte social joue un rôle essentiel dans l'apprentissage. Ils stipulent que l'enfant apprend en interaction avec autrui.

QUESTIONS DE RECHERCHE

La lecture fait débat. Régulièrement alimentée par les études menées sur le niveau, la question de l'apprentissage de la lecture ne fait pas l'unanimité dans le champ de la recherche et les débats relatifs à cette habileté sont loin d'être clos.

C'est dans ce contexte qu'il nous a semblé intéressant de faire le point d'une part, sur l'évolution de niveau en lecture et d'autre part sur l'apprentissage de la lecture.

En effet, il n'est pas chose aisée que de mesurer le niveau en lecture puisque celui-ci dépend de la définition que l'on donne du savoir-lire. Et le savoir-lire n'est pas défini de la même façon selon la conception que l'on se fait de de cet apprentissage.

Les nombreuses études menées sur le niveau en lecture le montrent : bien que la tendance semble aller dans le sens d'une baisse du niveau, ce constat est toutefois à prendre avec prudence pour des raisons méthodologiques. En effet, afin que les résultats d'une étude comparative soient fiables, il est nécessaire d'avoir un cadre méthodologique clairement défini pour que l'étude en question puisse être reproduite. Or nous l'avons vu, les études qui tendent à mesurer l'évolution des performances diffèrent aussi bien dans leur mode de recueil et de traitement des données que dans les compétences (compréhension ou décodage) et les objectifs visés (mesurer les acquis en référence aux programmes, capacités à explorer ses connaissances de la langue écrite, ...). Partant de ce constat, avant de nous intéresser plus particulièrement à la question de l'apprentissage de la lecture, nous avons tenté de compléter ce que l'on sait de l'évolution du niveau en lecture. L'utilisation du test de *L'alouette* (Lefavrais, 1965) nous a paru intéressante dans la mesure où il permet une telle comparaison puisque l'étalonnage initial est disponible et les conditions de recueil y sont précisées. Ainsi, au vu des résultats obtenus par les différents types d'évaluations nationales et internationales, nous avons cherché à savoir si à quarante ans à d'écart, il existe des variations intergénérationnelles du niveau en lecture. De cette façon, dans une perspective comparative,

nous nous sommes interrogée sur des éventuelles variations inter-générationnelles. Plus précisément, nous nous sommes demandée si **le niveau en lecture des élèves du CP eu CM2 a augmenté, baissé ou s'il est resté stable dans le temps** et s'il existe des variations dans le temps quant à la dispersion. Au-delà de la question de variations intergénérationnelles, nous nous sommes ensuite demandée s'il existe des variations intra-générationnelles. Autrement dit, nous nous sommes demandée, si le niveau en lecture des élèves en 2009 est influencé par les variables sociologiques classiques telles que le genre et le type d'établissement.

Ces études préliminaires, nous donnent ainsi le moyen de faire un état des lieux du niveau de lecture en France avant d'aborder plus précisément la question de l'apprentissage de la lecture.

Comme nous l'avons mentionné plus haut, les recherches menées sur l'apprentissage de la lecture nous permettent de constater qu'il existe plusieurs façons de concevoir cet apprentissage.

Dans les modèles les plus reconnus, les modèles développementaux, l'apprentissage de la lecture, considéré à travers les processus d'identification du mot, reposerait sur une progression prenant en compte différents aspects de la langue. Cette progression est alors envisagée de manière discontinue puisque l'enfant traverserait une série d'étapes (logographique, alphabétique, orthographique) durant cet apprentissage. Cette conception suppose que l'enfant, lorsqu'il rencontre un mot, aura recours essentiellement à la procédure par médiation phonologique. La construction de sens n'intervenant que plus tard dans l'apprentissage. Le test de *L'alouette*, qui repose sur une conception traditionnelle de la lecture, nous a paru particulièrement représentatif puisqu'il mesure essentiellement le niveau de décodage. Ainsi, **nous avons cherché à savoir dans quelle mesure le test de *L'alouette* nous renseigne sur le déroulement de l'apprentissage de la lecture et plus particulièrement de quelle façon il rend compte d'un apprentissage discontinu (ou continu) de la lecture.** L'utilisation de ce test présente un avantage majeur par rapport aux études menées sur l'identification du mot. En effet, alors que les recherches sur l'apprentissage de la lecture se font la plupart du temps sur des mots isolés, le test de Lefavrais présente des mots en contexte. Cette caractéristique rendrait alors disponible d'autres types d'indices pour identifier les mots, susceptibles de discuter la conception de l'apprentissage des modèles développementaux. Un autre avantage non négligeable, est que le fait d'étudier les mots en contexte nous situe dans des conditions textuelles de lecture, nous rapprochant ainsi des conditions réelles et naturelles de lecture. Par conséquent, sur la base

des études menées sur l'apprentissage de la lecture, **nous avons dans un premier temps essayé de voir si les effets des variables linguistiques classiques observés dans les recherches portant sur l'identification du mot isolé apparaissent en situation de lecture de texte.** Précisons que le test de *L'alouette* n'est pas constitué de façon anodine : l'introduction de faux indices parmi les indices pertinents amènerait les enfants à tomber dans ce que Lefavrais appelle des pièges. Les erreurs commises lors de la lecture seraient alors considérées comme résultant d'une défaillance dans le décodage, selon sa définition de l'acte lexicale et de son apprentissage. Ainsi, voyant ici l'occasion de tester à nouveau les modèles développementaux, **nous avons cherché à savoir quelle est l'influence des variables inhérentes au test de *L'alouette* en lecture et, de façon plus générale, de quelle façon cette influence nous renseigne sur l'apprentissage de la lecture et nous permet de discuter de la continuité ou de la discontinuité de cet apprentissage.**

METHODOLOGIE

Plan de la méthodologie

MODALITES DE RECUEIL DES DONNEES	79
1. Échantillon de départ	80
2. Matériel.....	80
3. Déroulement des passations.....	82
<i>Un temps de mise en condition</i>	<i>82</i>
<i>Un temps de lecture.....</i>	<i>83</i>
MODALITES DE CODAGE.....	84
1. Partie 1 : Le niveau en lecture des élèves du CP au CM2	84
2. Partie 2 : L'apprentissage de la lecture du CP au CM2.....	85
MODALITES D'ANALYSE	86
1. Partie 1 : le niveau en lecture des élèves du CP au CM2	86
<i>Echantillon</i>	<i>86</i>
<i>Indicateur utilisé.....</i>	<i>87</i>
2. Partie 2 : l'apprentissage de la lecture des élèves du CP au CM2	87
<i>Echantillon</i>	<i>87</i>
<i>Indicateurs utilisés</i>	<i>88</i>

MODALITES DE RECUEIL DES DONNEES

1. Échantillon de départ

Notre échantillon se composait au départ de 1403 élèves du CP au CM2¹ scolarisés dans trois types d'établissements (écoles publiques en Education Prioritaire², écoles publiques hors Education Prioritaire et écoles privées)³.

Nous avons ensuite enlevé les élèves qui n'ont pas pu être codés soit parce qu'ils avaient un niveau trop faible (12 ont niveau inférieur à la grande section de maternelle) ou trop fort (5 élèves ont un niveau supérieur), ce qui nous fait un échantillon de 1386 élèves.

Tableau 2 : répartition des élèves selon la classe et le type d'établissement

	CP	CE1	CE2	CM1	CM2	Total
Ecole en Education Prioritaire (EP)	68	65	69	77	70	349
Ecole hors Education Prioritaire (EP)	129	138	132	143	134	676
Ecole Privée	68	69	68	75	81	361
Total	265	272	269	295	285	1386

2. Matériel

Le support utilisé pour cette recherche est le test de *L'alouette* (Lefavrais, 1965). Il se compose d'une planche imprimée comportant un texte illustré de 265 mots (présence d'images autour du texte) et d'une liste de voyelles et syllabes⁴ (cf. ci-dessous).

¹ Les données ont été recueillies progressivement en Master 1, Master 2 et Doctorat par les binômes suivant : Carole Alayrac, Marie-Laure Aubert, Mélanie Decombe, Jessica Escabout, Corinne Félix, Sarah Frayard, Marie Gleizes, Emilie Kronidowski, Amélie Izard, Aurélie Laroussinie, Emilie Léotard, Laetitia Massines, Sabrina Maubourguet, Vanessa Palaysi, Zoé Penot, Julia Rousseau et Marie-Pierre Véron.

² Depuis 2006, l'éducation prioritaire recouvre deux structures d'aide spécifique remplaçant les ZEP et REP: les RAR (Réseaux « Ambition Réussite ») et les RRS (Réseaux de « Réussite Scolaire »).

³ La répartition était la suivante :

- écoles publiques en EP : 356 élèves (soit 26% de l'échantillon)
- écoles publiques hors EP : 685 élèves (soit 49% de l'échantillon)
- écoles privées : 362 élèves (soit 25% de l'échantillon)

⁴ « o u e i a »

« le la les un dans des do ti pu mi »

Cartographie 1 : test de *L'alouette* (Lefavrais, 1965)

L'alouette.







Sous la mousse ou sur le toit,
dans les haies vives ou le chêne fourchu,
le printemps a mis ses nids.
Le printemps a nids au bois.

Annie amie, du renouveau, c'est le doux temps.
 Amie Annie, au bois joli gamine le pinson.
 Dans les buis, gîte une biche, au bois chantant.
 Annie, Annie! au doigt joli, une églantine laisse du sang :
 au bout du temps des féeries viendra l'ennui.




L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
 L'hirondeau piaille sous la pente des bardeaux et, vif et gai, le geai
 sur l'écaille argentée du bouleau, promène un brin d'osier.
 Au verger, dans le soleil matinal, goutte une pompe dégelée.
 On voit un bec luisant qui trille éperdument des notes claires
 et, dans les pampres d'or que suspend la grille antique,
 on surprend des rixes de moineaux.
 Au potager s'alignent les cordeaux; l'if est triste à l'horizon
 et lourd et lent l'envol des corbeaux.








Un lac étire ses calmes rives et, quand le soir descend,
 le miroir de ses eaux reflète les poisons des brignoles perfides.
 Et, quand descend le soir, quand joue la pourpre du couchant,
 le ciel rougit ses eaux.
 Dans la moire de l'eau danse l'ombre d'un écueil.
 Tout est cris! Tout est bruits!





Une amarre est décochée... une barque est arrimée... des matelots
 jettent leurs cassettes sur le rivage...
 Tout est cris! Tout est bruits!
 Au clair de la lune mon ami Pierrot...
 Au clair de lune mon amie annie...
 Au clair de la lune mon ami Pierrot, prête-moi la plume pour écrire un mot.



o u e i a

le la les un dans des do ti pu mi

L'objectif de ce test est de fournir des indications de performance (précision et rapidité) dans une situation de lecture à voix haute. Ce test mesure essentiellement le niveau de décodage plutôt que la compréhension. En effet, la plupart des mots sont relativement rares et probablement inconnus des jeunes sujets. Lecocq⁵ (1991) souligne que « *cela équivaut en fait à contraindre l'enfant à lire des mots qu'il ne connaît pas, et ceci dans une situation naturelle de lecture* ». Ce test, qui permet d'élaborer un âge lexique, est principalement utilisé comme un outil de diagnostic des difficultés en lecture en situant le niveau de lecture d'un enfant par rapport à la médiane des enfants du même âge.

Le matériel comporte une planche illustrée à lire par l'enfant et un texte de protocole pour l'expérimentateur (double du texte sans les illustrations sur lequel est inscrit en bout de chaque ligne le nombre de mots). Ce texte sert à noter les erreurs commises par l'enfant (cf. CD annexes).

Nous utilisons en outre un chronomètre pour mesurer le temps de lecture et un dictaphone pour enregistrer la lecture de l'enfant.

Nous disposons par ailleurs d'une fiche individuelle concernant l'enfant (renseignements sur son âge, sa classe...) et d'un questionnaire rempli par l'enseignant concernant les supports pédagogiques utilisés en classe.

3. Déroulement des passations

Les passations sont réalisées à trois moments de l'année : décembre, février et avril. Cette étude n'est pas longitudinale, les élèves sont différents à chaque passation. Chaque passation se déroule en deux temps :

Un temps de mise en condition

Le chronomètre est mis en place avant que l'élève soit installé, afin de ne pas influencer sa lecture. On fait ensuite asseoir l'enfant à notre gauche afin d'observer son comportement lors de la lecture. Sans évoquer la lecture, nous bavardons d'abord avec lui afin de le mettre à l'aise et ce faisant nous pouvons nous rendre compte s'il présente des troubles de la parole. Nous remplissons à ce même moment avec lui sa fiche personnelle.

⁵ « Apprentissage de la lecture et dyslexie », Liège : Mardaga, 1991.

Un temps de lecture

Le texte illustré est placé devant lui, sans lui donner de consigne. Si, au bout de 10-15 secondes, le sujet ne s'est pas intéressé au texte, on essaie d'attirer son attention en lui montrant du doigt tout le texte et en lui demandant s'il veut bien essayer de le lire. Si l'enfant prend l'initiative de lire le titre, nous le laissons continuer sa lecture, par contre si l'enfant ne dit rien nous le lui lisons. À partir de ce moment, la lecture est commencée et le chronomètre est déclenché. La durée de lecture sera de trois minutes. Dès lors, deux cas peuvent se produire :

Si l'élève ne sait pas lire ou n'arrive pas à lire les deux premiers noms communs du texte (« mousse » et « toit »), nous lui demandons alors de lire la série de voyelles et de syllabes au bas de la page – 5 secondes sont accordées pour chaque déchiffrement. Si l'enfant met plus de temps ou n'arrive pas à les lire, on passe à la suivante. Lorsque la lecture est terminée, on comptabilise le nombre d'erreurs. Si l'élève n'a fait qu'une seule erreur, il est convié à lire le texte de *L'alouette*. Dans le cas contraire, l'épreuve est pour lui terminée.

Si l'élève sait lire les deux mots-clefs (« mousse » et « toit ») ou a déchiffré 14 ou 15 syllabes, il lui est demandé de continuer la lecture du texte.

Sur le texte destiné à l'expérimentateur (double du texte sans les illustrations et sur lequel est inscrit en bout de chaque ligne le nombre de mots), nous entourons toutes les minutes le mot que l'élève vient de lire afin de pouvoir, par la suite, compter le nombre de mots lus en une, deux et trois minutes. Nous retranscrivons instantanément les erreurs telles que l'enfant les a commises (ex. : pour le mot « mousse », il a prononcé « môze »). Les mots mal lus ou sautés isolément, comptent pour une erreur. Si l'enfant saute deux ou plusieurs mots accolés, on compte également une erreur. À la fin des trois minutes, nous notons l'endroit où le sujet s'est arrêté.

Précisons cependant que si le sujet arrive à lire le texte ou semble intéressé par celui-ci, nous le laissons continuer sa lecture jusqu'à la fin du texte, sans l'interrompre au bout des 3 minutes. Cependant, les erreurs commises au-delà des trois minutes ne sont pas prises en compte.

MODALITES DE CODAGE

1. Partie 1 : Le niveau en lecture des élèves du CP au CM2

Afin que la comparaison dans le temps soit la plus fiable possible, nous avons scrupuleusement repris les modalités de codage de Lefavrais (*cf.* CD annexes) :

- **Temps de lecture** codé en secondes. Le temps maximal est donc de 180 secondes.
- **Nombre de mots lus** comprend le nombre de mots biens lus, mal lus, sautés isolément durant les trois minutes. Le nombre maximum de mots lus est de 265.
- **Le nombre de mots sautés** correspond aux mots sautés accolés.
- **Le nombre d'erreurs** comprend les mots sautés isolément, les lignes sautées et les mots mal lus.
- **Le score apparent** correspond au nombre de mots réellement lus, c'est-à-dire le nombre de mots lus auquel on a ôté le nombre de mots sautés.
- **Le score réel** obtenu à partir du tableau de réduction des vitesses de lecture en croisant le score apparent avec le nombre d'erreurs commises.
- **L'âge de lecture** correspondant au score réel dans le tableau des niveaux de lecture

Dans une grille quantitative, les mots lus sont donc codés 0 et les erreurs 1.

Il se peut que certains élèves aient lu le texte en moins de trois minutes et ont sauté des lignes. Dans ce cas, il faut donc calculer soit le nombre de mots qu'ils auraient réellement lus en 3 minutes et le nombre d'erreurs commises (Méthode 1). Soit on calcule le temps qu'ils auraient mis pour lire le texte et le nombre d'erreurs qu'ils auraient commises (Méthode 2).

Exemple d'un élève :

- | | |
|------------------------------|--------------------------------|
| - Temps de lecture = 2mn03 | - Nombre de lignes sautées = 2 |
| - Nombre de mots lus = 265 | - Nombre d'erreurs = 18 |
| - Nombre de mots sautés = 18 | |

Pour ces calculs, il faut d'abord calculer le score apparent, c'est-à-dire le nombre de mots réellement lus. On va donc ôter du nombre de mots lus, les mots sautés. Le score apparent de notre exemple est donc de 265-18 soit 247.

Méthode 1 :

En 2mn03, cet élève a réellement lu 247 mots. En 3 minutes, il aurait donc réellement lu 361 mots $((180*247)/123)$.

En 2mn03, cet élève a commis 18 erreurs. En 3 minutes, il aurait donc commis 26 erreurs $((180*18)/123)$.

On obtient donc deux nouveaux indicateurs : le nombre de mots réellement lus (361) et le nombre d'erreurs (26). Le nombre de mots lus obtenu étant supérieur au nombre de mots lus maximum (265), il n'est donc pas possible de se servir de cet indicateur puisque la table de Lefavrais ne va que jusqu'à 265 mots lus. Il faut alors appliquer la deuxième méthode.

Méthode 2 :

Cet élève a réellement lu 247 mots en 2mn03. S'il avait réellement lu les 265 mots du texte, alors il aurait mis 132 secondes soit 2mn12 $((265*123)/247)$.

S'il avait lu tout le texte en 2mn12 alors il aurait commis 19 erreurs $((132*18)/123)$.

Les deux indicateurs obtenus ici sont le temps de lecture (2mn12) et le nombre d'erreurs (19).

On se reporte ensuite au tableau de réduction des vitesses de lecture (*cf. CD annexes*) pour trouver le score réel en croisant le temps de lecture (2mn12) et le nombre d'erreurs (19). Dans notre exemple, le score réel est donc de 2mn42.

2. Partie 2 : L'apprentissage de la lecture du CP au CM2

Pour les analyses quantitatives, nous avons codé les mots lus (0), les mots erronés (1) et les mots sautés. Contrairement à Lefavrais, nous avons considéré qu'un mot sauté isolément n'est pas un mot lu. Nous les avons donc considérés comme des absences de réponse.

Pour les analyses qualitatives, nous avons retranscrit en phonétique les mots erronés (*cf. CD annexes*).

Modalités d'analyse

1. Partie 1 : le niveau en lecture des élèves du CP au CM2

Echantillon

Pour plus de précision, nous avons souhaité disposer d'un échantillon représentatif de la population scolaire en France. Les données officielles de l'Education Nationale à la rentrée 2008 indiquent que 86,3% des établissements sont des écoles publiques (dont 10,4% en EP) et 13,7% des écoles privées.

Ainsi, en considérant que nos 680 élèves scolarisés en école publique hors EP représentent 75,9% des élèves, nous avons établi un nouvel échantillon qui comprend maintenant 890 élèves répartis de la façon suivante :

Ecoles publiques en EP : 93 élèves (soit 10,4% de l'échantillon)

Ecoles publiques hors EP : 680 élèves (soit 75,9% de l'échantillon)

Ecoles privées : 123 élèves (soit 13,7% de l'échantillon)

Afin de déterminer le nombre d'élèves nécessaire à chaque niveau scolaire de chaque établissement, nous avons procédé de la même manière, c'est-à-dire que nous avons effectué les mêmes calculs pour chaque niveau scolaire de chaque établissement. Ces élèves ont été choisis de manière totalement aléatoire par tirage au sort⁶.

La nouvelle répartition selon le niveau scolaire et l'établissement fréquenté est indiquée dans le tableau 3.

Tableau 3 : nombre d'élèves par type d'établissement selon le niveau scolaire

	CP	CE1	CE2	CM1	CM2	Total
EP	18	19	18	19	18	92
Public hors EP	129	138	132	143	134	676
Privé	23	25	24	26	24	122
Total	170	182	174	188	182	890

⁶ Tirage au sort effectué pour chaque classe de chaque établissement.

Indicateur utilisé

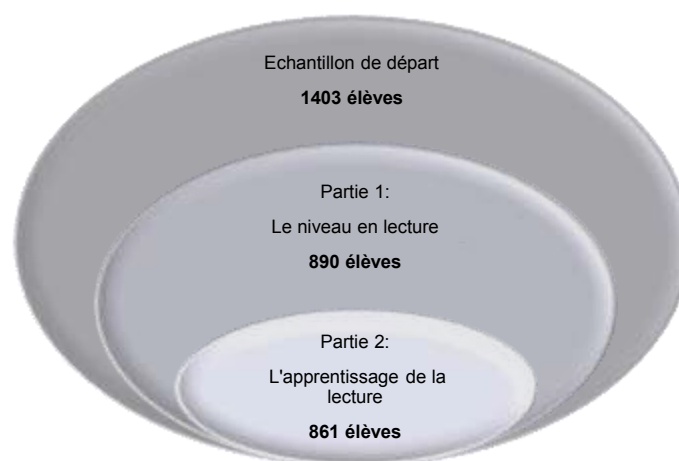
L'indicateur utilisé est l'âge de lecture. Celui-ci est déterminé par un score réel qui correspond au score apparent (nombre de mots lus moins le nombre de mots sautés) croisé avec le nombre d'incorrections. Dans la table de Lefavrais, l'âge lexique correspond à un niveau de lecture qui nous permet de situer le niveau en lecture d'un élève de 2009 par rapport à un élève de 1965 (*cf. CD annexes*).

2. Partie 2 : l'apprentissage de la lecture des élèves du CP au CM2

Echantillon

Cette partie étant destinée à analyser les erreurs de lecture du texte, nous avons par conséquent retiré de notre échantillon les élèves qui n'ont lu que les lettres et les syllabes situées au bas du texte ($N=27$ en CP et $N=2$ en CE1). Afin que l'échantillon reste représentatif de la population scolaire en France⁷, nous avons aussi retiré de façon aléatoire d'autres élèves.

Figure 4 : détail de l'échantillonnage



Ainsi, notre échantillon se compose de 861 élèves répartis de la façon suivante⁸ :

- 323 élèves en cycle 2 → 143 élèves de CP
→ 180 élèves de CE1
- 538 élèves en cycle 3 → 174 élèves de CE2

⁷ Soit 75,9% de l'échantillon en public hors EP, 10,4% de l'échantillon en école publique en EP et 13,7% en école privée en références aux données de l'éducation nationale à la rentrée 2008.

⁸ Cf. CD annexes pour la répartition des élèves selon le type d'établissement et le mois de passation.

→ 188 élèves de CM1

→ 176 élèves de CM2

Indicateurs utilisés

➤ **Le nombre de mots parcourus**

Rappelons que les élèves ont eu un temps limité pour lire le texte (3 mn). De ce fait, ils n'ont pas tous lu la même quantité de mots. Nous avons donc un premier indice de performance qui est le nombre de mots parcourus (*i.e.* les mots lus, qu'ils soient corrects ou erronés, et les mots omis) en trois minutes (*cf. Chapitre 6*).

➤ **La performance lexicale**

Nous avons ensuite fait le choix de quantifier la proportion de mots correctement lus. Nous avons divisé pour chaque élève le nombre de mots correctement lus et le nombre de mots parcourus : le rapport de ces deux indicateurs, nous permet d'obtenir ce que nous appelons « l'indice de performance lexicale » (IPL) (*cf. Chapitre 6*).

➤ **Les erreurs**

Le nombre de mots lus étant différent pour chaque élève, nous avons donc travaillé sur des rapports d'erreurs (*cf. Chapitres 9 et 10*).

➤ **Le nombre d'élèves qui ont commis au moins une erreur**

Pour certaines variables, plusieurs types d'erreurs sont possibles. En travaillant sur des rapports d'erreurs, nous aurions eu des taux d'erreurs supérieurs à 100. Par conséquent, pour certaines analyses (*cf. Chapitres 11 à 14*) nous avons travaillé sur des catégories d'élèves :

- Nombre d'élèves qui ont commis aucune ou au moins une erreur.
- Nombre d'élèves qui ont commis une seule erreur ou plusieurs erreurs.

ANALYSE DES RESULTATS

Partie I

Le niveau en lecture des élèves du CP au CM2

Plan de la partie

Chapitre 3 : Quarante ans plus tard, où en est la lecture ?	91
Chapitre 4 : Evolution de la dispersion à quarante ans de distance	103
Chapitre 5 : Le niveau en lecture en 2009	113
CONCLUSION	123

Chapitre 3



QUARANTE ANS PLUS TARD, OU EN EST LA LECTURE ?

Comparaison des résultats au test de *L'alouette* d'enfants du CP au CM2

Plan du chapitre

INTRODUCTION	92
METHODOLOGIE D'ANALYSE.....	93
<i>Population.....</i>	<i>93</i>
<i>Modalités d'analyse.....</i>	<i>93</i>
<i>Hypothèse générale.....</i>	<i>94</i>
RESULTATS.....	94
I. Le niveau en lecture entre 1965 et 2009	94
1. Résultats globaux.....	94
2. Résultats par classe.....	95
II. Le niveau en lecture entre 1965 et 2009 avec écoles en éducation prioritaire.....	98
1. Résultats globaux.....	98
2. Résultats par classe.....	99
CONCLUSION	100

INTRODUCTION

Depuis plusieurs années, dans notre société, le niveau de maîtrise de la langue et plus particulièrement celui de la lecture inquiète, et l'opinion dominante considère que les élèves étaient meilleurs autrefois qu'aujourd'hui. Déjà, en 1989, les sociologues Baudelot et Establet publiaient un ouvrage avec un sous-titre quelque peu provocateur (« *Le niveau monte : réfutation d'une vieille idée concernant la prétendue décadence de nos écoles* ») allant contre cette idée de baisse du niveau. Aujourd'hui, le débat est toujours présent et la diversité des outils de mesure ne permet pas d'apporter une réponse globale à cette question (cf. *Etat de la question, chapitre I*). Toutefois, la comparaison des résultats obtenus aux différentes évaluations permet de voir qu'à court terme le niveau en lecture semble stable, quelle que soit la compétence mesurée (compréhension, capacité d'utilisation) et la population testée (élèves de 11ans, 15 ans ou plus). Pour voir ce qu'il en est à plus long terme, seules les études comparatives offrent cette possibilité. Cependant, très peu d'études sont reprises¹. Néanmoins, les quelques résultats disponibles indiquent, qu'en 10 ans (1997-2007), le niveau des élèves en lecture et plus précisément en compréhension a légèrement diminué avec une baisse plus marquée pour les élèves les plus faibles.

Ainsi, dans un contexte idéologique où la question de la baisse ou de la hausse du niveau scolaire revient de façon récurrente dans l'espace public² et compte tenu de l'attention portée sur le niveau de maîtrise de la langue, il nous est apparu utile de tenter de voir plus précisément ce qu'il en est dans le cas de la lecture à l'école élémentaire. De cette façon, avec le souci d'un élargissement des comparaisons dans le temps, nous avons mené une étude comparative à quarante ans de distance. Un test de lecture, « *L'alouette* », permet une telle comparaison dans la mesure où l'étalonnage initial du test de Lefavrais (1965) est disponible.

Ayant fait à nouveau passer ce test entre 2006 et 2009, nous avons été particulièrement attentive à ce que le matériel utilisé, les conditions de passation et le codage respectent scrupuleusement les modalités originelles indiquées par Lefavrais, afin que la comparaison soit aussi rigoureuse que possible. De plus, la plupart des études en lecture portant essentiellement sur la compréhension, ce test nous a ainsi permis de voir du côté de la lecture-décodage à l'école élémentaire, si le niveau a augmenté, baissé ou est resté stable.

1 « Étude spécifique sur les élèves en difficulté en lecture à l'entrée en sixième », DEP, 1999.

« Lire, écrire, compter : les performances des élèves de CM2 à vingt ans d'intervalle 1987-2007 », DEPP, 2008

2 Selon un sondage mené par la TNS-SOFRES en 2008, 78% des français jugent que la maîtrise du français est insatisfaisante.

METHODOLOGIE D'ANALYSE

Population

L'étalonnage de Lefavrais ayant été élaboré à partir d'élèves d'écoles publiques, nous avons, par conséquent, gardé seulement ceux relevant de ce secteur³. Ainsi, les élèves scolarisés dans une école en EP (Education Prioritaire) représentent donc maintenant 12% de notre échantillon et ceux en école publique hors EP, 88%.

Tableau 4 : nombre d'élèves par type d'établissement selon le niveau scolaire

	CP	CE1	CE2	CM1	CM2	Total
EP	18	19	18	19	18	92
Public hors EP	129	138	132	143	134	676
Total	147	157	150	162	152	768

Lecture : 768 élèves composent notre échantillon : 92 sont scolarisés en école en Education Prioritaire et 676 sont scolarisés en école publique hors Education Prioritaire.

Modalités d'analyse

Ne disposant que de très peu d'informations sur la façon dont Lefavrais a constitué son échantillon, nous supposons que, conformément aux pratiques courantes en psychométrie, l'échantillon a été constitué en fonction des commodités pratiques de l'auteur, sans attention particulière à quelque facteur sociologique que ce soit, relatif par exemple aux professions et catégories socioprofessionnelles des parents (PCS). Toutefois, à défaut de recueillir les PCS, nous avons tenu compte du type d'établissement fréquenté, variable étroitement liée avec le milieu social⁴. Ainsi, dans un premier temps, nous avons comparé le niveau en lecture des élèves scolarisés en école publique sans tenir compte de ceux qui sont en EP. Puis, dans un second temps, nous avons comparé l'ensemble des élèves.

L'âge de lecture de 1965 sera notre âge théorique et celui de 2009 sera notre âge effectif. Tout écart entre ces deux indicateurs sera considéré comme indiquant une hausse ou une baisse du niveau en lecture entre hier (1965) et aujourd'hui (2009).

³ Les élèves scolarisés en école privée seront utilisés dans le chapitre suivant pour une comparaison inter-établissements.

⁴ Rappelons que, d'après les statistiques nationales (« *Repères et références statistiques* », 2008), les collégiens scolarisés dans des Réseaux Ambition Réussite (RAR) et des Réseaux de Réussite Scolaire (RRS) sont essentiellement d'origine sociale défavorisée. Ainsi, environ 74% des élèves en RAR ont des parents ouvriers ou inactifs, contre 58% en RRS et 35% en établissement classés hors Education Prioritaire (EP).

Nous avons effectué le traitement statistique à l'aide du logiciel SPSS. Les données de 2009 ne suivant pas une loi normale (*cf. CD annexes*) nous avons eu recours aux tests de Mann-Whitney (deux échantillons indépendants) et Kruskal-Wallis (plus de deux échantillons indépendants) pour comparer les tendances centrales des deux distributions. L'étalonnage de Lefavrais ayant été réalisé à partir des enfants médians, nous avons donc pris comme indice de tendance centrale la médiane plutôt que la moyenne.

Hypothèse générale

Au vu des résultats obtenus aux différents types d'évaluation en compréhension, **nous nous attendons alors à observer des variations inter-générationnelles du niveau en lecture-décodage.**

RESULTATS

Afin de voir s'il existe d'éventuelles variations dans le temps du niveau en lecture, nous avons dans un premier temps comparé les indices de tendance centrale entre 1965 et 2009.

I. Le niveau en lecture entre 1965 et 2009

Suivant l'opinion générale selon laquelle le niveau en lecture est en baisse, et compte tenu de la légère diminution du niveau en compréhension observée par les études comparatives, nous nous attendons, ici aussi, à observer une baisse du niveau en lecture-décodage.

1. Résultats globaux

Pour l'ensemble des élèves scolarisés en école publique, il semble qu'à quarante ans de distance, il existe un léger écart de niveau de lecture. Alors qu'en 1965, l'âge de lecture médian était de 8 ans et 7 mois, celui de 2009 est de 8 ans et 1 mois (*cf. tableau 5*). Les élèves d'aujourd'hui présentent donc un retard de six mois par rapport à ceux d'autrefois ($z(676) = -3.025$; $p < .01$). Plus précisément, plus de la moitié (444) d'entre eux ont un niveau de lecture inférieur ou égal à celui de 1965 et moins de la moitié ont un niveau supérieur (232).

Cet écart se réduit d'un mois lorsqu'on considère uniquement les élèves qui ont lu le texte ($z(655) = -3.119$; $p < .01$).

En revanche, pour ceux qui n'ont lu que les « syllabes », leur niveau en lecture est resté stable dans le temps ($z(21) = -1.662$; *NS.*).

Tableau 5 : médianes en mois des âges théoriques et effectifs de l'ensemble des élèves du public hors EP, pour la lecture du texte ou des « syllabes » en 1965 et 2009

	Ages de lecture médians théoriques en 1965 (âges)	Ages de lecture médians effectifs en 2009 (âges)	Test U de Mann-Whitney Z	Seuil de probabilité p.
Ensemble (N=676)	103 (8 ans et 7 mois)	97 (8 ans et 1 mois)	-3.025	<.01
Texte (N=655)	103 (8 ans et 7 mois)	98 (8 ans et 2 mois)	-3.119	<.01
« Syllabes » (N=21)	77 (6 ans et 5 mois)	77 (6 ans et 5 mois)	-1.662	>.05

Lecture : en 1965, les élèves qui ont lu le texte ont un âge médian théorique de 103 mois, ce qui équivaut à un âge de 8 ans et 7 mois.

En bref, s'il existe une légère baisse globale du niveau en lecture, elle n'est pas la même pour tous les élèves. Alors qu'elle est de cinq mois pour les élèves qui ont lu le texte, ceux qui n'ont lu que les « syllabes » ne sont ni en retard ni en avance par rapport aux élèves de 1965. Notre hypothèse selon laquelle le niveau en lecture a diminué est par conséquent validée et rejoint donc l'opinion générale selon laquelle les élèves d'aujourd'hui sont moins bons que ceux d'autrefois.

La question se pose alors de savoir si le retard général ainsi constaté apparaît aux différents niveaux scolaires pris en compte dans notre recherche ou s'il se limite à certains d'entre eux.

2. Résultats par classe

En 1997, une enquête a été menée par la DEPP sur un panel d'élèves à l'entrée en CP. Cinq ans plus tard, ces élèves ont été à nouveau observés à leur entrée en 6^{ème}. Les résultats montrent qu'il existe une corrélation entre les niveaux observés au CP et au début de la 6^{ème} : les élèves qui étaient en difficulté en CP, le sont toujours en 6^{ème}. Autrement dit, les difficultés repérées au début de la scolarité obligatoire ne tendraient pas à se réduire. Partant de ce constat, on peut penser que, s'il existe une baisse du niveau en lecture en CP, alors celle-ci devrait perdurer à tous les niveaux de l'école élémentaire.

Par ailleurs, les résultats des deux études comparatives menées en compréhension indiquent qu'en dix ans le niveau en lecture en fin d'école primaire a légèrement diminué et que le nombre d'élèves en difficulté à l'entrée en 6^{ème} a augmenté. Dès lors, nous supposons que si le niveau en lecture, et plus précisément en compréhension, a baissé en fin d'école primaire, alors nous devrions observer nous aussi une baisse du niveau de décodage en CM2.

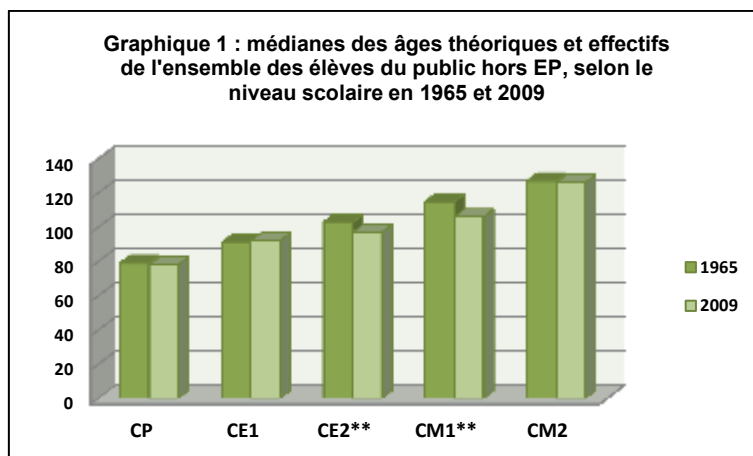
Notre analyse nous permet de constater qu'il existe des différences significatives selon le niveau scolaire considéré. Ainsi, alors que pour les élèves de CP, CE1 et CM2, le niveau en lecture est resté stable depuis 1965, celui des élèves de CE2 et CM1 a diminué (*cf.* graphique 1).

En effet, il apparaît que le niveau en cycle 2 n'a ni augmenté, ni baissé en quarante ans. Que ce soit en CP ($z(129) = -.223$; *NS.*) ou en CE1 ($z(138) = -.932$; *NS.*), les élèves d'aujourd'hui lisent aussi bien que ceux d'autrefois. Il n'y a pas non plus de différence significative entre les âges de lecture médians théoriques et effectifs des élèves de CP qui ont lu le texte ($z(109) = -.006$; *NS.*) ou les « syllabes » ($z(20) = -1.215$; *NS.*).

En revanche, alors qu'il y a quarante ans, les élèves de CE2 avaient un âge de lecture de 103 mois -soit un niveau équivalent à des CE2 au mois de décembre- ceux de 2009 ont un âge de lecture de 97 mois -soit un niveau équivalent à des CE1 au mois d'août-. Les élèves de CE2 d'aujourd'hui sont par conséquent en retard de six mois en lecture ($z(132) = -7.089$; $p < .01$).

Ce retard perdure et s'accroît en CM1 puisque les élèves de 1965 cumulaient un âge de lecture de 115 mois alors que ceux de 2009 ont un âge de lecture de 106 mois. Ainsi, en quarante ans, les élèves de CM1 manifestent un retard de neuf mois en lecture. Ils ont aujourd'hui un niveau équivalent à des CE2 au mois de mars ($z(143) = -3.783$; $p < .01$).

Cet écart constaté disparaît en fin de cycle trois, puisqu'il n'y a aucune différence significative entre les CM2 de 1965 et ceux de 2009 ($z(134) = -1.266$; *NS.*).



**significatif à .01

L'analyse conduite aux différents niveaux scolaires nous permet de voir qu'en dépit d'une baisse générale de six mois du niveau en lecture, celui-ci est resté constant au CP, CE1 et CM2, alors qu'il a diminué de six mois pour les CE2 et de neuf mois pour les CM1.

En ce qui concerne les CM2, l'hypothèse selon laquelle de plus en plus d'élèves quitteraient l'école primaire sans maîtriser la lecture est donc invalidée.

L'échantillon de Lefavrais était constitué d'élèves de l'école publique, y inclus les élèves qui ont par la suite donné lieu à la création des établissements en éducation prioritaire. Or la comparaison que nous venons d'effectuer repose sur le fait que l'éducation prioritaire n'existait pas en 1965. Néanmoins si les structures n'existaient pas, on peut supposer que le public concerné était déjà présent. Ainsi, la baisse de niveau constatée, relativement faible, pourrait s'expliquer par la non-prise en compte de cette population. Notre conclusion est par conséquent discutable. En effet, exclure cette partie de la population conduit à biaiser la comparaison puisqu'on retire précisément les élèves qui ont le plus de difficulté à apprendre à lire⁵.

Pour répondre à cette objection, nous nous proposons donc maintenant de comparer les âges de lecture de l'ensemble des élèves scolarisés en 2009 dans les écoles publiques, qu'elles soient situées en éducation prioritaire ou non, avec ceux obtenus en 1965 par les élèves évalués par Lefavrais.

⁵ A la rentrée 2009, 27,1% des élèves en RAR (« Réseaux Ambition Réussite ») et 21,3% des élèves en RRS (« Réseaux de Réussite Scolaire ») sont en retard à l'entrée en 6^{ème} contre 12,1% des élèves hors Education prioritaire.

II. Le niveau en lecture entre 1965 et 2009 avec écoles en éducation prioritaire

Alors que la comparaison du niveau en lecture-décodage indique un retard de six mois pour les jeunes générations, nous supposons que si on inclut les élèves scolarisés aujourd'hui en EP, la baisse du niveau sera plus importante.

1. Résultats globaux

De façon générale on observe, là aussi, une légère baisse de six mois du niveau en lecture ($z(768) = -4.224$; $p < .01$) (cf. tableau 6). En effet, environ 68% des élèves (521) ont un niveau de lecture inférieur ou égal à celui de 1965 alors que 32% (247) ont un niveau supérieur.

Ce retard est le même lorsqu'on regarde du côté des élèves qui ont lu le texte ($z(741) = -4.335$; $p < .01$).

En revanche, bien que les médianes en 1965 et 2009 soient les mêmes pour les élèves qui n'ont lu que les « syllabes », il apparaît toutefois une différence significative entre ces deux dates ($z(27) = -2.432$; $p < .05$) qui peut être liée à la dispersion des résultats.

Tableau 6 : médianes des âges théoriques et effectifs de l'ensemble des élèves du public (avec écoles en EP), pour la lecture du texte ou des syllabes en 1965 et 2009

	Ages de lecture médians théoriques en 1965	Ages de lecture médians effectifs en 2009	Test U de Mann-Whitney z	Seuil de probabilité p.
Ensemble (N=768)	103 (8 ans et 7 mois)	97 (8 ans et 1 mois)	- 4.224	$p < .01$
Texte (N=741)	103 (8 ans et 7 mois)	97 (8 ans et 7 mois)	- 4.335	$p < .01$
Syllabes (N=27)	77 (6 ans et 5 mois)	77 (6 ans et 5 mois)	- 2.432	$p < .05$

Lecture : en 1965, les élèves scolarisés en écoles publiques (écoles en EP y compris) qui ont lu le texte ont un âge de lecture médian théorique de 103 mois, ce qui équivaut à un âge de 8 ans et 7 mois.

L'hypothèse selon laquelle les élèves de l'éducation prioritaire feraient baisser davantage le niveau en lecture est, par conséquent, rejetée puisque avec ou sans cette population la jeune génération accuse un retard égal de six mois en lecture. Néanmoins, alors que le niveau des élèves les plus faibles – ceux qui ont lu les « syllabes » – scolarisés en école

publique hors EP est resté le même depuis 1965, des variations apparaissent lorsqu'on tient compte de ceux scolarisés en EP.

Reste à savoir si cette configuration se retrouve à chaque niveau scolaire.

2. Résultats par classe

L'analyse des résultats (*cf.* graphique 2) permet de voir qu'il existe des différences significatives en lecture pour les élèves en école publique selon la classe considérée.

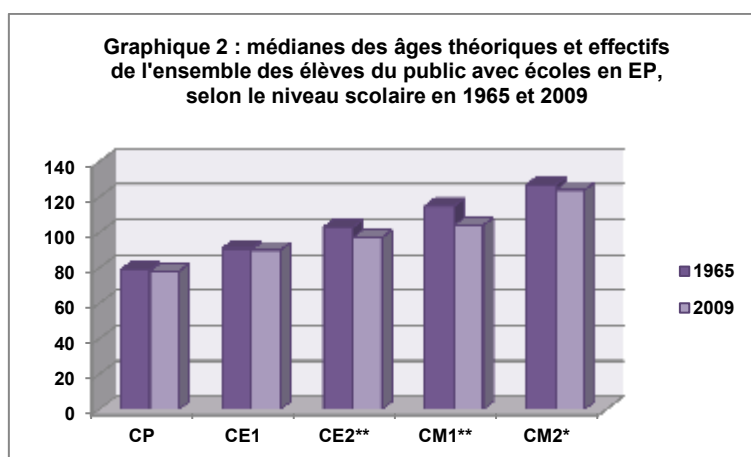
En effet, d'une manière générale, le niveau des élèves de CP est resté stable dans le temps ($z(147) = -0.988$; *NS.*). Alors que cette tendance se retrouve chez les élèves qui ont lu le texte ($z(122) = -0.537$; *NS.*), elle n'apparaît pas chez ceux qui ont lu uniquement les « syllabes ». En effet, comme pour l'ensemble des élèves, on remarque que pour ces derniers, en dépit d'une égalité des âges de lecture médians (77) il existe une différence significative ($z(25) = -2.040$; $p < .05$) entre les deux dates. Là aussi nous supposons une variation aux extrémités des distributions.

Cette stabilité se poursuit en CE1, puisque là aussi, nous ne constatons aucune différence en lecture entre 1965 et 2009 ($z(157) = -0.738$; *NS.*).

En revanche, le CE2 laisse apparaître un léger retard : en quarante ans, le niveau de lecture de ces élèves a diminué de six mois ($z(150) = -7.484$; $p < .01$). En 2009, nos élèves ont un niveau égal à celui des CE1 au mois d'août.

Cette différence se creuse dans la classe suivante puisqu'en CM1, les élèves accusent un retard de onze mois ($z(162) = -4.976$; $p < .01$). Autrement dit, leur niveau en lecture correspond à celui des CE2 au mois de mars en 1965.

Cet écart, qui se retrouve en fin d'école primaire, se réduit fortement ($z(152) = -2.255$; $p < .05$) : les élèves de CM2 ne présentent plus qu'un retard de trois mois en lecture par rapport à ceux de l'échantillon de Lefavrais.



*significatif à .05 ** significatif à .01

CONCLUSION

Notre hypothèse générale consistait à adopter la position dominante selon laquelle le niveau de lecture des enfants est aujourd'hui inférieur à ce qu'il était hier⁶. La comparaison des résultats entre 1965 et 2009 confirme cette hypothèse, mais de manière très modeste : **les élèves de 2009 présentent un retard de six mois seulement par rapport à ceux de 1965.** Cette tendance se retrouve aussi lorsqu'on tient compte uniquement des élèves qui ont lu le texte. Par ailleurs, en ce qui concerne ceux qui n'ont lu que les « syllabes » – les élèves les plus faibles – il apparaît une différence significative malgré une égalité des médianes. La baisse constatée est donc extrêmement faible. Elle n'est pas non plus constante à tous les niveaux scolaires considérés : ainsi, au CP – là où on pouvait s'attendre à la plus forte baisse, si l'on en croit ceux qui mettent volontiers en cause les méthodes d'enseignement de la lecture en ses tout débuts – les résultats montrent qu'il n'existe pas de réelle différence dans les âges lexiques : le niveau en lecture est stable. Cette constance se retrouve aussi chez les élèves de CE1 qui ne sont ni en avance ni en retard. En revanche, l'entrée en cycle 3 fait apparaître une baisse du niveau en lecture. En effet, les élèves de CE2 présentent un retard de six mois par rapport à ceux de 1965. Cet écart s'accroît de trois mois pour les élèves de CM1 puisqu'ils totalisent un retard de neuf mois en lecture. Cette baisse disparaît en CM2. Le retard observé

⁶ A titre d'exemple, selon un sondage effectué en septembre 2008 par l'institut de sondage TNS Sofres, 60% des Français pensent que le niveau scolaire est moins bon qu'il y a vingt ans contre 12% qui pensent qu'il est meilleur et 26% qui estiment qu'il n'y a pas de différence, 2% étant sans opinion. Par ailleurs, 78% jugent que la maîtrise du français n'est pas satisfaisante.

n'a donc pas le caractère progressif et cumulatif auquel on pouvait s'attendre suivant l'opinion d'une baisse générale du niveau commençant au CP et croissant d'année en année.

Plusieurs facteurs liés à l'enseignement de la langue et plus particulièrement à l'apprentissage de la lecture peuvent expliquer cette légère baisse du niveau. D'abord, il est possible que les exigences de la part des enseignants aient diminuées dans le temps. Mais aussi une diminution du nombre d'heures consacrées à l'enseignement de la lecture : primauté de la quantité au lieu de la qualité.

Toutefois, cette baisse du niveau peut aussi s'expliquer par des raisons méthodologiques. En effet, nous avons peu d'informations sur la façon dont Lefavrais a mené son étude. Par exemple, nous ignorons le moment de l'année au cours duquel le recueil de données a été effectué. Par conséquent, il se pourrait que le retard de six mois que nous avons observé ne soit pas en réalité un retard si Lefavrais a effectué ses passations à des moments différents des nôtres.

Notre analyse du niveau de lecture selon le type d'établissement fréquenté fait alors apparaître que, **lorsqu'on inclut les écoles situées en éducation prioritaire, la tendance générale ne varie pas ou très peu.** En effet, même si le niveau baisse d'un mois supplémentaire pour les élèves qui ont lu le texte lorsqu'on rajoute celles situées en EP, l'ensemble de l'échantillon ne présente quand même qu'un retard de six mois en lecture. **En revanche, des différences apparaissent lorsqu'on regarde ce qui se passe à chaque niveau scolaire.** Ainsi, alors que le niveau reste stable en CP et CE1 que ce soit avec ou sans les élèves de l'éducation prioritaire, celui des CM1 et CM2 connaît une baisse plus marquée. Les premiers passent de neuf à onze mois de retard et les seconds, dont le niveau est stable sans les élèves en EP, connaissent une diminution de trois mois lorsqu'on les rajoute. Le retard de six mois constaté pour les élèves de CE2 ne change pas dans les deux cas.

Par ailleurs, la comparaison avec les écoles en éducation prioritaire met en évidence un fait intéressant en ce qui concerne les élèves qui n'ont lu que les « syllabes ». Alors que leur niveau reste stable pour ceux qui sont scolarisés en école publique (sans RAR et RRS), on constate une différence significative lorsqu'on inclut celles situées en EP. Cette différence

non visible avec les médianes laisse supposer des variations internes à l'échantillon. Autrement dit, **la prise en compte des élèves en EP semble entraîner un accroissement de la dispersion.**

Précisons que ce constat peut s'expliquer en partie par une question d'échantillonnage. En effet, nous ne disposons d'aucunes informations quant aux variables sociologiques concernant les élèves de 1965. Toutefois, à défaut de recueillir les PCS, nous avons cependant tenu compte du type d'établissement dans notre échantillonnage, ce qui constitue un moyen grossier de prise en considération des différences sociologiques. Ainsi, ne disposant pas de la proportion d'élèves scolarisés en école élémentaire éducation prioritaire en 2009, nous avons utilisé le nombre d'établissements. Les élèves de l'éducation prioritaire sont par conséquent surreprésentés dans notre échantillon.

Afin de compléter cette étude de la tendance centrale, il nous a paru nécessaire de rajouter un examen de la dispersion des scores. Ainsi, aux informations qu'apporte la comparaison des médianes, nous nous proposons dans le chapitre suivant d'ajouter celles qui résultent de la comparaison de la dispersion des scores. Autrement dit, nous avons cherché à savoir si, entre les deux dates considérées la dispersion est restée constante ou non et de quelle façon les écarts ont évolué.

Chapitre 4



EVOLUTION DE LA DISPERSION A QUARANTE ANS DE DISTANCE

Plan du chapitre

INTRODUCTION	104
METHODOLOGIE.....	105
<i>Population.....</i>	<i>105</i>
<i>Modalités d'analyse.....</i>	<i>105</i>
<i>Hypothèse générale.....</i>	<i>105</i>
RESULTATS.....	106
I. La dispersion des scores.....	106
1. Résultats globaux	106
2. Résultats par classe.....	107
II. Caractéristiques de la tête et de la queue de la distribution.....	108
1. Caractéristiques du quart des meilleurs.....	108
2. Caractéristiques du quart des plus faibles	109
CONCLUSION	109

INTRODUCTION

Au-delà des effets observables pour l'ensemble de la population, nous nous sommes demandé si les différences éventuelles dans le temps sont les mêmes pour toute la population ou si des différences apparaissent quand on distingue les élèves les moins avancés des plus avancés.

En effet, la plupart des études comparatives récemment menées en lecture-compréhension font ressortir un point intéressant : celui d'une baisse plus marquée pour les élèves les plus en difficulté. Rappelons qu'en 2007, la DEPP a constaté qu'en vingt ans, la part des élèves en difficulté a augmenté de 11 points (de 10% en 1987 à 21% en 2007) alors que celle des meilleurs a diminué de 2 points (de 10% en 1987 à 8% en 2007). L'enquête PISA rejoint ce constat puisque les résultats à son évaluation en 2009 montrent que la proportion d'élèves en difficulté a doublé depuis la session qui a lieu en 2000.

Toutefois, ce constat n'est pas nouveau. Déjà en 1989, Baudelot et Establet concluaient que « *L'écart se creuse entre un peloton de tête plus étoffé qu'hier, plus performant que jamais ; et les autres, moins nombreux qu'hier, mais plus démunis que jamais* ». Ils confirment leurs propos avec les enquêtes PISA « *La médiocrité des scores moyens obtenus par la France est en grande partie imputable à l'ampleur des écarts entre la tête et la queue de peloton, et à la part importante d'élèves à la traîne. Autrement dit, l'essentiel des difficultés de l'école en France se situe en bas de la pyramide scolaire* »¹.

Partant de ces constats, et sur la base des résultats qui précèdent et notamment du « poids » que paraissent avoir les élèves de l'éducation prioritaire, on se demande maintenant si en plus d'un changement intergénérationnel il n'existe pas aussi des changements intra-générationnels. Plus précisément, nous nous attendons à observer un accroissement des écarts entre les élèves les moins bons et les meilleurs, qui résulterait d'une baisse plus importante pour les plus faibles.

¹ « L'élitisme républicain » (2009)

METHODOLOGIE

Population

Pour cette étude, nous avons repris le même échantillon que dans le chapitre précédent, c'est-à-dire, les élèves scolarisés en école publique en éducation prioritaire et hors EP.

Tableau 7 : nombre d'élèves par type d'établissement selon le niveau scolaire

	CP	CE1	CE2	CM1	CM2	Total
EP	18	19	18	19	18	92
Public hors EP	129	138	132	143	134	676
Total	147	157	150	162	152	768

Lecture : parmi les 768 élèves scolarisés en écoles publiques, 676 sont en écoles publiques hors EP et 92 en écoles publiques en EP.

Modalités d'analyse

Pour analyser la dispersion, nous avons utilisé les écarts-types, les valeurs extrêmes et un découpage des âges de lecture en quartiles². Nous comparerons alors tour à tour, pour les deux dates considérées, l'écart interquartile (différence entre le 3^{ème} et le 1^{er} quartile), puis la différence entre âges théoriques et âges effectifs pour les âges les plus faibles et pour les âges les plus forts respectivement. On considèrera dans ce dernier cas que l'étendue (différence entre la valeur maximale et minimale) et le sens de ces différences entre les deux dates considérées nous informent sur ce qui se passe pour les lecteurs les plus faibles et pour les lecteurs les plus avancés.

Comme dans les parties précédentes, notre analyse porte en premier lieu sur l'ensemble de l'échantillon, puis sur chaque niveau scolaire

Hypothèse générale

Nous supposons que la différence de niveau en lecture constatée dans le chapitre précédent est le résultat d'une hétérogénéisation des performances. D'une manière plus précise **nous nous attendons à observer un accroissement des écarts entre les élèves les**

² Un quartile représente chacune des 3 valeurs qui divisent les données classées en 4 parties égales, de sorte que chaque partie représente 1/4 de l'échantillon de population. Le 1^{er} quartile sépare les 25 % inférieurs des données ; le 2^e quartile est la médiane de la série et le 3^e quartile sépare les 75 % inférieurs des données. Pour notre analyse, nous avons classé chaque série de données (âges théoriques et âges effectifs) par ordre croissant (du plus faible au plus fort). Ainsi, les trois quartiles nous permettent d'obtenir quatre groupes. Le 1^{er} quartile correspond au 1^{er} élève du dernier quart des moins forts. Le second quartile correspond à l'élève médian. Et le 3^{ème} quartile correspond au dernier élève du premier quart des meilleurs.

plus faibles et les meilleurs résultant d'une baisse plus marquée pour les élèves les plus faibles.

RESULTATS

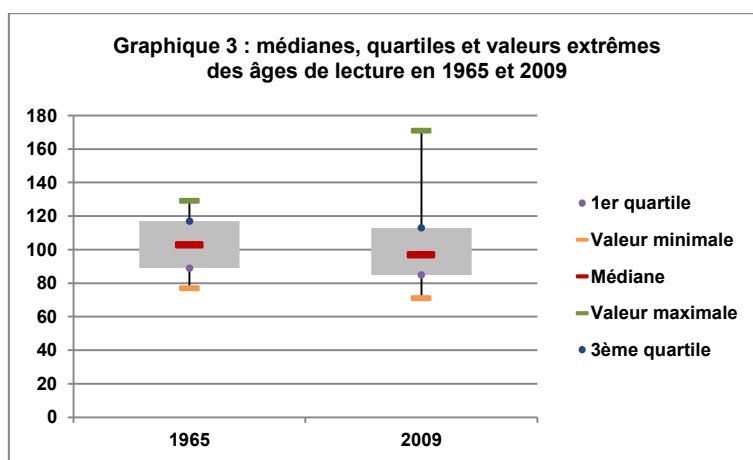
I. La dispersion des scores

1. Résultats globaux

De façon générale, la comparaison des écarts-types fait apparaître que les âges de lecture sont moins homogènes en 2009 qu'en 1965. En effet, l'écart type est passé de 17 à 20 en quarante ans. **La dispersion est donc plus forte aujourd'hui qu'hier.**

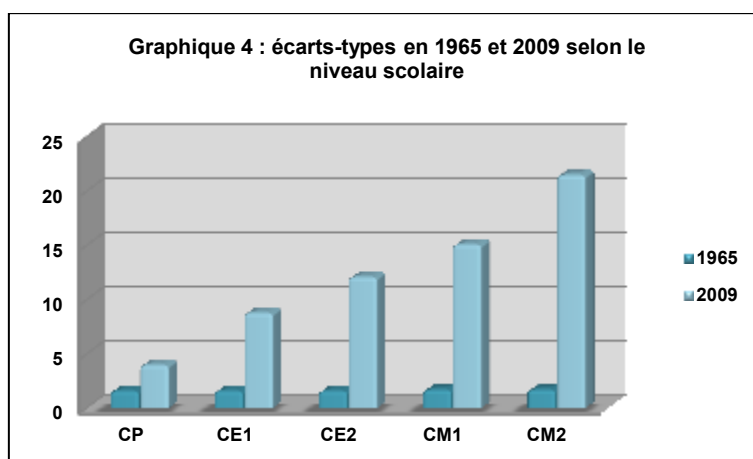
Plus précisément, alors que l'écart interquartile (écart entre le 1^{er} quartile et le 3^{ème} quartile) indique que la dispersion est la même en 2009 qu'en 1965, les deux extrémités de la distribution indiquent des différences (*cf.* graphique 3). En effet, en 1965 la valeur la plus faible était égale à 77, alors qu'elle est de 71 en 2009. D'une manière plus précise, si on regarde le quart des élèves les plus faibles, on constate qu'en moyenne ils ont un âge de lecture de 79 mois alors qu'en 1965 la moyenne était de 83 mois. **Les élèves les plus faibles aujourd'hui sont donc plus faibles que les plus faibles d'hier** (quatre mois d'écart). A l'autre bout de la distribution, on s'aperçoit de même que la valeur maximale est passée de 129 à 171 en quarante ans. Au niveau de lecture le plus élevé, **les élèves d'aujourd'hui apparaissent donc avoir de meilleurs résultats que ceux d'hier.** En effet, il y quarante ans, le quart des plus forts avaient un âge moyen de lecture de 119 mois alors qu'aujourd'hui ces élèves ont en moyenne 127 mois. On peut donc dire que, si à un niveau intermédiaire -celui que mesure l'écart interquartile- les scores apparaissent plus homogènes, les différences s'accroissent aux deux extrémités : les élèves les plus faibles en lecture sont encore plus faibles et les élèves les plus forts sont encore plus forts. Toutefois, contrairement à PISA et à l'étude comparative menée par la DEPP, nous constatons que cet accroissement des écarts ne résulte pas principalement d'une baisse plus marquée pour les élèves les plus faibles mais résulte surtout d'une hausse très prononcée des meilleurs. En effet, on constate une baisse moyenne d'un mois et demi pour les élèves les plus en difficulté contre une hausse de cinq mois pour les plus performants.

Reste à savoir si ceci apparaît à chaque niveau scolaire.



2. Résultats par classe

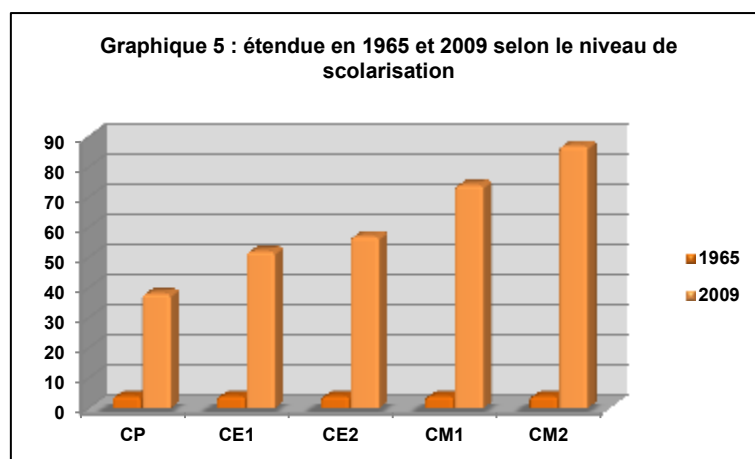
Comme pour l'ensemble de l'échantillon, il s'avère que la dispersion à chaque niveau scolaire est plus forte en 2009 qu'en 1965 (*cf.* graphique 4). Plus particulièrement, on remarque que les âges effectifs de lecture sont de plus en plus hétérogènes avec le niveau d'avancement des élèves. Alors que l'écart-type reste quasiment le même (entre 1,6 et 1,7) entre le CP et le CM2 pour 1965, il est de 4 pour les CP, 9 pour les CE1, 12 pour les CE2, 15 pour les CM1 et 20 pour les CM2 en 2009. La dispersion s'accroît donc d'année en année.



Lecture : en 1965 au CP, l'écart-type était de 1,6 point alors qu'en 2009, il est de 4 points.

Cette tendance semble être due à un accroissement des écarts entre les élèves les plus faibles et les élèves les plus forts. En effet, le graphique 5 nous permet de voir qu'en 1965, l'écart entre les élèves les moins bons et les élèves les plus forts était assez faible (égal à 4) et

restait constant quelle que soit la classe. En 2009, par contre, on constate que l'étendue varie de façon croissante d'une année à l'autre. Elle va de 38 points en CP, à 52 en CE1, puis 57 en CE2, 74 en CM1 et 87 points en CM2. Il semble donc que, en 2009, les écarts se creusent entre les élèves les moins bons et les meilleurs et que cet accroissement aille de pair avec l'avancée dans la scolarité.



Lecture : en 1965 au CP, l'étendue (différence entre la valeur la plus faible et la valeur la plus élevée) était de 4 points alors qu'en 2009, elle est de 38 points.

II. Caractéristiques de la tête et de la queue de la distribution

1. Caractéristiques du quart des meilleurs

En tête de la distribution, on trouve autant de garçons que de filles et ces dernières ne sont pas beaucoup plus en avance. En effet, on compte 104 garçons pour 118 filles. Les filles totalisent un âge de lecture de 127 mois contre 126 mois pour les garçons. Les deux groupes présentent une avance en lecture de huit mois.

Parmi les meilleurs, on compte aussi 14 élèves scolarisés en éducation prioritaire et 208 élèves hors EP. Les premiers sont en avance de deux mois en lecture (avec un âge effectif de 120 mois et un âge théorique de 118 mois) alors que les seconds totalisent une avance de neuf mois en lecture (avec un âge effectif de 127 mois et un âge théorique de 118 mois).

2. Caractéristiques du quart des plus faibles

Au fond de la distribution, on trouve là aussi quasiment autant de garçons (N=109) que de filles (N=113). Bien que les filles présentent une légère avance par rapport aux garçons (avec un âge effectif de 80 mois contre 79 mois pour les garçons), les deux groupes ont un retard de quatre mois en lecture.

Parmi les élèves les plus en difficulté, on trouve 35 élèves scolarisés en éducation prioritaire et 187 scolarisés hors EP. Les premiers, qui ont un âge de lecture moyen de 79 mois, cumulent un retard de sept mois en lecture. Les seconds, dont l'âge de lecture s'élève à 80 mois présentent quant à eux un retard moins important qui est de trois mois.

CONCLUSION

Notre analyse de la dispersion conduit à des conclusions tout aussi intéressantes. Ainsi, d'une manière générale, bien que **la dispersion ne varie guère à un niveau intermédiaire** – ce qu'indique l'écart interquartile – on constate cependant un accroissement de l'écart-type entre 1965 et 2009 (+ 3 points) et ce quel que soit le niveau scolaire. La dispersion est plus forte aujourd'hui qu'hier. Autrement dit, les scores des élèves aujourd'hui sont moins homogènes que ceux d'hier. Au-delà de cette tendance, l'analyse des écarts pour chaque niveau scolaire nous permet de voir que cette hétérogénéité augmente d'une année sur l'autre. Cet accroissement de la dispersion semble s'expliquer par les fortes variations aux extrémités de la distribution. En effet, quand on considère les écarts entre les valeurs non plus intermédiaires mais extrêmes – ce que mesure l'étendue –, un autre fait apparaît, celui de l'accroissement des différences aux deux extrémités de l'échantillon : **les faibles deviennent plus faibles et les forts deviennent plus forts**. Mieux encore : les écarts sont encore plus importants chez les forts ; tout se passe donc comme si le groupe des bons élèves creusait son avance. Il semble donc se produire un double processus aux deux bouts de la population : un léger décrochage des élèves en difficulté et, de manière encore plus marquée, un décollage des élèves les plus avancés.

Lorsqu'on regarde la tête et la queue de la distribution, nous remarquons deux choses. Tout d'abord, il y a pratiquement autant de filles que de garçons des deux côtés et les deux

groupes présentent autant d'avance ou de retard. Dans cette étude, le genre ne semble donc pas être un facteur qui influence les performances en lecture. Ensuite, nous avons constaté que bien que les élèves scolarisés en école en EP soient minoritaires, ils présentent un retard plus important que ceux scolarisés hors EP. Inversement, ces derniers sont davantage en avance. Il est donc possible qu'il existe un effet établissement.

Alors comment peut-on expliquer cet accroissement des écarts ?

Cet accroissement des écarts peut être envisagé sous types d'explications : des explications liées au système scolaire et des explications liées à la société.

En effet, deux facteurs scolaires peuvent expliquer ces résultats : le redoublement et le transfert. Ainsi, la mise en place du redoublement avait au départ pour objectif de faire répéter une année aux élèves les plus en difficultés afin qu'ils puissent acquérir le niveau nécessaire au passage dans la classe suivante. Or, selon les résultats obtenus lors de l'enquête PISA (2009), les pays dont le taux de redoublements est élevé ont de mauvais résultats. En effet, si on regarde les pays pour lesquels plus d'un quart des élèves indiquent avoir déjà redoublé au moins une fois, on constate que les performances des élèves se situent en-deçà de la moyenne des pays de l'OCDE. C'est le cas par exemple de l'Espagne et du Luxembourg dont les scores de ces deux pays lors de l'évaluation de compréhension de l'écrit sont respectivement de 481 et 472, soit 12 points et 21 points au-dessous de la moyenne des pays de l'OCDE (qui ont un score moyen de 493). En ce qui concerne la France, dont le score en compréhension de l'écrit est de 496 (différence non significative avec la moyenne des pays de l'OCDE), 25% des jeunes déclarent avoir redoublé au moins une fois. A l'inverse, si on regarde les pays qui pratiquent très peu, voire quasiment pas le redoublement, alors on constate que ceux-ci se trouvent être parmi ceux qui ont les meilleures performances. C'est le cas par exemple de la Finlande, 3^e au classement PISA 2009, dont 97% des jeunes précisent qu'ils n'ont jamais redoublés. C'est aussi le cas de la Corée qui ne pratique quasiment pas le redoublement. On remarque que ces deux pays se situent en tête du classement, avec des scores respectifs de 539 et 536, soit 46 points et 43 points au-dessus de la moyenne des pays de l'OCDE. La pratique du redoublement, ne permettrait donc pas d'améliorer les performances scolaires.

Ces résultats rejoignent ceux de PIRLS (2006).

D'autres recherches ont permis de mettre en avant le fait que le redoublement influence de façon négative les performances scolaires des élèves. A titre d'exemple, les chercheurs

Holmes & Matthews (1984) ont effectué, dans la lignée de Jackson (1975), une méta-analyse à partir de recherches précédemment menées sur l'efficacité du redoublement. Ils ont pour cela étudié 650 recherches et en ont retenu 44 qui présentaient des critères de validité suffisants. Leur constat est le suivant : les effets du redoublement sont tous négatifs au niveau scolaire.

Ces résultats recourent ceux des chercheurs français Seibel (1984) et Grisay (1993). En effet, en 1984, Seibel a observé l'évolution des résultats de 1100 élèves de CP, à deux épreuves standardisées (mathématiques et français) entre juin et décembre 1983. Bien que cette étude comporte quelques biais, les résultats font toutefois apparaître qu'en juin, les non redoublants moyens sont à peu près au même niveau que les nouveaux redoublants au test de français. En 1993, Grisay, a fait passer deux épreuves standardisées (mathématiques et français) à des élèves à l'entrée en sixième. Deux ans plus tard, ces élèves sont re-testés. Les résultats semblent aller là aussi dans le sens d'un effet négatif du redoublement.

Par conséquent, bien que la proportion d'élèves en retard a diminuée depuis 1960, le redoublement qui concerne encore 20% des élèves du primaire ne permettrait donc pas d'offrir une seconde chance aux élèves les plus en difficulté, ce qui était initialement son objectif.

Une autre explication liée au système scolaire pourrait concerner le transfert. En effet, d'après l'enquête PISA (2009), les pays où le transfert des élèves vers un autre établissement, est plus fréquent font état de performances moins élevées. Par exemple, en Finlande, pays qui se place en tête du classement, 3% des élèves sont dans ce cas alors qu'ils sont plus de 40% au Luxembourg.

Ainsi, d'après l'enquête PISA (2009), les établissements qui cherchent à regrouper les élèves qui ont des difficultés soit par redoublement soit par transfert ne contribuent pas à l'amélioration des performances.

Au-delà de ces explications scolaires, d'autres facteurs d'ordre plus sociologiques peuvent aussi rendre compte de cet accroissement des écarts. Il semblerait que cette évolution puisse s'expliquer avec les changements de la société.

En effet, on observe depuis quelques années une modification de la structure sociale en France : baisse du taux d'activité des personnes peu qualifiées et augmentation des revenus

des personnes les plus aisées. Ainsi, alors qu'entre 1982 et 2010, on constate une baisse de six points du taux d'activité des classes populaires, celui des classes moyennes et supérieures a augmenté respectivement d'un et quatre points. (INSEE). Il en est de même lorsqu'on compare les revenus des ménages : alors que le niveau de vie stagne ou est en baisse en 2010, il est à la hausse pour les plus aisés. Pour reprendre Baudelot et Establet (2009) : « *On retrouve donc au niveau des sociétés ce qu'on observait dans les écoles. Les résultats sont d'autant meilleurs que les inégalités sociales dans le pays sont faibles, et surtout que les écarts entre les moins bien lotis et la moyenne de la population sont resserrés* » (p.83). L'habitus lectural étant plus réduit pour les enfants issus de milieu peu favorisés, cela expliquerait alors cet accroissement des écarts.

D'autre part, on constate aussi que depuis une cinquantaine d'années la proportion d'immigrés en France augmente : elle est passée de 5% en 1946 à 8,4% en 2009 (INSEE).

Ainsi, il est possible que pour les enfants issus de l'immigration l'apprentissage de la lecture soit plus difficile. En effet, outre l'influence du milieu social, celui du milieu culturel l'est tout autant. Ces enfants se retrouvent dans une situation d'interculturalité (culture d'origine vs culture du pays d'accueil) dans laquelle ils vont devoir s'adapter.

L'enquête PISA (2009) révèle que dans les pays de l'OCDE, les élèves issus de l'immigration dits de la première génération, soit ceux qui sont nés à l'étranger de parents nés à l'étranger, accusent un score inférieur de 52 points en moyenne au score des élèves autochtones. Toutefois, elle précise qu'il n'y a pas de corrélation entre la proportion d'élèves issus de l'immigration et l'ampleur des écarts de performances entre eux et les autochtones. Baudelot et Establet (2009) confirment ce constat : « *La proportion d'immigrés ne joue aucun rôle dans le résultat global d'un pays, dans son taux d'échec global, dans la quantité de ses élites et dans l'écart entre les immigrés et les autres* » (p.90).

Il semblerait que dans l'étude que nous venons de mener, le genre ne soit pas une variable discriminante alors que le type d'établissement semble faire varier les résultats. Partant de ce constat, nous avons voulu savoir ce qu'il en est si on tient compte de l'ensemble de notre échantillon, c'est-à-dire en y incluant les élèves des écoles privées.

Chapitre 5



LE NIVEAU EN LECTURE EN 2009

Plan du chapitre

INTRODUCTION	114
METHODOLOGIE.....	114
<i>Population.....</i>	<i>114</i>
<i>Modalités d'analyse.....</i>	<i>115</i>
<i>Hypothèse générale.....</i>	<i>115</i>
RESULTATS.....	116
I. LE NIVEAU EN LECTURE VARIE-T-IL SELON LE GENRE ?	116
1. Résultats globaux	116
2. Résultats par classe.....	116
3. Résultats selon le type d'établissement	118
II. LE NIVEAU EN LECTURE VARIE-T-IL SELON LE TYPE D'ETABLISSEMENT ?	118
1. Résultats globaux	119
2. Résultats par classe.....	119
CONCLUSION	121

INTRODUCTION

En dehors du fait que les différents types d'évaluation permettent de situer les performances des élèves en fonction des objectifs visés ou des compétences mesurées, elles permettent aussi de comparer les performances des élèves en fonction de caractéristiques individuelles mais aussi socio-économiques.

En effet, au-delà du classement par pays qu'apportent les comparaisons internationales, elles permettent aussi de mettre en évidence les points forts et les points faibles de chaque pays. Ainsi, alors que certaines semblent être communes à tous (comme par exemple le genre), d'autres, quant à elles sont davantage propres à la France (caractéristiques sociales). C'est pourquoi, faisant suite au chapitre 3 dans lequel nous avons comparé le niveau en lecture des élèves à quarante ans de distance, nous avons voulu voir si dans notre échantillon, les performances en lecture sont influencées par des facteurs individuels, contextuels et/ou sociaux. L'analyse de la dispersion nous a permis de constater que le niveau en lecture des élèves est moins homogène aujourd'hui qu'autrefois, notamment en raison d'un accroissement des écarts résultant d'un léger décrochage des élèves les plus faibles et d'un fort décollage des élèves les meilleurs. Outre ce constat, nous avons donc cherché à savoir si d'autres facteurs peuvent rendre compte de cette hétérogénéisation.

METHODOLOGIE

Population

Pour cette étude, nous avons repris l'échantillon représentatif de départ (*cf. Méthodologie*), c'est-à-dire celui qui comprend les élèves des écoles privées. Ainsi, nous disposons maintenant de 890 élèves répartis de la façon suivante :

Tableau 8 : nombre d'élèves par type d'établissement selon le niveau scolaire

	CP	CE1	CE2	CM1	CM2	Total
EP	18	19	18	19	18	92
Public hors EP	129	138	132	143	134	676
Privé	23	25	24	26	24	122
Total	170	182	174	188	176	890

Lecture : 170 élèves sont scolarisés en CP. Parmi eux, 18 sont dans une école en éducation prioritaire, 129 sont dans une école publique hors éducation prioritaire et 23 sont dans une école privée.

Modalités d'analyse

Comme dans le chapitre précédent, nous avons utilisé comme indicateur l'âge de lecture obtenu à partir du score apparent déterminé dans la table de Lefavrais. En revanche, plutôt que de comparer les médianes, nous avons préféré ici les moyennes qui sont davantage sensibles aux valeurs extrêmes.

Le traitement statistique s'est effectué, à l'aide du logiciel SPSS. Notre distribution ne suivant pas une loi normale, nous avons par conséquent, eu recours à des tests non paramétriques : le test de Mann-Whitney et le test de Kruskal-Wallis. Pour plus de précision, nous avons ensuite utilisé des tests post-hoc (pour distributions non-gaussiennes) pour comparer deux à deux nos variables et voir ainsi d'éventuelles interactions (*cf. CD annexes*).

Hypothèse générale

Compte tenu que des variations intra-générationnelles existent, notamment entre les élèves les plus faibles et les meilleurs, **nous supposons maintenant que d'autres facteurs relatifs aux caractéristiques individuelles et au contexte scolaire influencent le niveau en lecture des élèves.**

RESULTATS

I. Le niveau en lecture varie-t-il selon le genre ?

Le constat selon lequel la variable « genre » est en lien avec les performances, n'est pas nouveau. Dans le cadre de la lecture, que ce soit en fin d'école primaire ou en fin de scolarité obligatoire, les filles obtiennent de meilleurs résultats que les garçons en compréhension (*cf. Etat de la question, chapitre 1*). Ce constat se retrouve aussi à l'échelle internationale puisque dans tous les pays participant à l'enquête PISA, les performances des filles en compréhension de l'écrit sont plus élevées que celles des garçons : en moyenne, elles devancent les garçons de 39 points, ce qui équivaut à plus d'un demi-niveau de compétence ou à une année d'études.

Partant de ce constat « universel », nous supposons alors que les filles de notre échantillon sont là aussi meilleures que les garçons en lecture-décodage.

1. Résultats globaux

Contre toute attente, l'analyse des résultats fait apparaître qu'en lecture, il n'y a pas de différence significative selon le genre ($z(890) = -.628$; *NS.*). En 2009, les garçons lisent aussi bien que les filles. En effet, d'une manière générale, elles ont un âge de lecture moyen de 101,5 mois et les garçons de 100,6 mois.

Le constat est le même lorsqu'on regarde la partie des élèves qui ont lu le texte. En effet, en dépit du fait qu'il y ait une majorité de filles (51%) qui aient lu le texte, il n'y a pas de différence significative entre les deux groupes ($z(861) = -.140$; *NS.*). Ainsi, les filles ont un âge de lecture de 102,1 mois et les garçons 101,7 mois.

En ce qui concerne les élèves les plus faibles – ceux qui ont lu les syllabes – là aussi, on ne trouve pas de différence significative ($z(29) = -1.903$; *NS.*) entre les performances en lecture des filles (âge moyen de lecture : 77 mois) et des garçons (âge moyen de lecture : 76 mois), bien qu'ils soient majoritaires dans ce groupe (62%).

2. Résultats par classe

Il apparaît donc que, les garçons lisent aussi bien que les filles quel que soit le niveau scolaire considéré (*cf. graphique 6*).

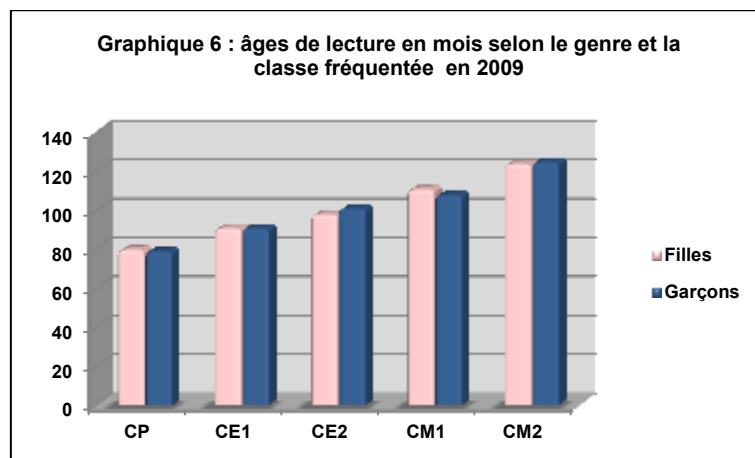
En effet alors qu'en CP les filles ont un âge moyen de lecture de 80 mois, les garçons ont quant à eux en moyenne 79 mois ($z(170) = -.920$; *NS.*). Ce constat est valable aussi bien pour ceux qui ont lu le texte ($z(143) = -.276$; *NS.*) que les syllabes ($z(27) = .115$; *NS.*).

En CE1, les garçons ont un âge moyen de lecture de 91 mois, comme les filles ($z(182) = -.321$; *NS.*). Ici aussi, le constat est le même que les élèves ont lu le texte ($z(180) = .341$; *NS.*) ou les syllabes ($z(2) = .317$; *NS.*)

En CE2, bien que les garçons aient un âge de lecture de 101 mois et les filles de 98 mois, il n'apparaît pas de différence significative selon le genre ($z(174) = .747$; *NS.*).

En CM1, les filles ont un âge moyen de lecture de 111 mois contre 108 pour les garçons. A ce niveau non plus nous ne constatons pas de différence significative ($z(188) = -1.293$; *NS.*).

Enfin, en CM2 les filles ont un âge moyen de lecture de 124 mois et les garçons 125 mois. Le constat est le même que pour les élèves de niveaux précédents : il n'existe pas de différence significative entre ces deux catégories ($z(176) = -.300$; *NS.*).



Par conséquent, **l'âge moyen de lecture des filles n'est pas significativement différent de celui des garçons, quel que soit le niveau scolaire considéré.** Autrement dit, les garçons lisent aussi bien que les filles quelle que soit la classe fréquentée.

Voyons, si nous observons des différences lorsqu'on compare les types d'établissements.

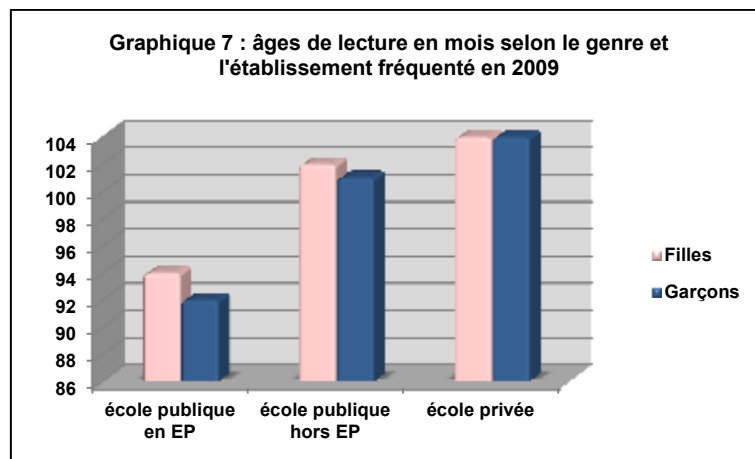
3. Résultats selon le type d'établissement

Il en est de même si on regarde par type d'établissement. Que ce soit en école publique, privée ou publique hors EP, les garçons lisent aussi bien que les filles (*cf.* graphique 7).

En effet, en école publique située en EP, les filles ont un âge moyen de lecture de 94 mois contre 92 pour les garçons ($z(92) = -.479$; *NS.*).

En école publique hors EP, alors que les filles ont un âge de lecture moyen de 102 mois, les garçons ont quant à eux, un âge de lecture moyen de 101 mois ($z(676) = -.520$; *NS.*).

En école privée, les filles et les garçons ont un âge moyen de lecture de 104 mois ($z(122) = .229$; *NS.*).



Ainsi, contrairement à une majorité d'études, il n'y a pas d'effet du genre sur les performances en lecture des élèves de notre échantillon. Notre hypothèse est par conséquent rejetée.

Au-delà du genre, nous avons cherché à savoir si le niveau en lecture de nos élèves est influencé par le type d'établissement fréquenté.

II. Le niveau en lecture varie-t-il selon le type d'établissement ?

Dans le chapitre précédent, nous avons constaté que lorsqu'on inclut les élèves scolarisés en EP, le niveau baisse pour les CM2 et plus fortement pour les élèves de CM1.

Ainsi peut-on également observer un effet établissement ?

Nous supposons que les performances en lecture varient de façon croissante selon que les élèves sont scolarisés en école publique en EP, hors EP et école privée.

1. Résultats globaux

L'analyse statistique nous permet de confirmer notre hypothèse selon laquelle les performances varient d'un type d'établissement à l'autre ($\chi^2(2) = 17.640$; $p < .01$). Les élèves scolarisés en école privée ont en moyenne un âge de lecture de 104 mois, ceux du public hors EP ont 102 et ceux en EP ont 93 mois. Pour plus de précisions, nous avons effectué une comparaison multiple par paires comparant deux à deux les trois types d'établissement. Les résultats font apparaître qu'il existe une différence significative entre les performances en lecture des élèves scolarisés en EP et ceux scolarisés en école publique hors EP ($\chi^2(1) = 3.834$; $p < .01$). Cette différence se retrouve aussi avec les élèves du privé ($\chi^2(1) = 3.927$; $p < .01$). En revanche, il apparaît que les élèves du public hors EP ont un niveau de lecture équivalent à ceux scolarisés dans le privé ($\chi^2(1) = 1.181$; *NS.*). Ce sont donc les élèves scolarisés dans les établissements en EP qui lisent le moins bien.

Voyons maintenant si cette différence apparaît à chaque niveau scolaire.

2. Résultats par classe

Si on regarde dans le détail, l'influence du type d'établissement par classe, on constate que la tendance générale se retrouve à pratiquement tous les niveaux scolaires.

Ainsi, en CP, l'âge de lecture varie avec le type d'établissement fréquenté ($\chi^2(2) = 9.695$; $p < .01$). Alors qu'en EP, les élèves ont en moyenne un âge de lecture de 77 mois, ils ont 79 mois en école publique hors EP et 81 mois en école privée. Comme pour l'ensemble de l'échantillon, on remarque une différence significative entre les âges moyen des CP en école EP et publique ($\chi^2(2) = 2.847$; $p < .05$) et entre les EP et les écoles privées ($\chi^2(2) = 2.896$; $p < .05$). Ce constat est le même pour les élèves de CP qui ont lu le texte ($\chi^2(2) = 7.123$; $p < .05$). En revanche, on ne constate aucune différence significative entre les établissements en ce qui concerne les élèves de CP qui ont lu les syllabes ($\chi^2(2) = .705$; *NS.*).

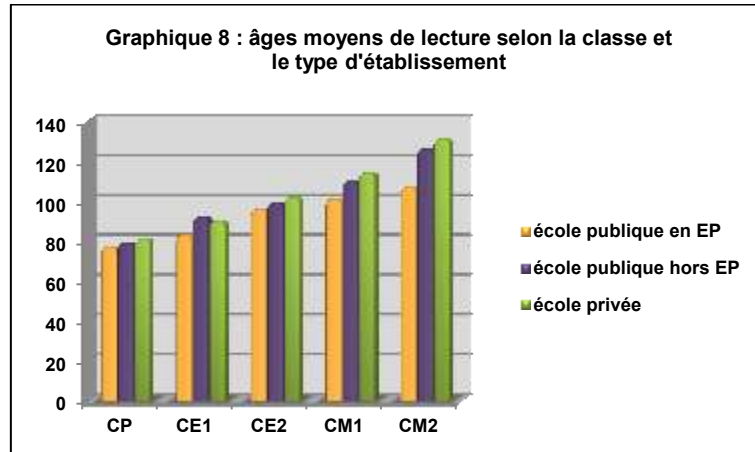
En CE1, la tendance est la même : les élèves en EP ont âge de lecture moyen de 83 mois, ceux en école publique hors EP 92 mois et ceux en école privée 90 mois ($\chi^2(2) =$

23.972 ; $p < .01$). A ce niveau aussi, on constate une différence entre les élèves en établissement EP et privé ($\chi^2 (1) = 2.995 ; p < .01$) et EP et public ($\chi^2 (1) = -4.722 ; p < .01$).

Alors qu'aux niveaux précédents, on observait des variations selon le type d'établissement fréquenté, il apparaît qu'en CE2, les différences ne soient pas significatives ($\chi^2 (2) = 5.497 ; NS.$).

On retrouve en CM1 la tendance générale : alors que les élèves en EP ont un âge moyen de lecture de 101 mois, ceux en école publique hors EP ont 110 mois et ceux en école privée 114 mois ($\chi^2 (2) = 8.590 ; p < .05$). A ce niveau aussi, on retrouve la différence entre les élèves en EP et en école publique hors EP ($\chi^2 (2) = -2.481 ; p < .05$) et entre les élèves en EP et en école privée ($\chi^2 (1) = 2.864 ; p < .05$).

Enfin, en CM2, les élèves en école publique en EP ont un âge moyen de lecture de 107 mois. Ceux scolarisés en école publique ont 126 mois et ceux en école privée 131 mois ($\chi^2 (1) = 17.003 ; p < .01$). Des différences significatives apparaissent encore à ce niveau entre les établissements en EP et public hors EP ($\chi^2 (1) = -3.626 ; p < .01$) et EP et privée ($\chi^2 (1) = 3.957 ; p < .01$).



Il semble donc y avoir un effet partiel du type d'établissement. En effet, le fait qu'il n'y ait pas de différence significative entre les élèves du privé et du public hors EP invalide en partie notre hypothèse et nous amène à envisager une hypothèse davantage relative au milieu social plutôt qu'au type d'établissement fréquenté. En effet, si on regarde de plus près la population qui fréquente ces différents types d'établissements, on constate qu'en éducation prioritaires les élèves sont essentiellement issus d'un milieu social peu favorisés. En effet,

selon les statistiques nationales, à la rentrée 2009, 74,4% des enfants d'ouvriers et d'inactifs étaient scolarisés dans un établissement en RAR et 57,5% en RRS et seulement 8% d'enfants de cadres et d'enseignants sont scolarisés en RAR, 18% en RRS. En revanche, si on regarde le public qui fréquente les écoles hors EP, alors on constate qu'il y a deux fois moins d'enfants d'ouvriers et d'inactifs (35%) et deux fois plus d'enfants de cadres (38%). Ainsi, il semblerait que ce soient les élèves issus d'un milieu social peu favorisé qui rencontrent le plus de difficultés en lecture.

CONCLUSION

L'objectif de ce chapitre était de voir s'il existe des différences dans les performances en lecture selon le genre et le type d'établissement.

Notre première analyse nous permet de constater contre toute attente que le genre n'influence pas les performances en lecture des élèves. En effet, alors que toutes les recherches menées à ce sujet trouvent une différence entre les filles et les garçons, il apparaît ici que ce n'est pas le cas : les garçons lisent aussi bien que les filles et ce, quel que soit le niveau scolaire ou le type d'établissement fréquenté. Avant les années soixante, les filles abandonnaient plus tôt les études à l'issue de la scolarité obligatoire. Mais depuis la généralisation de l'entrée en 6^e, les filles semblent vouloir rattraper leur retard et à partir de 1971, on comptait plus de filles bachelières que de garçons. Aujourd'hui, bien que les filles réalisent de meilleures performances que les garçons, on constate depuis quelques années un resserrement des écarts. Ainsi alors que l'écart concernant les titulaires du bac était de 12 points en faveur des filles en 2000, il est de 10 points en 2011 (DEPP, 2013). Les récents résultats de l'enquête PIRLS (2011) vont là aussi dans ce sens : alors qu'en 2006, l'écart entre les filles et les garçons était de 11 points, il est en 2011 de seulement 4 points.

Par ailleurs, notre seconde étude menée sur l'influence du type d'établissement fréquenté fait apparaître que ce sont les élèves scolarisés en éducation prioritaire qui éprouvent le plus de difficultés de lecture. En réalité, ce n'est pas tant le type d'établissement qui a une influence mais plutôt le milieu social d'origine. En effet, alors que les élèves qui

sont scolarisés dans des établissements en EP sont essentiellement issus de milieu défavorisé, ceux scolarisés hors EP sont essentiellement issus d'un milieu plus favorisé. Ce constat rejoint celui de l'enquête PISA (2009) qui indique que la France est le pays dont le milieu social exerce la plus grande influence sur le niveau scolaire des élèves. En effet, le système scolaire s'est vu évoluer au gré des changements dans la société et malgré la massification de l'école, il semblerait que la réussite scolaire dépende encore du milieu social d'origine. Ainsi, lors de l'enquête PISA, les variations des performances entre les élèves est à imputer à hauteur de 55% à des différences entre les élèves. Dans l'ensemble, les pays dans lesquels les élèves sont issus de milieux socio-économiques favorisés ont de meilleurs résultats et la France semble être l'un des pays où le milieu social influe le plus sur les performances scolaires comme le font remarquer les sociologues Badelot et Establet (2009) : « *Pour la première fois, le caractère élitiste du système scolaire est ainsi objectivé de façon comparative. Les enfants d'origine populaire sont davantage pénalisés à l'école en France que dans d'autres pays* » (p. 63). Ils concluent que « *le pays où la méritocratie républicaine est la plus fortement revendiquée est aussi l'un de ceux où les destins scolaires sont le plus fortement liés aux origines sociales et au capital culturel des familles* » (p. 61).

Au-delà du fait que cette comparaison nous confirme un effet du milieu social sur les performances des élèves, elle nous permet aussi de voir que la politique des zones d'éducation prioritaire mise en place en 1981 destinée à « *corriger cette inégalité [sociale] par le renforcement sélectif de l'action éducative dans les zones et les milieux sociaux où le taux d'échec scolaire est le plus élevé* »¹ (A. Savary, 1981) et dont l'objectif premier était « *d'obtenir une amélioration significative des résultats scolaires des élèves, notamment des plus défavorisés* »² (L. Jospin, 1990) n'a pas eu les effets escomptés, au contraire.

¹ Circulaire n°81-238 du 1^{er} juillet 1981 parue dans le Bulletin officiel n°27.

² Circulaire n°90-028 du 1^{er} février 1990 parue dans le Bulletin officiel n°3.

Partie I

LE NIVEAU EN LECTURE



CONCLUSION-DISCUSSION

L'objectif de ces chapitres était de voir quelle est, depuis quarante ans, l'évolution du niveau en lecture des élèves du CP au CM2 en France.

Alors que dans l'opinion publique, la position dominante consiste à considérer que le niveau en lecture des enfants est aujourd'hui inférieur à ce qu'il était hier, la comparaison objective des résultats entre 1965 et 2009 confirme cette opinion mais de manière très modeste : les élèves de 2009 présentent un retard de six mois par rapport à ceux de 1965. Lorsqu'on regarde plus en détail les résultats, on s'aperçoit que cette baisse n'a pas le caractère progressif ni cumulatif auquel on pouvait s'attendre suivant l'opinion d'une baisse générale du niveau commençant en CP et croissant d'année en année. En effet, le niveau en lecture semble être resté stable en CP, CE1 et CM2 alors qu'on constate une baisse de six mois en CE2 et de neuf mois en CM1.

La prise en compte des élèves en EP, lors d'une seconde comparaison, fait apparaître que d'une façon générale, le niveau ne varie pas. En revanche, le constat n'est pas le même lorsqu'on regarde ce qui se passe à chaque niveau scolaire. Bien que pour les élèves de CP et CE1 la tendance à la stabilité soit la même, on remarque que la baisse observée en CM1 est plus marquée. Elle passe de neuf mois à onze mois de retard. Egalement, alors que sans les élèves en EP, le niveau est resté stable en CM2, on voit apparaître une diminution de trois mois lorsqu'on tient compte de cette catégorie d'élèves. Le retard de six mois observé en CE2 est aussi le même dans cette comparaison.

Cette étude a permis de faire apparaître un fait tout aussi intéressant celui d'un accroissement de la dispersion.

En effet, notre seconde étude concernant l'évolution de la dispersion confirme ce constat. Si à un niveau intermédiaire, les scores des élèves ne varient guère, ce n'est pas le cas pour les scores aux extrémités de la distribution. Il apparaît que les faibles sont plus faibles aujourd'hui et les plus forts encore plus forts. Cet accroissement des écarts pouvant être expliqué par des facteurs scolaires (redoublements, transfert, ...) et extra-scolaires se traduit donc par un double processus aux deux bouts de la population : d'une part un décrochage des élèves les plus en difficulté et d'autre part, un décollage des élèves les plus performants.

Lorsqu'on regarde les caractéristiques de la tête et de la queue de la distribution, on constate que le genre ne serait pas un facteur discriminant contrairement au type d'établissement fréquenté.

C'est ce que nous avons voulu vérifier dans une troisième étude.

Ainsi, au-delà de variations intergénérationnelles, nous avons voulu savoir s'il existe des variations intragénérationnelles.

Contre toute attente, on remarque que les garçons lisent aussi bien que les filles, quel que soit la classe et le type d'établissement fréquenté.

En revanche, alors que le genre n'a pas d'influence sur les performances des filles et des garçons dans notre étude, le type d'établissement est quant à lui plus discriminant. En effet, les élèves scolarisés en EP, sont ceux qui lisent le moins bien en comparaison aux élèves des écoles publiques hors EP et privées. Toutefois, l'absence de différence entre les deux établissements situés hors zone prioritaire, laisse penser que ce n'est pas tant le type d'établissement qui a une influence mais plutôt le milieu social.

Outre le fait que la différenciation ainsi établie confirme l'importance qu'il y a à prendre en considération les différences sociologiques dans ce type d'analyse, le constat des différences selon que l'on inclut ou non l'EP pose la question de la comparabilité de notre échantillon avec celui de Lefavrais et partant de là, de la conclusion à tirer de la comparaison entre 1965 et 2009. On peut se demander, plus précisément, quelle est la plus valide des deux comparaisons effectuées – avec ou sans éducation prioritaire –, et ce choix est capital puisque qu'on observe des différences de niveau et surtout en ce qui concerne la dispersion. Si on considère que le public regroupé aujourd'hui en éducation prioritaire existait en 1965, bien que non identifié comme tel, la comparaison avec la population entière se justifie et on peut

en conclure alors que le niveau a légèrement baissé et que les scores sont plus hétérogènes; si on considère par contre que le public de l'éducation prioritaire n'existait pas en tant que tel en 1965 mais que le fait de l'avoir constitué a pu avoir des effets négatifs sur son niveau de performance, alors seule est valide la comparaison des classes sans EP, dans laquelle le niveau baisse aussi de six mois mais dont les scores sont plus homogènes. On se heurte ici aux difficultés de comparaison dans le temps d'une même épreuve, au fait que, en dépit de toutes les précautions prises aujourd'hui, quand on manque d'informations relatives aux conditions initiales de recueil des données, il est difficile de conclure simplement. On retrouve ici une question abondamment discutée à propos du niveau de l'orthographe par Manesse et Chervel (1989).

La difficulté qu'il y a à opérer des comparaisons temporelles nous incite donc à considérer comme très relatif le retard de six mois observé entre 1965 et 2009. Compte tenu des nombreuses incertitudes relatives à la comparabilité des deux échantillons, la seule conclusion claire qu'il nous paraît possible de tirer de cette comparaison, c'est qu'elle ne fait apparaître entre ces deux dates ni un retard massif, ni un net progrès, mais plutôt une différence faible qui peut alors être tenue pour négligeable et nous invite à penser que le niveau global de lecture est approximativement le même aujourd'hui qu'hier.

Rappelons enfin que si, tout au long de ce texte, nous avons essentiellement parlé de « lecture » à propos de *L'alouette*, il faut entendre ce terme dans une acception limitée. En vérité, le texte de *L'alouette*, dans la mesure où il fait principalement appel au décodage et sollicite très peu la compréhension, n'appréhende que très partiellement la lecture en totalité. Il n'est pas certain en effet que des résultats différents n'auraient pas été observés si le texte utilisé avait fait appel à des mécanismes de lecture de plus haut niveau. Des tests illustrant une conception moins limitée de la lecture existent, mais les résultats obtenus par le passé ne sont plus disponibles.

Ainsi, ce test mesurant principalement le niveau de décodage, il nous a paru intéressant de l'utiliser pour voir ce qu'il en est plus précisément de l'apprentissage de la lecture et ainsi mettre à l'épreuve les modèles d'apprentissage les plus reconnus : les modèles développementaux.

Partie II

L'apprentissage de la lecture des élèves du CP au CM2

Plan de la partie

ETUDES PRELIMINAIRES.....	129
Chapitre 6 : La performance lexicale	130
Chapitre 7 : Types de réponses erronées et processus de lecture	138
Chapitre 8 : Etude des mots qui ont posé le plus de difficultés de lecture	146
L'INFLUENCE DES VARIABLES LINGUISTIQUES.....	165
Chapitre 9 : L'influence de la longueur et de la fréquence d'un mot	166
Chapitre 10 : L'influence de la structure graphémique d'un mot	195
CONCLUSION.....	224
L'INFLUENCE DES VARIABLES LIEES AU TEST DE L'ALOUETTE	228
Chapitre 11 : L'influence des dessins lors de la lecture du texte de <i>L'alouette</i>	235
Chapitre 12 : L'influence des similitudes lors de la lecture du texte de <i>L'alouette</i> ...	259
Chapitre 13 : L'influence du co-texte sémantique lors de la lecture du texte de <i>L'alouette</i>	284
Chapitre 14 : L'influence des mots surdéterminés lors de la lecture du texte de <i>L'alouette</i>	303
CONCLUSION.....	320
Chapitre 15 : Processus de lecture et évolution du CP au CM2.....	322

Depuis plusieurs années, l'apprentissage de la lecture fait l'objet de nombreuses recherches, notamment en langue anglaise et, à l'heure actuelle, plusieurs modèles permettent de rendre compte de la façon dont se déroule cet apprentissage. Bien qu'ils aient été critiqués et remis en question plusieurs fois, les modèles dominants restent les modèles développementaux. Basés sur les modèles d'accès au lexique à double voie, ces modèles envisagent l'apprentissage de la lecture de façon discontinue. En effet, l'enfant traverserait une série de stades, chacun caractérisé par l'adoption d'une procédure, dont l'ordre de succession est strict et identique à tous. Le passage d'un stade à un autre ne pourrait s'effectuer que si les compétences du stade antérieur sont maîtrisées. Par exemple, dans le modèle de Frith (1985), l'enfant passerait d'abord par un stade logographique dans lequel les mots seraient traités comme des objets visuels et non phonologiques. Ensuite, il entrerait dans un stade alphabétique caractérisé par l'apprentissage des correspondances graphèmes-phonèmes. A ce stade, les mots sont traités de manière indirecte, par médiation phonologique. A ce moment de l'apprentissage, l'enfant se centre exclusivement sur le code indépendamment du sens dont la construction est toujours postérieure. Dans le troisième et dernier stade, le stade orthographique, l'enfant traiterait les mots avec un appariement direct avec un mot de son lexique mental par la voie grapho-sémantique.

L'intérêt majeur de ces modèles est qu'ils ont permis de mettre en évidence le rôle essentiel que joue la procédure par médiation phonologique lors de l'apprentissage de la lecture. Le problème reste que ces modèles réduisent le savoir lire à l'identification du mot, autrement dit, savoir lire c'est savoir décoder. La construction de sens est totalement écartée et n'interviendrait que plus tard dans l'apprentissage.

Il nous a donc semblé intéressant de mettre à l'épreuve ces modèles afin de mieux cerner le déroulement de l'apprentissage de la lecture.

QUESTIONNEMENTS

Le test de *L'alouette* nous a semblé particulièrement représentatif dans le sens où, basé sur une conception traditionnelle de la lecture, il mesure essentiellement le niveau de décodage, ce qui correspond à la phase alphabétique du modèle de Frith (1985). Dans ce contexte, les erreurs commises lors de la lecture de ce texte apparaîtraient alors comme une insuffisance dans les capacités de décodage.

On se demande donc **dans quelle mesure les erreurs commises lors de la lecture du texte de *L'alouette* nous informent quant au caractère discontinu ou continu de la lecture.**

Pour répondre à cette interrogation, les analyses qui suivent se déroulent en deux temps : une analyse quantitative et une analyse qualitative.

D'un point de vue quantitatif, nous considérerons qu'il y a continuité lorsque la progression des élèves est régulière, c'est-à-dire que la performance augmente progressivement et/ou les erreurs diminuent progressivement. Toute rupture dans cette progression sera alors envisagée comme un signe de discontinuité.

D'un point de vue qualitatif, nous considérerons qu'il y a continuité lorsque les types d'erreurs sont les mêmes d'un moment à l'autre de l'apprentissage. Si les types d'erreurs sont différents alors nous parlerons de discontinuité.

MODALITES D'ANALYSE

En analysant la façon dont les élèves lisent le texte de *L'alouette*, nous cherchons à savoir : - ***Qu'est-ce que les élèves savent déjà lire ?***

- ***Qu'est-ce qui fait difficulté ?***

- ***Face à ces difficultés quels sont les processus utilisés ?***

Afin d'observer les évolutions, nous détaillons chaque étude par cycle et plus précisément par niveau scolaire (du CP au CM2) pour voir qu'est-ce qui caractérise le début et le cours de l'apprentissage de la lecture.

Ainsi, dans une partie préliminaire, nous étudierons d'abord la performance lexicale des élèves. Puis, avant de nous intéresser plus particulièrement à ce qui fait difficulté aux élèves, nous avons effectué une typologie des erreurs commises et des processus hypothétiques de lecture. Une fois les différents types d'erreurs répertoriés et les processus identifiés, nous avons cherché à savoir quels sont les mots du texte qui ont été les plus difficiles à lire.

Nous étudierons ensuite tour à tour l'influence des variables linguistiques telles que la longueur, la fréquence et la structure graphémique.

Puis, profitant du fait que le test de Lefavrais ne soit pas constitué de façon anodine, nous essaierons de voir quelle est l'influence des variables inhérentes au test de *L'alouette*.

Etudes préliminaires

Plan

Chapitre 6 : La performance lexicale	130
Chapitre 7 : Types de réponses erronées et processus de lecture.....	138
Chapitre 8 : Etude des mots qui ont posé le plus de difficultés de lecture	146

Chapitre 6



LA PERFORMANCE LEXICALE

Plan du chapitre

INTRODUCTION	131
I. La performance lexicale.....	131
II. Quels sont les mots qui sont acquis ?	133
1. Les mots « définitivement » acquis	133
2. Les mots acquis partiellement.....	134
CONCLUSION	137

INTRODUCTION

Alors que la plupart des études menées sur l'apprentissage de la lecture cherchent à voir ce que l'enfant ne sait pas lire, nous avons fait le choix de nous intéresser dans un premier temps à ce que l'élève sait déjà lire en quantifiant la proportion de mots correctement lus. En outre, plusieurs erreurs ayant pu être commises lors de la lecture d'un même mot, la mesure de la performance permet de ne pas augmenter le taux d'erreurs commises par un élève. Néanmoins, même si on contrôle cette variable (performance), tous les enfants n'ont pas lu la même quantité de texte. Ainsi, pour rendre les choses comparables, nous avons divisé pour chaque élève le nombre de mots correctement lus et le nombre de mots parcourus des yeux (y compris les mots sautés) : le rapport de ces deux indicateurs, nous permet d'obtenir ce que nous appelons « l'indice de performance lexicale » (IPL). Nous avons ensuite cherché à savoir de quelle façon, d'un point de vue qualitatif, se traduit cette performance lexicale. Autrement dit, nous avons regardé pour chaque niveau scolaire quels sont les mots qui sont acquis, c'est-à-dire quels sont les mots qui ont été correctement lus par au moins 90% des élèves (*cf. CD annexes*).

I. La performance lexicale

L'analyse de la performance lexicale nous permet de voir qu'il existe une double évolution relative à la maîtrise de la lecture du texte de *L'alouette* : une évolution quantitative et qualitative.

En effet, il apparaît que le nombre moyen de mots parcourus en trois minutes augmente d'un cycle à l'autre et plus précisément, d'année en année (*cf. tableau 9*).

Tableau 9 : indices moyens de performance lexicale en fonction du nombre de mots parcourus et correctement lus et selon le niveau scolaire

	CP (N=143)	CE1 (N= 180)	CE2 (N=174)	CM1 (N=188)	CM2 (N=176)
Nb moyen de mots correctement lus	47	135	179	217	239
Nb moyen de mots parcourus	59	154	197	234	253
Indice moyen de performance lexicale (%)	75%	86%	90%	93%	94%

Lecture : les élèves de CP ont parcouru des yeux 59 mots en moyenne dont 75% (soit 47 en moyenne) ont été correctement lus.

Alors qu'en cycle 2, les élèves ont parcourus en moyenne 112 mots sur 265, ceux du cycle 3 en parcourent plus du double, soit 228 ($z = 19.905 ; p < .01$). Dans le détail, alors qu'en CP, les élèves parcourent, en moyenne, moins d'un quart du texte (soit 59 mots sur 265), les élèves de CE1 en lisent, quant à eux, un peu plus de la moitié (soit 154 mots sur 265). En CE2, nous constatons que le texte est lu en moyenne au trois quart (197 mots sur 265). En CM1 et CM2, les élèves ont parcouru quasiment la totalité du texte : 234 mots pour les premiers et 253 pour les seconds. Ces résultats, montrent donc que d'année en année, la lecture est plus rapide ($\chi^2(4) = 529,916 ; p < .01$).

Cette évolution est la même en ce qui concerne le nombre de mots correctement lus. En effet, alors que sur les 112 mots parcourus en moyenne, les élèves en cycle 2 en ont bien lu 96 (86%), ceux du cycle 3 en ont bien lu 212 sur les 228 parcourus (93%) ($z = 19.827 ; p < .01$). Plus précisément, les élèves de CP ont bien lu en moyenne 75% des mots parcourus (soit en moyenne 47 mots sur 59), ce taux passe à 86% pour les élèves de CE1, 90% pour les CE2, puis 93% pour les CM1 et 94% pour les CM2. Les élèves lisent donc de mieux en mieux d'année en année ($\chi^2(4) = 526,262 ; p < .01$).

Par conséquent, il apparaît que les élèves lisent de plus en plus vite et de mieux en mieux au fur et mesure de l'apprentissage de la lecture, ce qui n'est pas surprenant. En revanche, ce qui est plus surprenant, c'est que nous constatons un effet plafond dès le cycle 3. En effet, alors que la progression de la performance lexicale est de 11 points entre le CP et le CE1, elle est de 4 points en CE2, de 3 points en CM1 et d'un point CM2 ($\chi^2(4) = 323,228 ; p < .01$).

Comment se traduit cette double évolution en terme qualitatif ? Quels sont les mots qui sont acquis?

L'analyse détaillée par mots va nous permettre de répondre à ces questions. Ainsi, faisant suite à l'analyse quantitative de la performance lexicale des élèves, nous allons maintenant étudier de façon plus qualitative celle qui relève des mots.

II. Quels sont les mots qui sont acquis ?

Pour rappel, nous avons considéré comme acquis, les mots sur lesquels 10% ou moins des élèves se sont trompés¹.

1. Les mots « définitivement » acquis

Ainsi, en CP, sur les 59 mots parcourus en moyenne, seulement 6 d'entre eux (soit 10%) semblent être réellement acquis car ils ont été systématiquement bien lus par au moins 90% des élèves : « a, c'est, du, la, sous, une ». On constate qu'au début de l'apprentissage de la lecture ce sont donc essentiellement des mots courts (4 lettres et moins), des mots outils et les auxiliaires « être » et « avoir » conjugués au présent à la 3^{ème} personne du singulier.

En CE1, 35 mots sur les 154 parcourus semblent être acquis (soit 23%). Contrairement aux élèves de CP pour lesquels les mots maîtrisés ne sont, pour la plupart d'entre eux, que des mots outils, on constate qu'en CE1, le répertoire lexical des élèves s'enrichit et intègre maintenant des noms communs fréquents (tels que « amie », « bois », ...), plus longs (tels que « printemps », « mousse », ...) et qui contiennent parfois une lettre muette (« sang », « toit »). On retrouve là aussi des verbes conjugués toujours au présent de l'indicatif à la 3^{ème} personne du singulier (« voit », « mis », « fait »).

En CE2, cette liste s'accroît puisque 64 mots sur 197 (soit 32%) ont été bien lus par au moins 90% des élèves. On voit apparaître des mots au pluriel (« vives », « calmes »), des mots contenant des graphèmes complexes (« chêne », « fourchu », « grille », ...), des mots moins fréquents et plus longs (« renouveau », « l'alouette »), des mots fréquents contenant plusieurs lettres muettes (« doigt », « nids ») et des verbes conjugués à des temps autre que le présent de l'indicatif (« viendra », « chantant »).

En CM1, la tendance continue puisqu'on constate que 90 mots sur 234 (soit 38%) ont été bien lus par au moins 90% des élèves. A ce niveau de l'apprentissage de la lecture, les élèves lisent maintenant quelques mots rares (« gai », « antique ») et des mots à structure graphémique plus complexe (« paille », « nœud »).

¹ Lorsqu'un mot apparaît plusieurs fois, il faut que toutes les occurrences du mot aient posé des difficultés de lecture à moins de 10% des élèves. Pour plus de détails cf. *CD annexes*, tableau « évolution des acquis selon le niveau scolaire ».

Enfin, en CM2, 112 mots sur les 253 (soit 44%) en moyenne parcourus ont été bien lus par 90% des élèves. Des mots à structure syllabique complexe se rajoutent (par exemple « écaille »). On voit aussi apparaître des participes passés à valeur d'adjectif comme « dégelée » ou « argentée ».

Outre cette progression logique et linéaire relative à l'enrichissement du répertoire lexical et à l'amélioration de la lecture des élèves, cette analyse nous permet de constater que certains mots ne sont pas toujours définitivement acquis.

2. Les mots acquis partiellement

En effet, il arrive par exemple que des mots soient bien lus à un certain niveau d'apprentissage mais erronés à d'autres moments. Nous en avons relevé sept.

- « **le** » dans la phrase : « [...] dans les haies vives ou **le** chêne fourchu [...] »

Bien que ce mot ne semble pas présenter de difficulté particulière, on constate que 12% des élèves de CP l'ont mal lu, 17% des élèves de CE1 et de CE2, 7% des élèves de CM1 et 10% des élèves de CM2. Il est donc surprenant de voir que pour ce mot, autant d'élèves commettent une erreur de lecture ; c'est pourquoi nous avons analysé les erreurs commises lors de la lecture de ce mot. Alors qu'en CP, la majorité des élèves qui ont commis une erreur ne le lisent pas, ceux des autres classes prononcent essentiellement « les ». Ainsi, nous supposons que les élèves de CP se sont focalisés sur le mot d'après (« chêne ») et ont ainsi omis de lire ce mot outil. Pour les autres, nous supposons que les élèves aient lu « les » à cause des antécédents « les haies » et se soient ainsi laissés portés par le rythme de la phrase.

- « **du** » dans la phrase : « Annie amie, **du** renouveau, [...] »

En ce qui concerne ce mot, nous constatons que la proportion d'élèves qui l'ont mal lu augmente d'année en année. Ainsi, alors qu'en CP, ils sont 5% à avoir commis une erreur, ils sont 6% en CE1, 10% en CE2, 8% en CM1 et 11% en CM2. Cet article devrait normalement être correctement lu par la totalité des élèves en fin d'école primaire, or ce n'est pas le cas. Alors que du CP au CE2, ce mot est lu essentiellement « au », il est prononcé « de » à partir

du CM1. Il est possible que pour les premiers, les élèves aient parfois confondu deux lignes. En effet, la ligne juste en dessous semble commencer de la même façon « *Amie Annie, au bois joli* » ce qui a pu les amener à lire « au » au lieu de « du ». Pour les seconds, nous supposons que c'est le mot d'après qui les aient amenés à lire « de » : voyant que le mot qui suit est un mot long, ces élèves se sont peut-être là aussi focalisés sur ce dernier et ont donc remplacé la rime de cette syllabe (« du ») par celle de la syllabe d'après (« re »), ce qui les amène à lire « de ». Il est aussi possible que certains élèves se sont trompés à cause d'une anticipation de la locution adverbiale « de nouveau » confondue avec le mot « renouveau ».

- « **les** » dans la phrase : « *Dans **les** buis, gîte une biche, [...]* »

Alors que pour ce mot, seulement 7% des élèves de CP l'ont mal lu, ils sont 15% en CE1, 11% en CE2, 7% en CM1 et 8% en CM2. Les élèves qui ont commis une erreur de lecture ont lu essentiellement « le » quel que soit le niveau d'apprentissage.

Il semble donc que pour ces trois mots qui sont courts et qui ne semblent pas présenter de difficulté particulière, les élèves commettent des erreurs de lecture à cause des mots qui précèdent ou qui suivent. Autrement dit, ils semblent que les élèves aient été influencés soit par la lecture antérieure soit par anticipation des mots qui suivent. Outre ces trois mots outils, d'autres d'apparence plus difficiles, semblent eux aussi avoir été acquis à un certain moment.

- « **surprend** » dans la phrase : « *on **surprend** des rixes [...]* »

Les élèves de CP n'ayant pas été assez loin dans leur lecture, nous ne les avons donc pas considérés ici. Ainsi, 28% des élèves de CE1 n'ont pas bien lu ce verbe. Ils sont 18% en CE2. En CM1, ce mot semble être acquis pour la plupart des élèves puisqu'ils ne sont que 9% à avoir commis une erreur de lecture. Toutefois, cette proportion augmente en CM2 et passe à 11%. Il semble qu'en CE1 et CE2, ce soit la structure du mot, plus particulièrement le groupe consonantique « -rpr- » qui ait posé des difficultés car pour les premiers, ils sont presque la moitié à avoir éliminé le premier « r » et donc lu « suprend ». Pour les seconds, soit ils ne l'ont pas lu, soit, ils l'ont prononcé de la même façon. En revanche, en CM1 et CM2, les erreurs changent et les élèves lisent essentiellement « susprend ». Dans ce cas, nous supposons que le

mot « suspend » (proche graphiquement et phonémiquement) qui se situe quatre mots avant dans le texte, les ait influencés dans la prononciation.

- « **lent** » dans la phrase : « [...] *et lourd et lent l'envol des corbeaux* »

Ce mot étant situé au milieu du texte, nous n'avons pas tenu compte des élèves de CE1 car ils sont trop peu nombreux à avoir été jusque-là. Ainsi, il apparaît qu'en CE2, 9% des élèves qui ont lu ce mot se sont trompés. Ils sont autant en CM2 et légèrement plus nombreux en CM1 (10%). Tous, quel que soit le niveau d'apprentissage de la lecture, ont prononcé essentiellement « le vent ». Il s'emblerait que les élèves aient fait un amalgame avec le mot qui suit (« l'envol »). La répétition de la même syllabe (« len ») a pu les induire en erreur.

- « **étire** » dans la phrase : « *Un lac étire ses calmes rives [...]* »

Alors qu'en CE2, 6% des élèves n'ont pas correctement lu ce verbe, ils sont deux fois plus nombreux en CM1 (15%). En fin d'école primaire, ces élèves représentent 9% de ceux qui ont lu ce mot. Ici aussi, quel que soit le niveau scolaire, ils ont prononcé essentiellement « étiré ». Il se pourrait que les élèves aient substitué le *e* final par la syllabe vocalique « é » rencontrée en début de mot. Il est aussi possible que les élèves aient anticipé sur le mot d'après dont la rime est en « é ».

- « **joue** » dans la phrase : « [...] *quand joue la pourpre du couchant [...]* »

Ce mot semble être acquis en CE2 et CM2 puisqu'en moyenne moins de 10% des élèves (9% exactement) l'ont mal lu. En revanche, en CM1, ils sont 12%. Ce mot a été essentiellement sauté.

Il semble donc que pour ces quatre mots, plus longs et plus difficiles à lire que les précédents, les élèves aient commis des erreurs de lecture.

CONCLUSION

Cette brève analyse nous montre qu'il existe une différence entre ce que les élèves savent lire et ce qu'ils peuvent ; et cette différence semble être liée, certes aux caractéristiques internes aux mots mais aussi au test de *L'alouette*. En effet, comme nous avons pu le constater pour les mots outils, la capacité à identifier des mots ne semble pas être suffisante pour bien lire ce texte. C'est pourquoi, nous avons analysé dans un premier temps l'effet de variables linguistiques telles que la longueur, la fréquence ou encore la structure graphémique. Puis dans un second temps, nous avons cherché à voir quel est l'effet du test de *L'alouette* lors de l'apprentissage de la lecture.

Chapitre 7



TYPES DE REPONSES ERRONEES ET PROCESSUS DE LECTURE

Plan du chapitre

I. Les types de réponses erronées.....	139
1. Réponses erronées sur les unités infra-lexicales.....	139
2. Réponses erronées sur les unités lexicales.....	141
3. Réponses erronées sur les unités supra-lexicales.....	141
4. Réponses erronées au niveau textuel.....	142
II. Les processus de lecture.....	142
1. Le décodage.....	143
2. L'intention lexicale.....	144
3. L'abstention.....	144
4. Processus non identifiables.....	144

L'analyse des erreurs est un passage inévitable lorsqu'on étudie l'apprentissage de la lecture. Depuis plus de 40 ans, Goodman Y. et Goodman K. utilisent l'analyse des méprises afin de mieux appréhender le processus de lecture des enfants et des adultes. Ces chercheurs emploient le terme « méprise » pour parler des erreurs, car selon eux ce dernier est porteur d'une connotation trop négative. Ils pensent que les enfants ne font pas des erreurs par hasard et qu'elles seraient plutôt le reflet des différentes stratégies qui entrent en jeu dans la lecture.

Ainsi, afin d'étudier les réponses erronées des élèves de notre échantillon, nous avons répertorié, dans un premier temps, les erreurs relatives à la langue et, dans un second temps, les hypothèses explicatives des différents processus utilisés par les élèves.

I. Les types de réponses erronées

Les types de réponses erronées font référence à la langue et peuvent être classées selon qu'elles portent sur les unités infra-lexicales, lexicales ou supra-lexicales.

1. Réponses erronées sur les unités infra-lexicales

Les réponses portant sur les unités infra-lexicales concernent les éléments qui composent le mot (lettre, graphème, syllabe) et qui conduisent à la verbalisation d'un mot ou d'un logatome, c'est-à-dire une suite de phonèmes sans signification.

➤ Ajout

Lorsque la verbalisation comporte des lettres supplémentaires différentes de celles qui composent le mot d'origine :

Exemple : « mousse » → [mwɛtə] ou [brɪsə]

➤ Absence

Lorsque la verbalisation ne contient pas tous les éléments du mot d'origine. Cela peut être une absence de :

- lettres : « une » → [œ], « fait » → [fi]
- syllabes : « amarre » → [marə], « printemps » → [prɛ],

➤ Déplacement

Lorsque la verbalisation contient les phonèmes correspondant aux lettres du mot d'origine mais dans un ordre différent :

Exemple : « rivage » → [vɪrɑʒə] ou [gɑrɪvə]

➤ Remplacement

Lorsque certains éléments du mot ont été remplacés par d'autres qui ne présentent pas de ressemblance graphique ou phonologique avec eux. Cela peut être un remplacement de :

- Lettre(s) : « joli » → [ʒonɪ] ou [ʒorɪ]
- Phonème(s) : « goutte » → [gɛtə] ou [gɛ̃tə]
- Syllabe(s) : « promène » → [amɛnə] ou [promiɛrnə]

➤ Réutilisation

Lorsque la verbalisation contient plusieurs fois un élément du mot d'origine. Cela peut être une réutilisation de :

- Lettre(s) : « amie » → [mami] ou « potager » → [pɔtɑʒe]
- Phonème(s) : « descend » → [dɑ̃sɑ̃] ou « fourchu » → [fʊɾfʊ]

➤ Erreur de prononciation

Lorsque certains éléments de la réponse ne tiennent pas compte de certaines règles de prononciation. Cela peut concerner les :

- lettres muettes : « dans » → [dɑ̃s] ou « doigt » → [dʁwɑʒ]
- lettres dont la prononciation dépend du contexte orthographique : « poisons » → [pwasɔ̃] ou « rivage » → [rɪvɑgə]
- la flexion de verbes au pluriel : « s'alignent » → [salɪnɑ̃]

➤ Amalgame

Lorsqu'un élément de la réponse présente une similitude graphique et/ou phonémique avec un élément de départ. Cela peut être un amalgame au niveau des :

- Lettres : « bois » → [dwa] ou « vif » → [viv]
- Phonème(s) : « au » → [u] ou « chantant » → [fɑ̃tɑ̃]

➤ Réponse erronée portant sur un graphème composé

Lorsque la verbalisation comporte des erreurs portant sur un graphème composé :

- Le graphème est décomposé en deux éléments : « paille » → [pailə]
- Un élément du graphème composé est rapporté à un autre graphème : « paille » → [pɛl]

2. Réponses erronées sur les unités lexicales

Les réponses portant sur les unités lexicales se situent au niveau des mots.

➤ Ajout(s) de mot(s)

Lorsque la verbalisation s'accompagne d'un mot supplémentaire.

Exemple : « au clair de lune » → [o klɛR də la lynə]

➤ Omission(s) de mot(s)

Lorsqu'il manque un mot à la verbalisation :

Exemple : « au clair de la lune » → [o klɛR də lynə]

➤ Déplacements(s) de mot(s)

Lorsque deux mots sont verbalisés dans le désordre :

Exemple : « au bois joli » → [o ʒoli bwa]

➤ Remplacements(s) de mot(s)

Lorsque la réponse est un mot qui ne présente aucune ressemblance graphique et/ou phonémique avec le mot de départ :

Exemple : « mon amie annie » → [mɔ̃ ami PjɛRO]

3. Réponses erronées sur les unités supra-lexicales

Les réponses portant sur les unités supra-lexicales se situent au-delà des mots.

➤ Omission de lignes

Lorsqu'une ligne n'a pas été verbalisée.

➤ Omission de phrases

Lorsqu'une phrase n'a pas été verbalisée

➤ Erreur de liaison

Lorsque deux mots sont liés de façon abusive : « les haies » → [leze]

Nous pouvons ajouter un quatrième niveau, le niveau textuel. Nous précisons que nous n'avons pas traité ce niveau dans notre recherche.

4. Réponses erronées au niveau textuel

➤ La lecture hachée

➤ Le non-respect de la ponctuation

Nous précisons que ces types d'erreurs ne sont pas exclusifs et qu'un même mot peut en comporter plusieurs. Ainsi, pour le mot « nids », un élève a lu [ɛ̃b]. On observe dans sa verbalisation un déplacement des lettres « i » et « n », suivi d'une prononciation d'une lettre muette, le « d », elle-même incorrecte puisque lue [b]. La lecture de cet élève comporte donc trois erreurs.

Tous ces types d'erreurs relèvent donc de la langue et se retrouvent dans les processus de lecture des élèves. C'est ce que nous allons étudier maintenant.

II. Les processus de lecture

Alors que les types d'erreurs font référence à la langue, les comportements, quant à eux sont relatifs à l'élève. Nous distinguerons trois comportements de lecture : le **décodage** conduisant à la prononciation d'un logatome (c'est-à-dire un non-mot), l'**intention lexicale** dont le résultat est la prononciation d'un mot et l'**abstention**. Dans certains cas, la verbalisation est un mot mais on ne peut pas savoir si cette réponse est le résultat d'une erreur

de décodage ou d'une intention lexicale, ce qui nous fait une quatrième catégorie : les **processus non identifiables** (cf. *CD annexes*).

1. Le décodage

L'intention de l'élève c'est de déchiffrer. En conséquence, sa verbalisation est un logatome dû à :

➤ une mise en **correspondance graphèmes-phonèmes erronée** résultant d'une erreur:

- de prononciation liée à des similitudes graphiques et/ou phonémiques

Exemple: « fourchu » → [vURʃy]

- due à une méconnaissance de certaines règles de prononciation

Exemple : « gîte » → [git]

- de prononciation d'une lettre muette

Exemple : « nid » → [nidə]

➤ **décodage incomplet** : « printemps » → [prɛ̃tə]

➤ un **mauvais découpage du mot** entraînant une lecture :

- littérale : l'élève ne fusionne pas les graphèmes composés, ce qui le conduit à lire une à une les lettres qui composent le mot.

Exemple : « bois » → [bois], paille → [pailə]

- séquentielle : l'élève décode le mot deux éléments par deux éléments.

Exemple : « geai » → [ʒəe].

La plupart du temps, ce décodage séquentiel s'effectue sous forme de régularisation d'une ou plusieurs syllabes selon le type CV¹.

Exemple : « printemps » → [puritəmɛps], « paille » → [palə].

¹ Consonne- Voyelle

2. L'intention lexicale

L'intention de l'élève est de donner du sens. Le résultat de la verbalisation est donc un mot. Cela se traduit par un :

- ajout de lettres différentes de celles qui composent le mot d'origine
Exemple : « sous » → [sosisə]
- ajout de mots
Exemple : « et quand descend » → [e kã ã desã]
- remplacement de lettre, de phonème, de syllabe ou de mots
Exemples : « les » → [me], « mis » → [mẽ], « cordeaux » → [kado], « paille » → [Rɛlɛvə]

3. L'abstention

L'intention de l'élève est de ne pas se tromper. De ce fait, la non-lecture est un moyen pour éviter de fournir une réponse erronée.

L'abstention peut porter sur un mot, une phrase ou une ligne.

4. Processus non identifiables

Ici, la verbalisation est un mot mais nous ne pouvons dire si l'élève a eu une intention lexicale ou si cette erreur est le résultat d'un processus de décodage. On retrouve les erreurs concernant les :

- amalgames visuels: « bois » → [dwa] : intention lexicale ou amalgame entre les lettres b/d ?
- amalgames de sons : « vif » → [viv] : intention lexicale ou amalgame entre les sons [f]/[v] ?
- déplacements : « rivage » → [viraʒə] : intention lexicale ou déplacement des lettres ?
- déplacements accompagnés d'omissions : « printemps » → [prɛ̃sə] : intention lexicale ?
- Prononciation de graphèmes contextuels: « poisons » → [pwasõ] : intention lexicale ou erreur de prononciation ?
- Prononciation de lettres muettes : « dans » → [dãs] : intention lexicale ou erreur de prononciation ?
- Prononciation de la flexion de certains verbes au pluriel : « s'alignent » → [salinã] : intention lexicale ou erreur de prononciation ?

En **résumé**, lors de l'apprentissage de la lecture, l'élève va devoir faire face à un certain nombre de difficultés liées à la langue qui vont l'amener à fournir des réponses erronées. Ces erreurs peuvent être commises à quatre niveaux : infra-lexical (à l'intérieur du mot), lexical (au niveau du mot), supra-lexical (au-delà du mot) et textuel (au niveau du texte). Ces erreurs aboutissent soit à des verbalisations de non-mots soit de mots. Partant de ce constat, nous avons formulé quatre hypothèses explicatives concernant les processus utilisés en lecture par l'élève. Ainsi, dans le premier cas, nous supposons que l'élève a eu essentiellement recours à la voie grapho-phonétique pour traiter les mots et que son intention principale est de déchiffrer, ce qui expliquerait que la réponse fournie est un non-mot. Dans le second cas, nous pensons que l'élève a eu recours à la voie grapho-sémantique pour traiter les mots et que son intention est de donner du sens, ceci expliquant que les réponses fournies sont des mots. Par ailleurs, il se peut que l'élève n'ait rien prononcé. Par conséquent, nous avons émis l'hypothèse que l'élève s'est abstenu de lire pour éviter de fournir une réponse erronée, son intention est donc de ne pas se tromper. Enfin, pour certaines erreurs, les processus qui en sont à l'origine sont difficilement repérables, nous les avons donc classés dans la catégorie des processus non identifiables.

Faisant suite à cette description des erreurs et des hypothèses explicatives des différents processus de lecture, nous avons voulu voir quels sont, dans le texte de *L'alouette*, les mots qui sont les plus difficiles à lire? Et pourquoi? Et quel est le comportement de l'élève face à ces mots?

Chapitre 8



ETUDE DES MOTS QUI ONT POSE LE PLUS DE DIFFICULTES DE LECTURE

Plan du chapitre

METHODOLOGIE D'ANALYSE	147
I. Ensemble des élèves (N=861)	147
II. Les élèves de CP (N=143)	148
III. Les élèves de CE1 (N=180)	151
IV. Les élèves de CE2 (N=174)	156
V. Les élèves de CM1 (N=188)	159
VI. Les élèves de CM2 (N=176)	162
CONCLUSION	163

Dans une perspective exploratoire, l'objectif de ce chapitre est de voir quels sont les mots qui sont difficiles à lire, c'est-à-dire ceux qui entraînent le plus d'erreurs. Une fois, ces différents mots répertoriés, nous avons cherché à savoir quel est le comportement des élèves face à des mots qui présentent des difficultés de lecture : *essaient-ils de produire du sens ou simplement de déchiffrer ?* Autrement dit, à partir des mots qui ont entraîné le plus d'erreurs de lecture, nous avons cherché à savoir *quel est le comportement adopté par les élèves.*

METHODOLOGIE D'ANALYSE


Les élèves n'ayant pas tous lu le même nombre de mots (*cf. Méthodologie*), nous avons dû effectuer un tri pour ne garder que les erreurs les plus significatives. De cette façon, nous avons dans un premier temps sélectionné les mots qui ont été lus par au moins la moitié des élèves. Dans un second temps, nous avons gardé parmi ces mots ceux sur lesquels au moins la moitié des élèves se sont trompés. Cette analyse a été réalisée d'une part pour l'ensemble de notre échantillon puis, d'autre part, pour chaque niveau scolaire (*cf. CD annexes*).

I. Ensemble des élèves (N=861)

De façon générale, quatre mots ont entraîné des difficultés de lecture pour au moins 50% des élèves (*cf. tableau 10, dernière colonne*). Il s'agit des mots « *féeries* », « *piaille* », « *geai* » et « *poisons* ».

Tableau 10 : mots qui ont entraîné le plus de difficultés de lecture et pourcentage d'élèves qui ont commis une erreur sur ces mots (par rapport au nombre d'élèves qui ont lu ce mot)

Mots du texte	CP (N=143)	CE1 (N=180)	CE2 (N=174)	CM1 (N=188)	CM2 (N=176)	Ensemble (N=861)
« <i>haies</i> »	70 (140)	56 (180)	50 (174)	36 (188)	27 (176)	47 (858)
« <i>printemps</i> »	56 (140)	8 (180)	1 (174)	0 (188)	0 (176)	11 (858)
« <i>printemps</i> »	50 (127)	7 (180)	0 (174)	0 (188)	0 (176)	9 (845)
« <i>féeries</i> »	84 (45)	75 (169)	60 (174)	49 (188)	38 (176)	57 (752)
« <i>l'ennui</i> »	62 (45)	64 (168)	36 (174)	26 (188)	15 (176)	36 (751)
« <i>piaille</i> »	70 (32)	65 (161)	61 (172)	46 (187)	36 (176)	52 (723)
« <i>geai</i> »	65 (23)	66 (155)	64 (169)	56 (187)	41 (176)	57 (710)
« <i>d'osier</i> »	61 (18)	62 (151)	40 (164)	22 (186)	14 (176)	34 (695)
« <i>poisons</i> »	67 (3)	89 (61)	89 (107)	86 (165)	84 (169)	86 (505)
« <i>au clair de lune</i> »	100 (1)	88 (17)	88 (48)	91 (105)	91 (142)	90 (313)
« <i>la</i> »	100 (1)	38 (13)	49 (41)	54 (100)	48 (138)	50 (293)

 Mots qui ont été lus par moins de la moitié des élèves

Lecture : en CP, sur les 143 élèves, 140 ont lu le mot « *haies* ». Parmi eux, 70%, ont commis une erreur de lecture.

En effet, 86% des élèves qui ont lu le mot « *poisons* » ont commis une erreur, comme 57% des élèves qui ont lu les mots « *féeries* » et « *geai* » et 52% qui ont lu le mot « *piaille* ».

Toutefois, si on regarde dans le détail chaque niveau scolaire, on se rend compte que les mots qui ont entraîné le plus de difficulté de lecture ne sont pas les mêmes d'une classe à l'autre. En effet, on remarque plus particulièrement que la lecture de certains mots a été difficile seulement pour les lecteurs débutants alors que d'autres concernent plutôt les lecteurs intermédiaires ou plus avancés.

C'est pourquoi, nous nous proposons de regarder pour chaque classe, quels sont les mots qui ont entraîné le plus de difficultés de lecture.

II. Les élèves de CP (N=143)

En CP, trois mots semblent avoir posé des difficultés majeures de lecture (cf. tableau 11). Il s'agit des mots « *haies* » et « *printemps* » qui apparaît à deux reprises dans le texte.

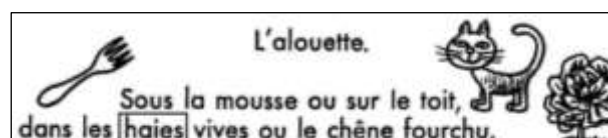
Tableau 11 : mots sur lesquels au moins la moitié des élèves de CP ont commis une erreur de lecture

N° mot	Mot	Nb élèves qui l'ont lu	Nb élèves qui ont commis une erreur	% élèves qui ont commis une erreur
10	haies	140	98	70
17	printemps	140	79	56
23	printemps	127	64	50

Lecture : en CP, 140 élèves sur 143 ont lu le mot « haies ». Parmi eux, 98, soit 70%, ont commis une erreur de lecture.

En effet, 70% des élèves de CP qui ont lu le mot « haies » ont commis une erreur de lecture. En ce qui concerne le mot « printemps », on remarque que lors de sa première apparition ce sont 56% des élèves qui l'ont mal lu et 50% à la seconde apparition.

- « *haies* »



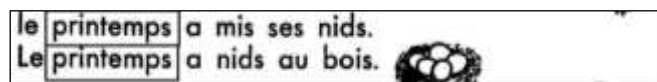
Face au mot « *haies* », **50 % des élèves de CP ne l'ont pas lu** (soit 49 sur 98). Ces derniers ont peut-être préféré ne rien prononcer au risque de se tromper.

Parmi les élèves qui ont mal lu ce mot, **29% ont commis une erreur de décodage** (soit 28 sur 98). La plupart ont associé le « h » aspiré au son [ʃ] et ont donc lu [ʃe], [ʃɛ], [ʃɛvə], [ʃɥi], ...etc (32%). Quelques-uns, ont commis une erreur de liaison avec l'article qui précède et ont donc lu [leze] (25%). Il se pourrait que ces élèves aient abordé le cas de la liaison mais ne connaissent pas les exceptions telles que l'absence de liaison en présence d'un « h » aspiré. Enfin, parmi les erreurs de décodage, on observe aussi une lecture littérale, parfois incomplète, amenant les élèves à prononcer [a], [aə], [aiə] ou encore [əe] (21%). Il semblerait que pour ces élèves, la difficulté réside dans le traitement du groupe de voyelles « aie » qui font le son [e].

Alors que certains élèves essaient de déchiffrer le mot « haies », d'autres, c'est-à-dire **21% ont verbalisé un mot** (21 sur 98). La plupart d'entre eux (57%) a lu [ʃa] et certains ont prononcé: [œj], [ɛʀ], [ʃɛz] ou [ʃascœʀ]. Il apparaît que ces élèves, ont, à partir d'indices graphiques, verbalisé un mot qui leur est certainement plus familier. Dans le premier cas, il est fort probable que l'image du chat située à proximité du mot « haies » les ait amenés à commettre cette erreur (*cf.* texte ci-dessus).

On observe ainsi un certain nombre d'erreurs commises sur le mot « haies » dues certainement à ses particularités grapho-phonémiques. En effet, on remarque que ce mot se compose de plusieurs lettres qui représentent un seul phonème : le « h » aspiré qui ne se prononce pas, les lettres « ai » pour le son [ɛ], le « e » qui marque le féminin et le « s » du pluriel. Ainsi, face à ces difficultés, il semble que les élèves ne se comportent pas tous de la même façon : la moitié ne prononcent rien, d'autres, essaient de déchiffrer alors que d'autres encore essaient de produire un mot.

- « printemps »



Concernant la première apparition du mot « printemps », **73% des élèves a commis une erreur de décodage** (58 sur 79). La plupart ont effectué un décodage littéral incomplet du mot, souvent la première syllabe. Les erreurs les plus fréquentes sont les prononciations [pʀitə] (16%), [pʀɛ̃tə] (10%), [pʀintə] (9%) ou encore [pʀinətə] (3%). Il est possible, que débordés par l'information, les élèves n'aient pas su comment la traiter et se soient découragés pour lire le mot en entier.

On remarque que **23% des élèves ne l'ont pas lu** (18 élèves). Comme pour le mot « haies », il est possible que les élèves aient préféré ne rien prononcer plutôt que de commettre une erreur de lecture.

Les élèves qui ont verbalisé un mot sont seulement 4% (3 élèves). Ces derniers ont lu [pRĕsə], [pĕRmi] ou [pĭRə]. Néanmoins, on se demande si ces productions sont dotées d'une intention lexicale étant donné qu'elles réutilisent les lettres du mot d'origine.

La seconde apparition du mot « printemps » totalise aussi 70% d'erreurs de décodage (45 élèves) et on retrouve les mêmes erreurs que lors de la première apparition du mot.

Ici, aussi on observe 23,5% de non lectures (15 élèves).

Enfin, les verbalisations de mots représentent 6,5% et comme lors de la première apparition du mot, on se demande si ce sont réellement des intentions lexicales étant données qu'on retrouve des lettres du mot d'origine dans les erreurs des élèves.

Le mot « printemps » est un mot long, mais certainement plus fréquent que le mot « haies » ou du moins il est plus familier aux élèves. Il est donc possible que ce soit sa longueur qui ait fait difficulté ; ce qui expliquerait la majorité des décodages partiels. D'autre part, les lectures incomplètes semblent démontrer que c'est aussi la structure graphémique de la 2^{ème} syllabe qui pose problème. En effet, cette syllabe contient deux phonèmes [t] et [ã]. L'attaque est un phonème consonantique et la rime un phonème vocalique. On remarque que cette dernière, comme pour le mot « haies » est représentée par plusieurs lettres, plus précisément, quatre lettres : « em » pour le son [ã] et les lettres muettes « p » et « s ». Il est donc possible que toutes ces informations aient empêché les élèves de CP de lire correctement ce mot.

En **résumé**, il semble qu'en CP, les mots les plus difficiles à lire sont ceux qui contiennent des graphèmes composés et des lettres muettes. Face à ces difficultés, la plupart des élèves semblent vouloir essayer de les déchiffrer avant tout, alors que certains semblent préférer peut-être ne rien prononcer au risque de se tromper. Il est possible que ces élèves préfèrent ne rien dire plutôt que de commettre une erreur. Enfin, quelques-uns ont verbalisé des mots. Pour certains d'entre eux, l'intention de donner du sens semble être claire alors que pour d'autres, il nous est difficile de voir s'il y a eu réellement une intention lexicale. On remarque que pour un mot (« haies »), lorsqu'un indice autre que l'écrit est présent, les

productions de mots sont plus nombreuses. Cela signifierait donc que, face à des mots difficiles à lire, les élèves auraient recours à d'autres indices plus accessibles que l'écrit.

III. Les élèves de CE1 (N=180)

On remarque que pour les élèves de CE1, le mot « printemps » ne semble plus poser de difficulté de lecture à la majorité des élèves puisqu'en moyenne seulement 7,5% des élèves commettent encore une erreur de lecture. En revanche, d'autres mots viennent s'ajouter. Il s'agit des mots « *féeries* », « *geai* », « *piaille* », « *l'ennui* », « *d'osier* » et on retrouve en dernier le mot « *haies* » (cf. tableau 12).

Tableau 12 : mots sur lesquels au moins la moitié des élèves de CE1 ont commis une erreur de lecture

N° mot	Mot	Nb élèves qui l'ont lu	Nb élèves qui ont commis une erreur	% élèves qui ont commis une erreur
68	<i>féeries</i>	169	127	75
96	<i>geai</i>	155	102	66
85	<i>piaille</i>	161	104	65
70	<i>l'ennui</i>	168	108	64
105	<i>d'osier</i>	151	94	62
10	<i>haies</i>	180	100	56

Lecture : en CE1, 169 élèves sur 180 ont lu le mot « *féeries* ». Parmi eux, 127, soit 75%, ont commis une erreur de lecture.

Ainsi, 75% des élèves qui ont lu le mot « *féeries* » ont commis une erreur et 66% ont mal lu le mot « *geai* ». Pour le mot « *piaille* », ils sont 65% à s'être trompés lorsqu'ils l'ont lu. Le mot « *l'ennui* » a entraîné à peu près la même proportion d'élèves à commettre une erreur de lecture puisqu'ils sont 64%, tout comme le mot « *d'osier* » sur lequel 62% des élèves se sont trompés. Enfin, concernant le mot « *haies* », bien qu'ils soient moins nombreux, les élèves de CE1 sont encore plus de la moitié à l'avoir mal lu : 56% au lieu de 70% en CP.

Dans le détail, on constate que, pour quatre mots sur six : « *féeries* », « *piaille* », « *d'osier* » et « *haies* » ; la plupart des élèves qui se sont trompés ont verbalisé un non-mot suite à une erreur de déchiffrage.

- « *féeries* »

Annie, Annie! au doigt joli, une églantine laisse du sang :
au bout du temps des **féeries** viendra l'ennui.

En effet, le mot « *féeries* » a entraîné **78% d'erreurs de décodage**. La majorité des élèves (64%) a lu [fɛri] : il se pourrait que les élèves aient reconnu le morphogramme « fée » et l'aient prononcé comme tel. Cette prononciation, bien qu'acceptable, n'est pas la prononciation commune qui est [fœri]. Ce mot peut donc être qualifié d'irrégulier puisque que sa prononciation ne dépend pas des règles orthographiques. Ainsi, si on n'a jamais rencontré ce mot nous serions tentés de lire [fɛri].

20% des élèves ont verbalisé un mot (25 sur 127). La plupart d'entre eux a prononcé [fɛrie] (19). Ici, il est possible que les élèves aient effectué un déplacement du phonème [e] à la fin du mot et aient pensé au mot « férié ». Compte tenu qu'il n'y pas a eu d'ajout(s) de lettre(s), nous ne pouvons affirmer s'il y a eu une intention lexicale ou pas. Les élèves qui semblent avoir eu une intention lexicale sont peu nombreux (N=3). Ils ont prononcé [fɛrmjɛ], [fœrikə] ou [fœvriɛ]. On remarque que leur production se base un décodage partiel du mot (le début) auquel ils ont ajouté des lettres différentes de celles qui composent le mot d'origine. Cet ajout d'éléments laisse penser que ces élèves ont eu l'intention de donner du sens à leur lecture et ont ainsi verbalisé un mot qui leur est plus familier.

Enfin, on remarque que, **très peu d'élèves n'ont pas essayé de lire** ce mot (2,5%).

Le caractère irrégulier du mot « *féeries* » semble donc avoir posé des difficultés de lecture aux élèves de CE1, et face à cette difficulté, ces derniers ont essentiellement essayé de le déchiffrer sans chercher à y donner du sens.

- « *piaille* »

L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau **piaille** sous la pente des bardeaux et, vif et gai, le geai
sur l'écaille argentée du bouleau, promène un brin d'osier.

En ce qui concerne le mot « *piaille* », **77% des élèves qui se sont trompés ont commis une erreur de décodage** (soit 80 élèves sur 104) et ont essentiellement lu [pɛjə], [pjɛjə] ou [pjɛlə]. Il semblerait que les élèves aient été perturbés par la suite des lettres « iai » et peut-être par la répétition de la lettre « i ». En effet, on remarque que pour deux de ces verbalisations, la lettre « i » a été supprimée ou réutilisée. Cette difficulté semble les avoir amenés à commencer la lecture du mot selon le type de syllabe le plus courant en français

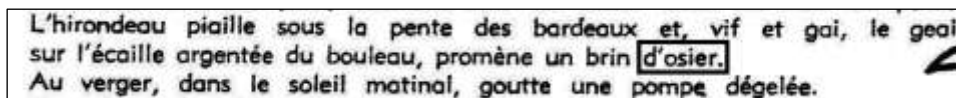
(CV) et à découper le mot de la façon suivante : « pai_lle », « pi_ai_ille » ou encore « pi_ai_lle ».

Par ailleurs, **16% des élèves de CE1 qui se sont trompés ont verbalisé un mot** : ils ont lu [pajə] (17 élèves). On remarque que le mot prononcé, est phonémiquement proche du mot d'origine. De plus, si on regarde la phrase précédente, on constate que le mot « paille » est situé juste avant. Ainsi, il est probable que les élèves aient crus que ce mot se répétait.

Enfin, **6% des élèves se sont abstenus de lire** le mot « paille » (6 élèves).

Il se pourrait donc que pour le mot « paille », la difficulté réside dans la suite des lettres « iai » et notamment la répétition du « i ». Ainsi, face à cette difficulté, les élèves de CE1, ont effectué un découpage erroné du mot, les amenant ainsi à verbaliser un non mot.

- « d'osier »



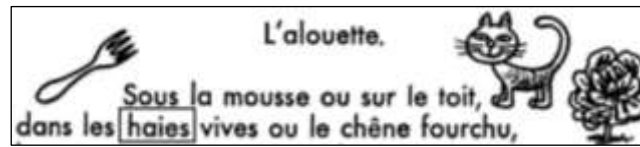
En ce qui concerne le mot « d'osier », **75% des élèves ont commis une erreur de décodage** due à une mauvaise conversion grapho-phonémique (soit 70 sur 94). Ils ont prononcé [dosie]. Il semble que pour ce mot, ceux qui ont commis une erreur, ne connaissent pas encore la règle de prononciation du « s » qui, entre deux voyelles, se prononce [z].

Ici, **22% se sont abstenus de lire** le mot « d'osier » (21 élèves).

Et seulement **3% ont verbalisé un mot** en prononçant [dozɛj], [dwazijɔ̃] ou [dwazo] (3 élèves). Il semble que les élèves qui ont effectué les deux premières erreurs ont essayé de donner du sens à leur lecture puisqu'on observe un ajout de lettres différentes de celles qui composent le mot d'origine. En revanche, pour celui qui a lu [dwazo], il est aussi possible que son erreur soit due à un déplacement et une suppression de lettres et que son intention de départ n'était pas de donner du sens.

Il apparaît donc que pour le mot « d'osier », en CE1, la difficulté principale réside dans la prononciation du graphème contextuel « s » qui entre deux voyelles se lit [z] puisque les trois quart des élèves ont commis une erreur de conversion grapho-phonémique.

- « haies »



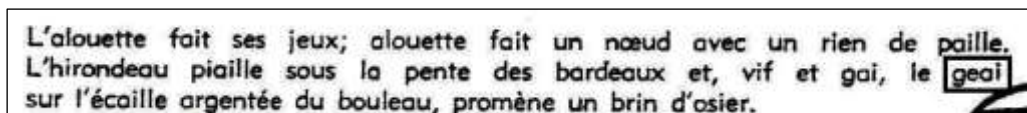
Pour finir, alors qu'en CP le mot « haies » entraînait essentiellement des abstentions de lecture, on constate qu'en CE1, **59% des élèves ont commis une erreur de décodage**. De plus, il apparaît que le type d'erreur a changé d'une classe à l'autre. En effet, alors qu'en CP, les élèves qui se sont trompés, ont effectué une mise en correspondance des graphèmes et des phonèmes erronés (« haies » lu [ʃe] ou [ʃε]). On remarque qu'en CE1, ce type d'erreur a quasiment disparu (4%). A ce niveau les élèves lisent correctement le mot « haies » mais commettent une erreur de liaison avec l'article qui précède et lisent [leze]. Il se pourrait qu'à ce niveau scolaire, les élèves aient abordé les cas de liaison mais peut-être pas les cas particuliers, notamment celui du « h » aspiré qui empêche toute liaison avec le mot précédent.

Par ailleurs, on remarque aussi que **la verbalisation de mots a diminué depuis le CP**. Alors qu'au niveau précédent, 21% des élèves qui se sont trompés ont prononcé un mot, ils sont 13% en CE1 et ont lu pour la plupart d'entre eux [ɛlə], [ɛklɛR] ou [ɛR]. On constate que plus aucun élève n'a recours en CE1 au dessin du chat pour lire le mot « haies ».

Enfin, on observe encore **28% des élèves qui ne prononcent pas ce mot**.

Ainsi, il apparaît qu'en CE1, le mot « haies » pose encore des difficultés de lecture. Néanmoins, elles ne se situent pas au même niveau. En effet, alors que pour les lecteurs débutants, la difficulté principale est dans la conversion des cinq lettres en un seul phonème. Il apparaît ici, que la lecture du mot semble être acquise et que c'est plutôt le phénomène de liaison qui a entraîné des erreurs.

- « gai »



Contrairement aux mots précédents, dont l'essentiel des élèves a verbalisé un non mot, on constate que pour le mot « gai », **la plupart des élèves qui ont commis une erreur, ont prononcé un mot** : [gɛ] (47% d'entre eux). On observe ici le même phénomène qu'avec le mot « piaille » lu [pajə]. En effet, le mot prononcé par les élèves est phonémiquement proche du mot d'origine. De plus, le mot « gai » se situe juste avant dans le texte. Ainsi, comme

précédemment, on se demande si les élèves n'auraient pas cru que le mot « gai » se répétait ou si, cette erreur est due à une mauvaise conversion grapho-phonémique de la lettre « g ». Parmi, les élèves qui ont verbalisé un mot, on constate que 9% ont lu [ʒœdi], [ʃjẽ], [ʒest], [ʒeã] (soit 9 élèves). L'ajout de lettres différentes de celles qui composent le mot d'origine semble confirmer leur intention de produire un mot.

Les élèves qui ont verbalisé un non mot sont encore nombreux puisqu'ils sont 34% (35 élèves sur 102). Leurs erreurs relèvent d'un décodage littéral puisqu'ils ont lu [ʒea], [ʒœe] ou encore [ʒee]. On remarque ici que, comme le mot « haies », le mot « geai » se structure par une suite de voyelles. Il est donc fort probable que, face à cette suite de voyelles, les élèves n'aient pas su comment les traiter.

Enfin, **9% des élèves n'ont pas lu le mot « geai »** (soit 9 élèves).

Ainsi, il apparaît que pour le mot « geai », la difficulté est la même que pour le mot « haies », il s'agit de la suite de voyelles qui forment un phonème : le [ɛ]. Bien que la difficulté soit la même pour ces deux mots, elle n'est pas toutefois traitée de la même façon. En effet, on remarque qu'en CP, la majorité des élèves préfèrent s'abstenir de lire alors qu'en CE1, les élèves semblent se servir d'indices (similitude avec un autre mot) pour produire un mot.

- « *l'ennui* »

Annie, Annie! au doigt joli, une églantine laisse du sang :
au bout du temps des féeries viendra **l'ennui.**

Le mot « *l'ennui* », quant à lui, totalise **autant de verbalisations de non mots que de mots**. En effet, 48% des élèves qui ont mal lu ce mot ont commis une erreur de décodage (52 élèves). Ils ont pour l'essentiel d'entre eux prononcé [lənɥi]. Cette erreur peut être due à l'article élide « l' » et à la double consonne « n ». Comme pour le mot « paille », il semblerait que les élèves aient voulu décoder ce mot en commençant la lecture selon le type CV, ce qui les aurait amenés à effectuer un découpage erroné du mot.

Les élèves qui ont verbalisés un mot sont donc également 48%. Parmi eux, 32% (35 élèves) ont lu [lənɥi] et 16% [lanɥi]. Ici, on se demande si cette erreur n'aurait pas été influencée par la présence du dessin de la nuit situé un peu plus bas dans le texte.

Enfin, seulement **4 % des élèves se sont abstenus de lire** le mot (4 élèves).

Il apparaît donc que la difficulté principale du mot « l'ennui » soit l'article élide amenant les élèves effectuer un découpage erroné du mot ; certains prononçant un non mot alors que d'autres, prononcent un mot basé sur indice iconique et non graphique.

En **résumé**, alors que les élèves de CP, face à un mot difficile à lire, essaient de déchiffrer et quand ils n'y arrivent pas, ne lisent pas, les élèves de CE1 font essentiellement du décodage. Les non réponses ne sont plus le type d'erreur principal. Par ailleurs, on observe que les productions de mots sont toutefois plus nombreuses que chez les élèves de CP. On voit apparaître que les élèves de CE1, pour la lecture de certains mots, ont tendance à utiliser les similitudes grapho-phonémiques alors que les élèves de CP avaient recours aux dessins.

IV. Les élèves de CE2 (N=174)

En CE2, les mots qui ont posé le plus de difficultés de lecture sont à peu près les mêmes qu'en CE1. On retrouve ainsi les mots « *geai* », « *piaille* », « *féeries* » et « *haies* ». En revanche, les mots « d'osier » et « l'ennui » semblent être acquis et ils sont remplacés par le mot : « *poisons* ».

Tableau 13 : mots sur lesquels au moins la moitié des élèves de CE2 ont commis une erreur de lecture

N° mot	Mot	Nb élèves qui l'ont lu	Nb élèves qui ont commis une erreur	% élèves qui ont commis une erreur
178	poisons	107	95	89
96	geai	169	108	64
85	piaille	172	105	61
68	féeries	174	104	60
10	haies	174	87	50

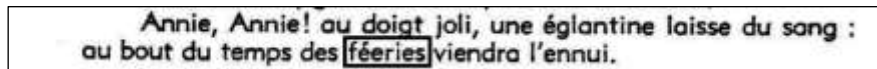
Lecture : en CE2, 107 élèves sur 174 ont lu le mot « poisons ». Parmi eux, 95, soit 89%, ont commis une erreur de lecture.

Ainsi, le mot qui a entraîné le plus de difficultés de lecture est « poisons » puisque 89% des élèves qui l'ont lu se sont trompés. Le mot « geai » quant à lui engendre presque autant d'erreurs de lecture par rapport au niveau précédent (64% en CE2 contre 66% en CE1). Il en est de même pour le mot « piaille » puisque 61% des élèves de CE2 ont commis une erreur de lecture contre 65% en CE1. Le mot « haies », lui aussi, entraîne autant d'erreurs qu'au niveau précédent puisqu'ils sont 50% à s'être trompés contre 56% en CE1. En revanche, on remarque

la proportion d'élèves à avoir commis une erreur de lecture sur le mot « féeries » a diminué : ils sont passés de 75% en CE1 à 60% en CE2.

Voyons dans le détail les types d'erreurs commises lors de la lecture de ces mots.

- « *féeries* »



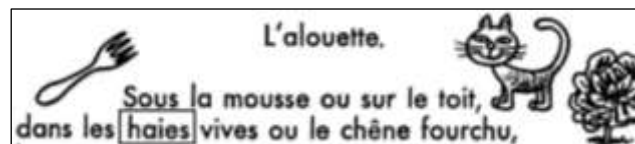
Les erreurs commises sur le mot « *féeries* » sont, comme en CE1, **essentiellement des erreurs de décodage**. Il a été principalement lu [feri] (68%).

25% des élèves ont verbalisé un mot. Pour la majorité des élèves, on retrouve la même lecture qu'en CE1, à savoir [ferie]. Seulement trois élèves ont ajouté des lettres différentes de celles qui composent le mot d'origine pour verbaliser un mot et lire ainsi : [fermje], [feerik] ou [fevrie].

La proportion d'élèves qui n'ont pas lu ce mot est la même qu'en CE1, ils sont 3%.

Ainsi, au vu des types d'erreurs commises, il semble que le mot « féeries » ne fasse pas encore partie du vocabulaire des élèves, ce qui explique ce fort taux d'erreurs de décodage.

- « *haies* »



Le type d'erreur commis sur le mot « *haies* » est aussi le même qu'au niveau précédent. Les élèves ont **essentiellement effectué une liaison abusive** en prononçant [leze] (64%).

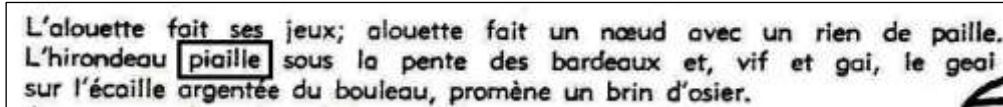
On remarque que **les erreurs lexicales sont moins nombreuses** tout comme les non lectures. En effet, alors qu'en CE1, 13% lisaient [fa], ils ne sont plus que 6% en CE2.

De la même façon, alors que 28% des élèves n'ont pas lu ce mot au niveau précédent, ils ne sont plus que 6% en CE2.

Il apparaît donc que, bien que le mot « haies » semble être acquis, le phénomène de liaison et notamment, les exceptions, ne sont pas encore maîtrisées, d'où le nombre important d'erreurs dues à une liaison abusive. De plus, on remarque que les élèves qui ont produit un

mot, ont prononcé un mot en rapport avec le dessin du chat présent autour du texte. Il semble donc, qu'à ce niveau scolaire aussi, lorsqu'un mot paraît difficile à lire, les élèves ont recours à d'autres indices, plus accessibles que l'écrit.

- « *piaille* »



L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau **piaille** sous la pente des bardeaux et, vif et gai, le geai
sur l'écaille argentée du bouleau, promène un brin d'osier.

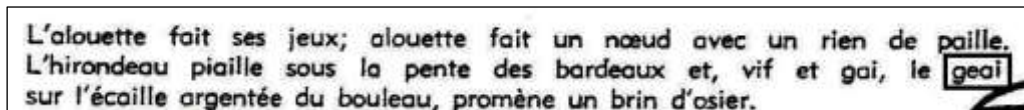
Le mot « piaille » entraîne toujours **essentiellement des erreurs de décodage** (68%). On retrouve ainsi les prononciations [pjɛjə] ou [pjɛlə].

On constate que le nombre d'élèves à avoir lu [pajə] a presque doublé : ils sont passés de 16% à 29%.

La proportion d'élèves qui n'a pas lu le mot reste en revanche à peu près le même : 6% en CE1 contre 4% en CE2.

On remarque donc qu'en CE2, la juxtaposition des voyelles dans le mot « piaille » pose encore des difficultés de traitement amenant les élèves à commettre des erreurs de décodage. Néanmoins, on constate que la proportion d'élèves qui ont produit un mot a quasiment doublé d'une classe à l'autre.

- « *geai* »



L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau piaille sous la pente des bardeaux et, vif et gai, le **geai**
sur l'écaille argentée du bouleau, promène un brin d'osier.

Concernant, le mot « geai », il a été lu [ge] par une majorité d'élèves (68%), comme en CE1. Ici, aussi, on se demande si cette prononciation est due à une méconnaissance de la règle contextuelle qui régit la prononciation du « g » ou si c'est le mot « gai » situé juste avant qui a interféré dans la lecture de ce mot.

Bien qu'ils soient moins nombreux (23% au lieu de 35% en CE1), on observe encore des erreurs de décodage et on retrouve les prononciations [gəe] ou [ʒəe].

Les élèves qui n'ont pas lu ce mot, sont aussi nombreux qu'au niveau précédent (9%).

Ainsi, il apparaît que le mot « geai » entraîne essentiellement des verbalisations de mots.

- « *poisons* »

Un lac étire ses calmes rives et, quand le soir descend,
le miroir de ses eaux reflète les **poisons** des brignoles perfides.

Enfin, le mot « *poisons* » a **essentiellement amené les élèves à verbaliser un mot** et donc à lire [pwasɔ̃] (99%). On constate que les élèves ont verbalisé un mot qui leur est certainement plus familier, toutefois, nous ne pouvons écarter l'idée que cette lecture peut être aussi le résultat d'une erreur de conversion grapho-phonémique. En effet, il est possible que, comme pour le mot « d'osier » en CE1, les élèves aient commis cette erreur à cause d'une méconnaissance des règles de prononciation de la lettre « s ». Cependant, nous supposons que si cette règle n'était pas maîtrisée alors, nous aurions dû retrouver le mot « d'osier » lu principalement [dosie] en CE2. Ce mot n'apparaissant pas ici, il est donc fort probable que les élèves aient voulu dire « poissons ». De plus, ce mot apparaît dans un co-texte sémantique piège : il est entouré de mots faisant partie du champ lexical de l'eau.

En **résumé**, il apparaît que face à des mots difficiles à lire, les élèves de CE2 (comme les élèves de CE1) essaient essentiellement de les déchiffrer sans chercher à attribuer du sens. Néanmoins, on constate que les productions de mots sont plus nombreuses qu'aux niveaux précédents et que pour ces verbalisations, les élèves semblent avoir recours à d'autres indices tels que les dessins, la ressemblance grapho-phonémique ou encore le co-texte sémantique.

V. Les élèves de CM1 (N=188)

En CM1, le nombre de mots qui a posé des difficultés de lecture diminue et on ne répertorie plus que quatre mots. Les mots « haies », « féeries », « paille » et « l'ennui » semblent être acquis pour une majorité des élèves de CM1. En revanche, les erreurs sur les mots « *poisons* » et « *geai* » sont encore présentes et on voit apparaître deux mots : « *lune* » et « *la* ».

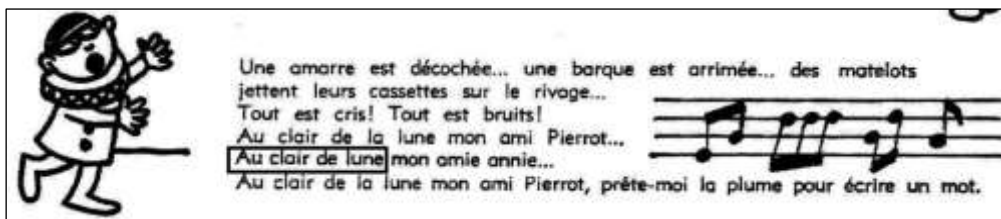
Tableau 14 : mots sur lesquels au moins la moitié des élèves de CM1 ont commis une erreur de lecture

N° mot	Mot	Nb élèves qui l'ont lu	Nb élèves qui ont commis une erreur	% élèves qui ont commis une erreur
246	lune	105	96	91
178	poisons	165	142	86
96	geai	187	105	56
260	La	100	54	54

Lecture : en CM1, 105 élèves sur 188 ont lu le mot « lune ». Parmi eux, 96, soit 91%, ont commis une erreur de lecture.

En effet, parmi les mots qui ont entraîné le plus d'erreurs en CM1, on trouve en premier le syntagme « au clair de lune » qui a été mal lu par 91% des élèves. On retrouve le mot « poisons » sur lequel 86% des élèves ont commis une erreur de lecture. Le mot « geai » pose encore des difficultés puisque 56% des élèves ont commis une erreur de lecture. Enfin, on trouve le mot « la » sur lequel 54% des élèves de CM1 ont commis une erreur.

- « au clair de lune »



En ce qui concerne le syntagme « au clair de lune », presque tous les élèves (99%) ont lu « au clair de la lune » : le mot « la » a été ajouté. On remarque que ce syntagme fait partie d'un paragraphe qui rappelle faussement les paroles de la chanson « Au clair de la lune ». De plus, ce paragraphe est entouré de dessins en rapport avec la chanson: le dessin d'un Pierrot à gauche et des notes de musique à droite. Il est donc possible que les élèves qui se sont trompés aient crus qu'il s'agissait vraiment de la chanson.

Il semble donc ici, que ce soit le contexte (co-texte sémantique et dessins) qui ait amené les élèves à commettre une erreur.

- « *poisons* »

Un lac étire ses calmes rives et, quand le soir descend,
le miroir de ses eaux reflète les **poisons** des brignoles perfides.

En ce qui concerne le mot « *poisons* », on retrouve le même type d'erreur qu'au niveau précédent : les élèves sont une majorité à avoir lu [pwasõ]. Bien que nous ne pouvons savoir clairement si cette erreur est due à une méconnaissance de la règle de prononciation du « s » entre deux voyelles qui se lit [z] ou si c'est une intention lexicale, il est tout de même fort probable, qu'à ce niveau scolaire, les élèves ont cherché à donner du sens et ont prononcé ce mot par rapport au contexte dans lequel il apparaît .

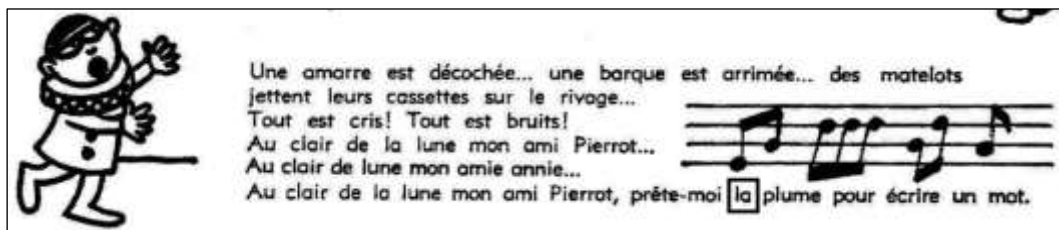
Ainsi, comme pour l'erreur commise sur le syntagme « au clair de lune », il apparaît ici que le contexte, notamment le co-texte sémantique, a probablement amené les élèves à commettre ce type d'erreur.

- « *geai* »

L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau piaille sous la pente des bardeaux et, vif et gai, le **geai**
sur l'écaille argentée du bouleau, promène un brin d'osier.

Le mot « *geai* » est encore essentiellement lu [ge]. Comme pour le mot « *poisons* », il est peu probable que cette erreur soit due à une méconnaissance de la règle de prononciation du « g » lorsqu'il est précédé d'une voyelle. Nous supposons qu'à ce niveau scolaire, les élèves se soient trompés à cause de la similitude avec le mot qui précède « *gai* ». La lecture du premier aurait alors interféré sur la lecture du second.

- « *la* »



Une amarre est décochée... une barque est arrimée... des matelots
jettent leurs cassettes sur le rivage...
Tout est cris! Tout est bruits!
Au clair de la lune mon ami Pierrot...
Au clair de lune mon amie annie...
Au clair de la lune mon ami Pierrot, prête-moi **la** plume pour écrire un mot.

Enfin, en ce qui concerne le mot « *la* », tous les élèves qui ont commis une erreur, ont lu [ta]. Ce type d'erreur semble être assez étrange pour des élèves de CM1, puisque c'est un mot qui ne présente aucune difficulté particulière. Si on regarde de plus près l'environnement dans lequel ce mot apparaît, on constate, que le contexte est ici aussi un piège. En effet, comme pour le groupe de mots « au clair de lune » ce mot fait partie du syntagme « prête-moi *la*

plume » qui fait suite aux paroles « pièges » de la chanson « Au clair de la lune ». Ainsi, comme précédemment, il est possible que les élèves aient cru que ce sont réellement les paroles de la chanson qui sont écrites et ont donc été amenés à lire « prête-moi **ta** plume ».

En **résumé**, il apparaît qu'en CM1, les mots qui ont entraîné le plus d'erreurs, sont des mots qui apparaissent dans un contexte particulier : ces mots ne sont pas prédictibles et apparaissent donc comme des pièges pour les élèves puisque la majorité des élèves a verbalisé des mots en cohérence syntaxique et sémantique avec le contexte. Ce constat, laisse penser qu'à ce niveau scolaire, la recherche de sens s'élargit et se situe au-delà du mot, c'est-à-dire du syntagme ou de la phrase.

VI. Les élèves de CM2 (N=176)

On remarque que les élèves de CM2 ont essentiellement commis des erreurs lors de la lecture du syntagme « au clair de lune » et du mot « poisons ».

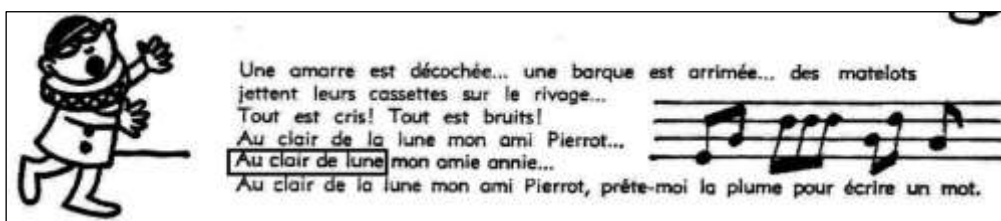
Tableau 15 : mots sur lesquels au moins la moitié des élèves de CM2 ont commis une erreur de lecture

N° mot	Mot	Nb élèves qui l'ont lu	Nb élèves qui ont commis une erreur	% élèves qui ont commis une erreur
246	lune	142	129	91
178	poisons	169	142	84

Lecture : en CM2, 142 élèves sur 176 ont lu le mot « lune ». Parmi eux, 129, soit 91%, ont commis une erreur de lecture.

En effet, 91% des élèves ont commis une erreur de lors de la lecture du syntagme « au clair de lune » et 84% ont commis une erreur de lecture sur le mot « poisons ».

- « *au clair de lune* »



Tous les élèves de CM2 qui ont commis une erreur sur le syntagme « *au clair de lune* » ont, comme en CM1, rajouté le mot « la » et ont donc lu « au clair de la lune ». Ce type d'erreur semble donc confirmer l'idée selon laquelle les élèves ont été influencés par le contexte qui rappelle la chanson.

- « *poisons* »

Un lac étire ses calmes rives et, quand le soir descend,
le miroir de ses eaux reflète les **poisons** des brignoles perfides.

Concernant le mot « *poisons* », on remarque que quasiment tous les élèves (99%) ont lu [pwasɔ̃]. Le fait que cette erreur se retrouve encore en fin d'école primaire semble donc confirmer le fait que ce n'est plus la prononciation du « s » qui pose des difficultés mais plutôt le contexte dans lequel le mot apparaît.

En **résumé**, il apparaît qu'en fin d'école primaire, en CM2, les mots qui ont entraîné le plus d'erreurs sont, comme en CM1, des mots qui s'inscrivent dans un contexte piège.

CONCLUSION

Cette analyse avait pour objectif de voir quels sont les mots qui ont entraîné le plus de difficultés de lecture et comment, face à ces difficultés, les élèves procèdent.

Nous avons constaté qu'au début de l'apprentissage de la lecture, les mots qui posent le plus de difficultés sont des mots longs, peu fréquents (ou du moins pas familiers) et qui présentent une structure graphémique assez complexe. En effet, on remarque que les erreurs apparaissent lorsque plusieurs lettres représentent un phonème. Nous avons remarqué que les lettres muettes (« **haies** »), les suites de voyelles (« **haies** ») ou de consonnes (« **printemps** ») amènent les élèves à commettre des erreurs de lecture.

Pour les lecteurs intermédiaires, les mots qui sont les plus difficiles à lire contiennent encore des graphèmes composés (« **printemps** »). On voit apparaître des mots contenant des graphèmes irréguliers (« **féeries** ») ou contextuels (« **geai** », « **d'osier** »).

Pour les lecteurs plus avancés, on constate que les mots qui ont entraîné le plus de difficulté changent. Il ne semble plus que ce soit la structure grapho-phonémique qui soit en cause mais plutôt le contexte dans lequel ils s'inscrivent. En effet, ces mots ne sont pas

prédictibles en fonction du contexte, autrement dit ce ne sont pas des mots auxquels on s'attend (« au clair de la lune, mon ami Pierrot, prête-moi **la** plume pour écrire un mot »).

Face à ces difficultés les stratégies de lecture semblent être différentes avec le niveau d'avancement en lecture.

Ainsi, on observe qu'au début de l'apprentissage de la lecture, les apprentis lecteurs semblent vouloir avant tout essayer de déchiffrer les mots sans donner du sens à leur lecture. On remarque que les élèves qui ne prononcent rien sont aussi très nombreux. Il est possible que pour ces derniers, il est préférable de ne rien dire plutôt que de se tromper. Enfin, on constate que le peu d'élèves qui ont verbalisé un mot ont eu recours à un dessin qui se situe autour du texte (celui du chat). Il est donc possible que, lorsque les lecteurs débutants se retrouvent face à un mot qu'ils ne savent pas traiter, ils ont recours à un indice plus accessible que l'écrit, ici, le dessin.

A un niveau intermédiaire, on constate que les élèves font encore essentiellement du décodage. En revanche, les non réponses ne sont plus le type d'erreur principal. Autrement dit, les élèves essaient de déchiffrer même s'ils se trompent. Par ailleurs, on note aussi, que les productions de mots deviennent plus nombreuses. Cela signifierait que les élèves commencent à vouloir attribuer du sens à leur lecture. A ce niveau, on note que certaines productions de mots semblent être dues à des dessins mais aussi à des similitudes grapho-phonémiques avec d'autres mots. Ainsi, il est possible que face à des mots difficiles à lire, les lecteurs intermédiaires, comme les lecteurs débutants, cherchent d'autres indices pour les mots.

En revanche, à un niveau plus avancé, il semblerait que la recherche de sens s'élargit. Ainsi, les élèves essaieraient de produire du sens non plus au niveau du mot, mais au niveau du syntagme voire de la phrase.

Par conséquent, il apparaît que certaines variables linguistiques (longueur, fréquence, structure graphémique) semblent avoir un effet lors de la lecture de mots et entraînent des difficultés de lecture. Alors que certaines erreurs semblent dues au mot d'autres semblent être davantage caractéristiques au test même de *L'alouette* (dessins, similitudes grapho-phonémiques, co-texte sémantique). Ainsi, nous nous proposons dans les chapitres qui suivent d'étudier tour à tour l'effet de ces variables afin de voir dans quelle mesure elles nous renseignent sur l'apprentissage de la lecture.

L'influence des variables linguistiques

Plan

Chapitre 9 : L'influence de la longueur et de la fréquence d'un mot.....	166
Chapitre 10 : L'influence de la structure graphémique d'un mot.....	195
CONCLUSION	224

Chapitre 9

L'INFLUENCE DE LA LONGUEUR ET DE LA FREQUENCE D'UN MOT

Plan du chapitre

INTRODUCTION	167
PREDICTIONS	168
METHODOLOGIE DE TRAITEMENT DES DONNEES.....	169
I. Est-ce que la longueur et la fréquence d'un mot font difficulté ?	171
<i>Le nombre d'erreurs varie-t-il avec la longueur et la fréquence d'un mot ?</i>	<i>171</i>
<i>L'effet de longueur et de fréquence diminue-t-il avec l'avancée dans la scolarisation ?</i>	<i>175</i>
II. Les processus de lecture selon la longueur et la fréquence d'un mot.....	177
<i>Les élèves de CP (N = 143)</i>	<i>178</i>
<i>Les élèves de CE1 (N = 180)</i>	<i>180</i>
<i>Les élèves de CE2 (N = 174)</i>	<i>182</i>
<i>Les élèves de CM1 (N = 188)</i>	<i>184</i>
<i>Les élèves de CM2 (N = 182)</i>	<i>186</i>
III. Evolution des processus de lecture	188
<i>Evolution des processus de lecture des mots longs peu fréquents (ou rares)</i>	<i>188</i>
<i>Evolution des processus de lecture des mots courts fréquents (ou très fréquents)</i>	<i>190</i>
CONCLUSION	192

INTRODUCTION

Plusieurs recherches portant sur l'apprentissage de la lecture et plus précisément sur l'identification des mots ont permis de mettre en évidence un effet de la longueur du mot lors de la lecture. Ainsi, dans leur étude sur l'acquisition de la lecture¹ Content et Leybaert (1992) montrent qu'il existe une influence de la longueur. Ils comparent les performances des enfants en 2e, 3e et 6e année de primaire pour des mots de haute et basse fréquences et des pseudo-mots², de longueurs variées. Les résultats montrent qu'en 2e année, l'effet de longueur est très important sur les temps de lecture des mots (fréquents ou rares) et des pseudo-mots, alors qu'en 4^e année, les temps de réponses aux pseudo-mots sont davantage affectés par la longueur que les mots. Concernant les erreurs, l'effet de longueur est plus marqué pour les pseudo-mots, que ce soit en 2e ou 4e année. Ils en concluent que la longueur a un effet sur l'identification des pseudo-mots et sur les temps de réponse. Ils supposent que le faible effet de la longueur sur les mots tient à un rôle moindre de l'assemblage, ce qui impliquerait l'existence d'une composante d'adressage.

En 1999, Mousty, Leybaert et Alegria ont mené une étude longitudinale en lecture et en orthographe sur 217 enfants testés en 2e et 4e année³. L'outil de mesure, BELEC (Batterie d'Évaluation du Langage Écrit) permet d'identifier les difficultés rencontrées par certains enfants en lecture et en écriture. Les résultats aux évaluations en lecture montrent que les items courts (5 lettres) sont mieux lus et plus rapidement tant en 2e année qu'en 4e année. Cet effet est plus marqué chez les enfants de 2e année.

Plus récemment, quelques études (Ferrand & New, 2003; Valdois, et *al.*, 2006) ont été menées afin d'éclairer les mécanismes sous-jacents à la reconnaissance visuelle et à la lecture de mots polysyllabiques. Les conclusions sont concordantes mettant en évidence un effet de longueur durant la lecture de pseudo-mots.

Bien que la variable longueur apparaisse dans plusieurs études relatives à l'apprentissage de la lecture (notamment lors de la reconnaissance d'un mot) elle n'est, à notre connaissance, jamais analysée seule. Souvent cette variable est étudiée avec la

¹ « *L'acquisition de la lecture : influence des méthodes d'apprentissage* » (1992).

² Ici, la fréquence et la lexicalité peuvent être considérées comme des variables contrôles.

³ « *Évaluation des habiletés de lecture et d'orthographe au moyen de BELEC. Données longitudinales auprès d'enfants francophones testés en 2e et 4e années* »

fréquence. En effet, d'après la loi de Zipf (1935), la longueur d'un mot est étroitement liée à la fréquence : les mots courts tendent à être plus fréquents que les mots longs.

Adoptant une autre perspective, celle d'une lecture plus proche de ses conditions ordinaires de réalisation, nous avons souhaité étudier l'effet de ces variables non pas sur des mots isolés mais en situation de lecture textuelle.

Travaillant sur l'apprentissage de la lecture, nous avons considéré la longueur d'un mot en nombre de lettres (et non de syllabes), mettant ainsi l'accent sur l'aspect visuel du texte écrit, tel qu'il apparaît aux enfants. Au début de l'apprentissage de la lecture la longueur est donc considérée comme une variable typographique.

Dans ce contexte, nous avons choisi de mener une analyse portant, non pas sur les temps de réponse, mais sur les erreurs qui pourraient être dues à ces variables. Autrement dit, **nous avons tenté de voir d'un côté, si le nombre d'erreurs et les difficultés de lecture varient avec la longueur et la fréquence d'un mot et, d'autre part, de quelle manière ces difficultés évoluent au long de l'apprentissage de la lecture.**

PREDICTIONS

On se demande donc, dans quelle mesure les erreurs liées à la longueur et à la fréquence des mots du texte de *L'alouette* nous informent quant au caractère continu ou discontinu de la lecture.

Nous supposons qu'au début de l'apprentissage de la lecture, la longueur écrite et la fréquence d'un mot ont un effet lors de la lecture de ce mot. Autrement dit, plus un mot est long et rare et plus il serait difficile à lire.

De ce fait, d'un point de vue quantitatif, nous nous attendons à ce que les mots longs rares (ou peu fréquents) entraînent davantage d'erreurs que les mots courts fréquents (ou très fréquents). L'apprentissage du code devrait faire diminuer ce type d'erreur dans le temps.

D'un point de vue qualitatif, nous supposons que, plus un mot est long et rare et plus l'enfant aura tendance à effectuer dans ses débuts un décodage du mot. La maîtrise croissante du code grapho-phonétique devrait faire diminuer ce type d'erreur dans le temps et la maîtrise de la procédure lexicale devrait permettre à l'enfant de le reconnaître plus facilement et ainsi de le lire plus aisément.

METHODOLOGIE DE TRAITEMENT DES DONNEES

Pour la variable longueur, nous avons divisé les mots du texte en deux catégories : les mots courts (contenant quatre lettres ou moins) et les mots longs (contenant cinq lettres ou plus). Le texte comporte alors 162 mots courts et 103 mots longs (*cf. CD annexes*).

Pour la variable fréquence, à l'aide de la base de données lexicales Manulex⁴, nous avons divisé les mots du texte en quatre catégories⁵ : les mots rares, peu fréquents, fréquents et très fréquents (*cf. CD annexes*).

Le croisement de ces deux variables (longueur et fréquence) nous amène donc à un total de quatre catégories que nous avons utilisé dans notre étude: les mots courts fréquents (ou très fréquents), les mots courts peu fréquents (voire rares), les mots longs fréquents (ou très fréquents) et les mots longs peu fréquents (ou rares).

La longueur étant liée à la fréquence, nous avons au préalable voulu voir si cette corrélation est valable pour les mots du texte de *L'alouette*. Ainsi, nous avons effectué le test de Bravais Pearson afin de voir s'il existe une relation entre la fréquence et la longueur d'un mot. Les résultats confirment cette hypothèse : plus le nombre de lettres d'un mot augmente et plus la fréquence diminue ($r = -.743$; $p < .01$) (*cf. CD annexes*).

Les enfants n'ayant pas tous lu le même nombre de mots, nous avons par conséquent travaillé sur les rapports entre le nombre d'erreurs commises sur les mots courts/longs très fréquents/fréquents/ peu fréquents ou rares et le nombre de mots courts/longs très fréquents/fréquents/ peu fréquents ou rares lus.




Concernant le traitement statistique des données, nous avons eu recours à des tests non paramétriques. En effet, notre distribution ne suivant pas une loi normale nous nous sommes servie du test de Wilcoxon pour comparer les taux d'erreurs commises sur les mots courts et mots longs à un même niveau (deux groupes appariés), du test de Friedman (quatre groupes appariés car quatre catégories de mots) et des test de Mann-Whitney et Kruskal-Wallis pour les comparaisons des erreurs sur les mots courts d'une part, et des mots longs d'autre part, selon le niveau de scolarisation (groupes indépendants car cinq classes) (*cf. CD annexes*).

⁴ Ortéga, É., & Lété, B. (2010). *eManulex: Electronic version of Manulex and Manulex-infra databases*. Retrieved from <http://www.manulex.org>.


⁵ Les catégories ont été effectuées à partir des Indices de Fréquence Standard (SFI). Plus l'indice est élevé et plus le mot est fréquent. Ainsi, sur la base des quartiles indiqués dans la base de données, du CP au CM2, nous avons constitué quatre catégories : les mots rares dont le $SFI < 22,95$; les mots peu fréquents dont le $22,95 < SFI < 34,40$; les mots fréquents dont le $34,40 < SFI < 43,12$ et les mots très fréquents dont le $43,12 < SFI < 84,96$.

Cartographie 2 : mots longs peu fréquents dans le texte de L'alouette


L'alouette.





Sous la mousse ou sur le toit,
 dans les haies vives ou le chêne **fourchu**
 le printemps a mis ses nids.
 Le printemps a nids au bois.





Annie amie, du renouveau, c'est le doux temps.
 Amie Annie, au bois joli **gamine** le pinson.
 Dans les buis, gîte une biche, au bois chantant.
 Annie, Annie! au **doigt** joli, une **églantine** laisse du sang :
 au bout du temps des **féeries** viendra l'ennui.





L'alouette fait ses jeux; alouette fait un **nœud** avec un rien de paille.
 L'hirondeau **piaille** sous la pente des **bardeaux** et, vif et gai, le geai
 sur l'écaille argentée du bouleau, promène un brin d'osier.
 Au verger, dans le soleil matinal, **goutte** une pompe **dégelée**.
 On voit un bec luisant qui **trille** éperdument des notes claires
 et, dans les **pompres** d'or que suspend la grille antique,
 on surprend des **rixes** de moineaux.
 Au potager s'alignent les cordeaux; l'if est triste à l'horizon
 et lourd et lent l'envol des corbeaux.

Un lac étire ses calmes rives et, quand le soir descend,
 le miroir de ses eaux reflète les **poisons** des **brignoies** **perfidés**.
 Et, quand descend le soir, quand joue la pourpre du couchant,
 le ciel rougit ses eaux.
 Dans la **moire** de l'eau danse l'ombre d'un écueil.
 Tout est cris! Tout est bruits!

Une **amarre** est **décochée**.. une barque est **arrimée**.. des matelots
 jettent leurs cassettes sur le rivage...
 Tout est cris! Tout est bruits!
 Au clair de la lune mon ami Pierrot...
 Au clair de lune mon amie annie...
 Au clair de la lune mon ami Pierrot, prête-moi la plume pour écrire un mot.

o u e i a
 le la les un dans des do ti pu mi

Mots longs peu fréquents (ou rares)

I. Est-ce que la longueur et la fréquence d'un mot font difficulté ?

Le nombre d'erreurs varie-t-il avec la longueur et la fréquence d'un mot ?

Considérant que le temps de lecture d'un mot est d'autant plus long qu'il contient de lettres et qu'il est peu fréquent, nous supposons que la performance dans le traitement de celui-ci s'amoinde en fonction du temps. Cela nous amène donc à penser que, d'une façon générale, les mots courts et fréquents sont plus faciles à lire que les mots longs et rares. Ainsi, nous nous attendons plus précisément à trouver un nombre plus important d'erreurs lors de la lecture de mots longs peu fréquents (ou rares) que de mots courts fréquents (ou très fréquents).

On remarque que, pour l'ensemble des élèves, les mots longs et rares sont ceux qui entraînent le plus de difficultés (*cf.* tableau 16). Plus précisément, on compte 32% d'erreurs lors de la lecture de mots longs peu fréquents (ou rares), 25% d'erreurs lors de la lecture de mots courts peu fréquents (ou rares), 18% d'erreurs lors de la lecture de mots longs fréquents (ou très fréquents) et 7% d'erreurs lors de la lecture de mots courts fréquents (ou très fréquents). Notre hypothèse selon laquelle les mots longs peu fréquents sont plus difficiles à lire que les mots courts fréquents est donc vérifiée ($\chi^2 = 26.175$; ddl = 3 ; $p < .01$).

Par ailleurs, on remarque qu'à fréquence égale les mots longs sont plus difficiles à lire que les mots courts. Ainsi, si on compare les mots courts et les mots longs peu fréquents (ou rares), on constate 25% d'erreurs pour les premiers contre 32% pour les seconds ($\chi^2 = 28.517$; ddl = 1 ; $p < .01$). Il en est de même avec les mots fréquents (ou très fréquents) : 7% d'erreurs sur les mots courts contre 18% sur les mots longs ($\chi^2 = 10.936$; ddl = 1 ; $p < .01$). Ce constat semble donc confirmer un effet de longueur lors de la lecture de certains mots.

De la même façon, si on compare les nombres d'erreurs commis sur des mots de même longueur mais de fréquence différente, on constate que 32% des mots longs peu fréquents ont été mal lus contre 18% des mots longs fréquents ($\chi^2 = 15.238$; ddl = 1 ; $p < .01$). En revanche, la fréquence ne semble pas avoir d'impact lors de la lecture de mots courts. En effet, bien que 25% des mots courts peu fréquents (ou rares) ont été mal lus et 7% pour les mots courts fréquents, il n'existe pas de différence significative entre ces deux catégories ($\chi^2 = 2.342$; ddl = 1 ; *NS.*).

Il semble donc que, d'une manière générale, la fréquence d'un mot associée à la longueur écrite entraîne davantage de difficultés de lecture. Autrement dit, plus un mot est long et peu fréquent (voire rare) et plus il est difficile à lire.

Si on regarde de plus près les résultats par cycle, on remarque que la tendance est la même (*cf.* tableau 16)

Tableau 16 : pourcentages d'erreurs commises selon la longueur et la fréquence des mots et le cycle de scolarisation

Cycle	% d'erreur sur les mots courts fréquents (ou très fréquents)	% d'erreur sur les mots courts peu fréquents (ou rares)	% d'erreur sur les mots longs fréquents (ou très fréquents)	% d'erreur sur les mots longs peu fréquents (ou rares)
Cycle 2	11	40	32	44
Cycle 3	5	18	9	24
Ensemble	7	25	18	32

Lecture : 44% des mots longs peu fréquents (ou rares) qui ont été lus en cycle 2, n'ont pas été bien lus.

En effet, en cycle 2, on constate que les mots longs peu fréquents (ou rares) sont ceux qui ont entraîné le plus de difficultés de lecture. On compte plus exactement, 44% d'erreurs sur les mots longs peu fréquents (ou rares), 40% d'erreurs sur les mots courts peu fréquents, 32% sur les mots longs fréquents et 11% sur les mots courts fréquents. En cycle 2, les mots longs peu fréquents (ou rares) sont donc plus difficiles à lire que les mots courts fréquents (ou très fréquents) ($\chi^2 = -15.284$; ddl = 1 ; $p < .01$).

Comme pour l'ensemble des élèves, on remarque que si on compare les mots courts et les mots longs de fréquence équivalente, ce sont les mots longs qui entraînent le plus de difficultés de lecture : pour les mots fréquents, 32% des mots longs ont été mal lus contre 11% des mots courts ($\chi^2 = -8.556$; ddl = 1 ; $p < .01$) ; pour les mots peu fréquents, 40% des mots courts ont été mal lus contre 44% des mots longs ($\chi^2 = -10.734$; ddl = 1 ; $p < .01$). Ainsi, en cycle 2, à fréquence équivalente, les mots courts sont mieux lus que les mots longs.

En ce qui concerne la fréquence, on remarque que parmi les mots courts, 11% des mots fréquents ont été mal lus contre 40% des mots rares ($\chi^2 = -4.550$; ddl = 1 ; $p < .01$). De la même façon, parmi les mots longs, on remarque que 32% des mots fréquents (ou très fréquents) ont été mal lus contre 44% des mots peu fréquents (ou rares) ($\chi^2 = -6.729$; ddl = 1 ; $p < .01$). Ainsi, à longueur équivalente, les mots peu fréquents (ou rares) sont mieux lus que les mots courts.

Il existe donc bien un effet de longueur et de fréquence en cycle 2. D'ailleurs, 44% des mots longs peu fréquents (ou rares) lus ont entraîné des erreurs contre 11% pour les mots courts fréquents (ou très fréquents) ($\chi^2 = -15.284$; ddl = 1 ; $p < .01$). Notre hypothèse se vérifie donc en cycle 2. Voyons maintenant, ce qu'il en est en cycle 3.

Le constat est le même en cycle 3 puisqu'on constate que les mots longs peu fréquents (ou rares) sont ceux qui ont entraîné le plus de difficultés de lecture. Ainsi, on compte que 24% des mots longs peu fréquents ont été mal lus comme 18% des mots courts peu fréquents, 9% des mots longs fréquents (ou très fréquents) et 5% des mots courts fréquents (ou très fréquents).

Comme en cycle 2, on remarque que si on compare les mots courts et les mots longs de même fréquence, ce sont les mots longs qui sont le plus souvent mal lus.

De la même façon, si on compare les mots fréquents (ou très fréquents) et peu fréquents (ou rares) de même longueur, ce sont les mots peu fréquents qui sont le plus souvent mal lus.

Ici aussi semble se confirmer un effet de la longueur et de la fréquence d'un mot en lecture en cycle 3. Ainsi, on observe que 5% des mots courts fréquents (ou très fréquents) ont été mal lus contre 24% des mots longs peu fréquents (ou rares) ($\chi^2 = -21.264$; ddl = 1 ; $p < .01$).

Ainsi, ces premiers résultats confirment notre hypothèse puisqu'**à fréquence égale, les mots courts sont mieux lus que les mots longs et à longueur égale, les mots fréquents (ou très fréquents) sont mieux lus que les mots peu fréquents (ou rares)**. La quantité d'information à traiter lors de la lecture de mots longs entraînerait alors plus de risques de commettre une erreur de lecture. De la même façon, la fréquence élevée d'un mot, permettrait de le retrouver plus facilement dans son lexique mental. Autrement dit, les mots courts, quelle que soit leur fréquence sont mieux lus que les mots longs et inversement, les mots fréquents, quelle que soit leur longueur écrite, sont mieux lus que les mots rares. Par conséquent, **les mots courts fréquents (ou très fréquents) entraînent moins de difficulté de lecture que les mots longs peu fréquents (ou rares)**.

Maintenant, reste à savoir si cet effet est présent à tous les niveaux scolaires.

Dans le détail, on remarque que cette tendance se retrouve à chaque niveau scolaire : les mots longs peu fréquents (ou rares) entraînent systématiquement davantage d'erreurs par rapport aux mots courts fréquents (ou très fréquents).

Tableau 17 : pourcentages d'erreurs commises selon la longueur et la fréquence des mots et la classe

Classe	% d'erreur sur les mots courts fréquents ou très fréquents	% d'erreur sur les mots courts peu fréquents ou rares	% d'erreur sur les mots longs fréquents ou très fréquents	% d'erreur sur les mots longs peu fréquents ou rares
CP (N = 143)	15	42	46	54
CE1 (N = 180)	8	39	21	37
CE2 (N = 174)	6	28	13	30
CM1 (N = 188)	5	15	8	24
CM2 (N = 176)	4	12	6	20

Lecture : en CP, 54% des mots longs peu fréquents qui ont été lus, n'ont pas été bien lus.

En CP, on constate que les mots longs peu fréquents (ou rares) ont entraîné le plus de difficultés. En effet, on remarque qu'il y a 54% d'erreurs sur cette catégorie de mots contre 46% pour les mots longs fréquents ($\chi^2 = -2.868$; ddl = 1 ; $p < .05$), 42% pour les mots courts peu fréquents ($\chi^2 = -4.781$; ddl = 1 ; $p < .01$) et 15% pour les mots courts fréquents ($\chi^2 = -7.253$; ddl = 1 ; $p < .01$). En revanche, la fréquence ne semble pas jouer un rôle sur les mots courts car on ne constate pas de différence significative entre le nombre d'erreurs commises sur les mots courts fréquents et mots courts rares ($\chi^2 = -2.473$; ddl = 1 ; *NS.*). On ne constate pas non plus de différence significative entre les mots courts rares et les mots longs fréquents ($\chi^2 = -1.912$; ddl = 1 ; *NS.*). La longueur et la fréquence influencent donc la lecture en CP.

En CE1, on note un effet de longueur et de fréquence mais pas comme on pouvait s'y attendre. En effet, on remarque que les mots les plus difficiles à lire ne sont pas les mots longs et peu fréquents mais les mots courts et peu fréquents (39% d'erreurs contre 37%). A ce niveau, la fréquence influence davantage que la longueur ($\chi^2 = 199.450$; ddl = 3 ; $p < .01$).

En CE2, il apparaît que les mots longs peu fréquents sont ceux qui ont entraîné le plus de difficultés de lecture puisqu'on compte 30% contre 28% sur les mots courts peu fréquents, 13% d'erreurs sur les mots longs fréquents et les mots courts fréquents. Ainsi, l'effet de fréquence et longueur se retrouve en CE2 ($\chi^2 = 312.629$; ddl = 3 ; $p < .01$).

Cette tendance se poursuit en CM1 puisqu'encore 24% des mots longs peu fréquents ont entraîné des erreurs contre 8% des mots longs fréquents, 15% des mots courts peu fréquents et 5% des mots courts rares ($\chi^2 = 317.257$; ddl = 3 ; $p < .01$).

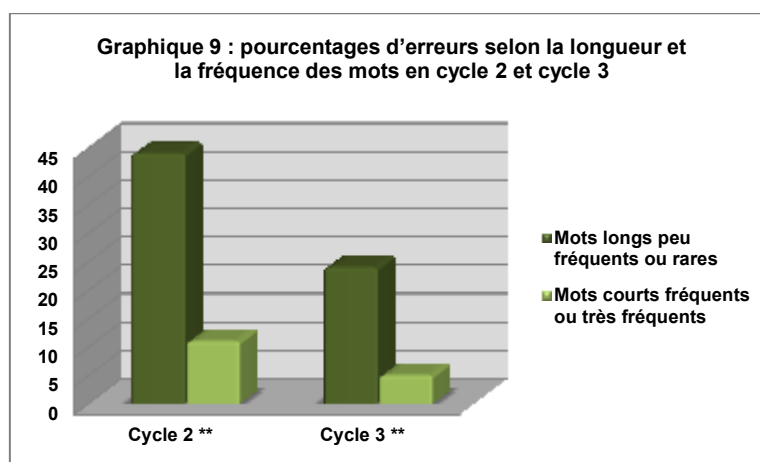
Enfin, en CM2, le constat ne change pas puisque 20% des mots longs peu fréquents ont entraîné des difficultés de lecture contre 6% des mots longs fréquents, 12% des mots courts peu fréquents et 4% des mots courts fréquents.

En **résumé**, notre hypothèse selon laquelle les mots longs peu fréquents (ou rares) sont plus difficiles à lire que les mots courts fréquents (ou très fréquents) se vérifie donc ici et ce quel que soit le niveau scolaire considéré.

Par ailleurs, il semble que les mots longs peu fréquents entraînent des difficultés de traitement plus importantes au début de l'apprentissage de la lecture (54% d'erreurs) plutôt qu'en fin (20% d'erreurs). Cet effet devrait toutefois diminuer de façon significative au fur et à mesure de cette acquisition.

L'effet de longueur et de fréquence diminue-t-il avec l'avancée dans la scolarisation ?

D'une façon générale, il semble que l'effet de longueur et de fréquence d'un mot diminue d'un cycle de scolarisation à l'autre.

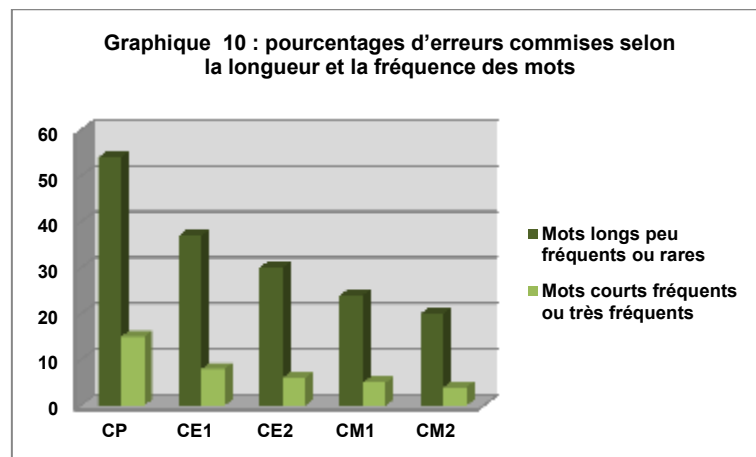


En effet, alors qu'en cycle 2, 44% des mots longs peu fréquents (ou rares) ont été mal lus, ils ne sont plus que 24% en cycle 3 (-20 points ; $z = -9.206, p < .01$). Dans de moindres proportions, on constate que le nombre d'erreurs sur les mots courts fréquents (ou très fréquents) diminue lui aussi entre ces deux périodes (-6 points ; $z = -10.791, p < .01$).

Le nombre d'erreurs sur les mots longs peu fréquents diminue donc, de façon significative, entre le cycle 2 et le cycle 3. Cette forte baisse laisse penser qu'il y aurait une rupture dans la progression des élèves, synonyme d'un apprentissage discontinu de la lecture.

L'analyse de cette diminution entre les niveaux scolaires va nous permettre de préciser ce constat.

Si on regarde dans le détail l'évolution des erreurs liées à la longueur et la fréquence des mots, on remarque que cette diminution n'est pas, la plupart du temps, significative d'une classe à l'autre.



En effet, bien que le nombre d'erreurs commises sur les mots longs peu fréquents (ou rares) diminue de 17 points entre le CP et le CE1 ($\chi^2(4) = 2.204$, *NS*) et de 7 points entre le CE1 et le CE2 ($\chi^2(4) = 2.052$, *NS*), il n'y a qu'entre le CE2 et le CM1 que la diminution est significative (- 6 points ; $\chi^2(4) = 3.435$, $p < .01$). Entre le CM1 et le CM2, la baisse constatée de 4 points n'est pas significative au niveau statistique ($\chi^2(4) = 2.218$, *NS*).

Bien que cette diminution soit constante entre le CP et le CM2, elle n'est pas pour autant régulière. Ceci nous amène donc à penser à une rupture effective entre les CE2 et les CM1 qui confirmerait la mise en place d'un stade alphabétique au cours duquel l'enfant aurait recours à la médiation phonologique. Ces résultats semblent donc être en concordance avec l'idée d'un apprentissage discontinu de la lecture, comme le suggèrent les modèles développementaux.

En **conclusion**, il semblerait nos prédictions de départ soient vérifiées. D'une part, cette analyse nous permet de vérifier que plus un mot contient de lettres et plus sa fréquence est basse, plus il est difficile à lire, et ce quel que soit le niveau scolaire considéré : les élèves ont commis significativement plus d'erreurs lors de la lecture de mots longs peu fréquents (ou rares) que de mots courts fréquents (ou très fréquents). La non maîtrise de la procédure

alphabétique amènerait donc les élèves à mettre plus de temps pour traiter les mots longs à basse fréquence, ce qui entraînerait davantage de risques de commettre une erreur de lecture. Le fait que la longueur et la fréquence d'un mot aient encore une influence à la fin du cycle 3 laisse supposer que les élèves n'aient pas encore tout à fait automatisé cette procédure et ne soient pas encore totalement passés au stade suivant. D'autre part, nous avons constaté une progression des élèves entre le CP et le CM2. En effet, bien qu'il n'y ait pas de différence significative d'un niveau à l'autre, excepté entre les CE2 et CM1, on constate que le nombre d'erreurs commises diminue systématiquement d'une classe à l'autre. Cette diminution semble confirmer notre supposition de départ selon laquelle nous avons envisagé une baisse de ce type d'erreur avec l'apprentissage du code. Néanmoins, cette progression semble irrégulière puisqu'il n'y a qu'entre le CE2 et le CM1 que cette baisse est significative. Ce constat laisse penser qu'il existerait une rupture entre les élèves de cycle 2 et ceux de cycle 3, synonyme d'un apprentissage discontinu de la lecture. Il se pourrait que cette rupture soit un signe de changement de stratégie.

Dans cette optique, nous avons procédé à une analyse qualitative des erreurs. Ainsi, en plaçant l'élève au centre de notre recherche, nous avons cherché à savoir, à partir des erreurs commises, comment l'enfant procède face à cette difficulté. Autrement dit, nous nous sommes demandée quelles sont les stratégies adoptées.

II. Les processus de lecture selon la longueur et la fréquence d'un mot

S'il existe un effet de longueur et de fréquence, alors nous nous attendons à ce que les erreurs de décodage soient plus nombreuses lors de la lecture de mots longs peu fréquents (ou rares) que de mots courts fréquents (ou très fréquents). En effet, s'il est vrai que les enfants passent par une étape dite « alphabétique », on devrait alors trouver davantage d'erreurs de décodage sur les mots longs peu fréquents que sur les mots courts fréquents au début de l'apprentissage de la lecture. La maîtrise du code grapho-phonétique devrait réduire les écarts de ce type d'erreur entre les deux catégories de mots.

Si on considère que les enfants passent ensuite par un stade orthographique, alors nous supposons que l'effet de longueur et de fréquence s'amoinde avec la mise en place de la voie

lexicale et nous devrions voir apparaître ce type d'erreur à un niveau plus avancé de l'apprentissage.

Par conséquent, s'il existe un effet de longueur et de fréquence, la maîtrise de ces deux procédures (alphabétique et orthographique) devrait faire diminuer les écarts entre les erreurs de décodage commises sur les mots longs rares et les mots courts fréquents et augmenter les erreurs de type lexical.

Les élèves de CP (N = 143)

On remarque qu'en CP, **67% des erreurs commises sur les mots longs rares sont des erreurs de décodage**. Ce type d'erreur se traduit essentiellement par des erreurs de conversion grapho-phonémique, soit d'un graphe contextuel soit d'un graphème composé. On remarque que ces décodages sont aussi souvent incomplets. En effet, si on regarde le mot « fourchu », sur lequel 45% des élèves qui l'ont lu ont commis une erreur, on constate que parmi eux, 65% ont commis une erreur de décodage. Il a été lu [fo], [foʃy], [foʀʃy] ou encore [fʏʀʃy]. Ces erreurs semblent être dues à la difficulté de traitement du graphème composé « ou ». On observe aussi une grande quantité de décodage partiels tels que [fu], [fʏʀ] ou encore [fʏʀʃ]. Il semble que les élèves, après avoir essayé de lire la première syllabe, se soient ensuite arrêtés, probablement débordés par le nombre d'informations à traiter. En revanche, pour le mot « gamine » qui a posé des difficultés à 41% des élèves de CP, il semble que c'est la lecture du graphème « g » qui a été difficile puisqu'il a été lu essentiellement [ʒaminə].

Parmi les erreurs les plus fréquentes lors de la lecture des mots longs rares en CP, se trouvent aussi les **abstentions qui représentent 17% des incorrections**. Comme pour les erreurs de décodage, ce sont les mots « fourchu » et « gamine » qui ont été le plus souvent sautés par les élèves. Ainsi, 27% des élèves de CP n'ont pas lu le mot « gamine » et 22% n'ont pas lu le mot « fourchu ». Il est possible que la longueur écrite et la faible fréquence de ces mots aient perturbé les élèves et que ces derniers aient préféré ne rien lire plutôt que de se tromper.

Le troisième type d'erreur répertorié sur les mots longs peu fréquents en CP est d'ordre **lexical**. En effet, certains élèves ont énoncé un mot porteur de sens contrairement aux erreurs de décodage entraînant une production d'un non-mot. Ainsi, **8% des élèves** qui ont commis une erreur semblent avoir essayé de remplacer un mot par un autre qui leur est plus familier. Par exemple, parmi les élèves qui ont lu le mot « fourchu », la plupart a prononcé [fʏʀʃɛt] (5

élèves sur 8). On remarque que le dessin d'une fourchette est situé non loin du mot « fourchu ». Il est donc possible que le dessin les ait influencés. D'autres, ont lu [fURmi], [flœR] ou encore [fURYRƏ]. On note que pour toutes ces productions lexicales, les élèves ont effectué un décodage partiel du mot pour ensuite prononcer un mot qui leur est certainement plus familier.

Le dernier type d'erreur concerne **les verbalisations de mots mais nous ignorons le processus de départ** (décodage ou intention lexicale). Elles **représentent en moyenne 7%** des erreurs. C'est le cas, par exemple du mot « rixes » lu par 7 élèves. Parmi eux, quatre ont prononcé [RiSk]. Pour cette erreur il est difficile de savoir si les élèves ont voulu dire « risque » ou s'ils ont inversé le son [kS].

Ainsi, cette analyse semble aller dans le sens de notre hypothèse puisque la prédominance des erreurs dues à un mauvais décodage lors de la lecture de mots longs rares laisse supposer la mise en place d'une procédure alphabétique qui expliquerait un possible effet de longueur et de fréquence. Si on considère que les erreurs issues de processus non identifiés sont en réalité des intentions lexicales, alors cette catégorie représente 15% des erreurs sur les mots longs rares, ce qui reste inférieur aux erreurs de décodage (67%) et aux abstentions (17%). La longueur et la fréquence semblent donc avoir une influence. Toutefois, cette hypothèse doit être confirmée par une analyse des erreurs sur les mots courts fréquents. En effet, si la longueur et la fréquence influencent la lecture, alors nous devrions trouver moins d'erreurs de décodage sur les mots courts fréquents (ou très fréquents).

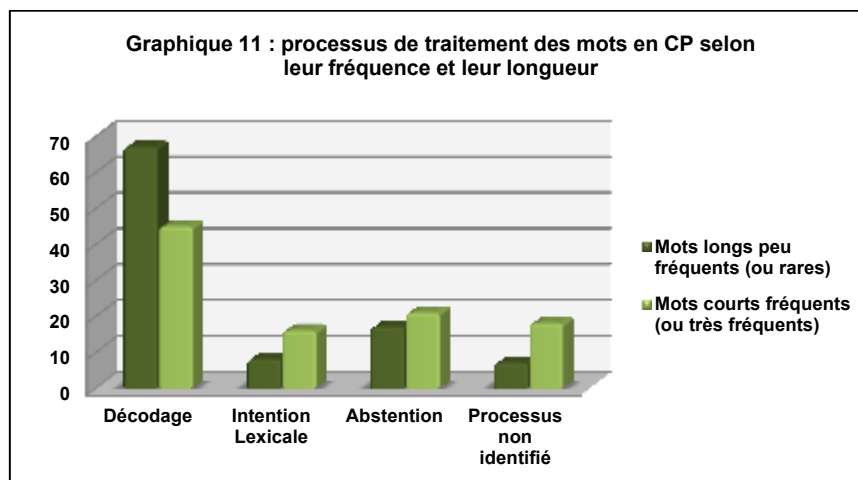
En effet, il apparaît que pour les élèves de CP, les erreurs de décodage sont plus nombreuses sur les mots longs rares que sur les mots courts fréquents : 67% contre 45% pour les mots courts fréquents. Ce premier constat semble donc confirmer notre supposition.

En ce qui concerne les mots qui n'ont pas été lus, ils représentent 17% des erreurs sur les mots longs rares contre 21% des erreurs sur les mots courts fréquents. Ils sont donc plus nombreux pour cette catégorie.

Les intentions lexicales sont elles aussi plus nombreuses sur les mots courts fréquents puisqu'elles sont de 16% contre 8% sur les mots longs rares.

C'est aussi le cas des erreurs issues de processus non identifiés qui représentent 18% des erreurs commises sur les mots courts fréquents contre 7% sur les mots longs rares.

En **résumé**, il apparaît qu'en CP, les mots longs rares entraînent non seulement principalement des erreurs de décodage et en entraînent davantage que les mots courts fréquents. Ces résultats semblent donc aller dans le sens d'un effet de longueur et fréquence d'un mot. Il apparaît donc qu'à ce stade de l'apprentissage de la lecture, l'enfant procède principalement à un décodage lorsqu'il rencontre un mot long peu fréquent plutôt qu'un mot court fréquent. Les mots contenant plus de lettres et moins fréquents sont donc davantage traités par une stratégie d'assemblage.



D'autre part, le fait que les verbalisations de mots soient présentes à ce niveau, laisse supposer une existence de la stratégie orthographique pour le traitement de certains mots. **Il apparaît donc que la longueur et la fréquence d'un mot ont une influence sur la stratégie de décodage pour les élèves de CP.**

Les élèves de CE1 (N = 180)

En CE1, comme au niveau précédent, **le type d'erreur dominant sur les mots longs rares (ou peu fréquents) concerne le décodage (64%)**. Parmi les mots qui ont entraîné le plus d'erreurs de décodage, on trouve le mot « féeries », qui a entraîné 78% d'erreur de décodage : la plupart d'entre eux a lu [fəri] (82%). On observe aussi que 77% des élèves se sont trompés lors de la lecture du mot « paille » et ont commis une erreur de décodage. Ils ont lu essentiellement [pjɛjə], [pjɛlə] ou encore [pɛjə]. Ces derniers semblent avoir commencé leur lecture en associant les deux premières lettres (selon le type de syllabe le plus

fréquent en français : consonne-voyelle), puis auraient continué leur décodage de la même façon. Cette procédure les aurait donc amenés à effectuer une resyllabation du mot.

Le second type d'erreur concerne **les verbalisations de mots issues de processus non identifiés qui représentent en moyenne 21%** des erreurs commises sur cette catégorie de mots. C'est le cas par exemple du mot « rixes » puisque 54% des élèves qui ont mal lu le mot « rixes » ont prononcé [ʀisk]. Dans ce cas, il est difficile de savoir si cette erreur est due à une erreur de conversion grapho-phonémique ou si l'élève a voulu dire « risque ».

Les abstentions représentent en moyenne 9% des erreurs commises et ce sont surtout les mots « dégelée », « pampres » et « trille » qui n'ont pas été lus.

Enfin, **les intentions lexicales représentent 5%** des erreurs commises sur les mots longs rares et concernent essentiellement (12%) le mot « hirondeau » lu [iʀɔ̃dɛ]. Le décodage du début et milieu du mot a pu amener les élèves à vouloir dire « hirondelle ». De plus, comme pour le mot « fourchu », le dessin d'une hirondelle est situé juste à côté. Il est donc possible que la présence du dessin ait eu une influence sur la lecture. A ce niveau scolaire, les élèves effectuent donc peu de lexicalisation.

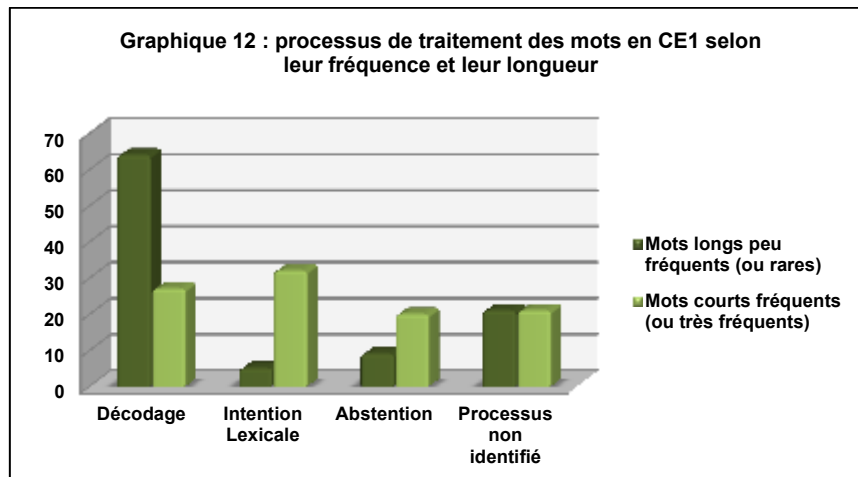
Par conséquent, l'essentiel des erreurs commises lors de la lecture de mots longs rares étant dues à un mauvais décodage, il semble donc que l'effet de longueur et de fréquence constaté en CP soit encore présent à ce niveau scolaire. Si on considère les verbalisations de mots issues de processus non identifiés comme étant des intentions lexicales, alors cette nouvelle catégorie représente 26% en moyenne des erreurs commises sur les mots longs rares ; ce qui reste inférieur aux erreurs de décodage (64%). Comme précédemment, notre hypothèse doit se confirmer par une analyse des erreurs sur les mots courts fréquents.

En ce qui concerne les erreurs de lecture commises sur les mots courts fréquents, on remarque que les erreurs de décodage sont moins nombreuses pour cette catégorie de mots puisqu'elles sont de 27% (contre 64% pour les mots longs rares). De plus, on remarque qu'elles ne sont plus le type d'erreur principalement commis lors de la lecture des mots courts fréquents.

En effet, le premier type d'erreur commis sur les mots courts fréquents est l'intention lexicale qui concerne 32% des erreurs. Si on rajoute les verbalisations de mots dont nous ignorons le processus de lecture, alors cette catégorie s'élève à 53%. Ce qui signifie que plus de la moitié des élèves ont verbalisé un mot.

Les non-lectures de mots courts fréquents sont de 20%.

En **résumé**, il apparaît qu'en CE1, les erreurs de décodage lors de la lecture de mots longs rares sont encore le type d'erreur principalement commis. De plus, le fait que ce type d'erreur soit moins important lors de la lecture de mots courts fréquents semble ici aussi valider notre supposition quant à un effet de longueur et de fréquence.



Autrement dit, à ce stade de l'apprentissage, les enfants utiliseraient encore essentiellement la voie alphabétique pour traiter les mots longs peu fréquents que les mots courts fréquents. Si on regarde les verbalisations de mots, on note qu'elles sont de 26% pour les mots longs peu fréquents et 53% pour les mots courts fréquents. Ainsi, ce constat semble confirmer **l'existence constatée en CP de la procédure lexicale** pour lire certains mots.

Les élèves de CE2 (N = 174)

En CE2, **les erreurs de décodage continuent d'être le type d'erreur principalement commis** lors de la lecture de mots longs peu fréquents (52%). On retrouve les erreurs commises sur les mots « piaille » et « fêtes ». Concernant le mot « piaille » qui a causé 40% d'erreurs, il a été lu essentiellement [pɛjə], [piɛjə] ou encore [piɛlə]. Ce type d'erreur semble nous conforter dans la mise en place de la voie phonologique qui ne serait pas encore tout à fait automatisée, ce qui expliquerait ce décodage graphème par graphème. Le mot « fêtes » a induit 44% d'élèves en erreur, ces derniers l'ont lu principalement [fɛri] (68%) ou [fɛrie] (22%).

Le second type d'erreur est, comme en CE1, les **verbalisations de mots dont nous ignorons le processus qui représentent 35%** des erreurs. La plus courante porte sur le mot « poisons » essentiellement lu [pwasɔ̃]. Il est possible que les élèves aient voulu dire « poisson » car le co-texte sémantique dans lequel ce mot apparaît invite à lire « poisson » mais il n'est pas non plus impossible que les élèves se soient trompés lors de la lecture du « s » qui entre deux voyelles, doit se lire [z].

Les **intentions lexicales, quant à elles, représentent 8%** des erreurs. Comme précédemment, elles portent essentiellement sur le mot « hirondeau » lu [iʁɔ̃dɛl].

Enfin, **les abstentions qui représentent 6%** des erreurs commises sur les mots longs rares concernent encore essentiellement le mot « pampres ».

Ainsi, la prédominance des erreurs de décodage lors de la lecture de mots longs rares semble donc indiquer que l'effet de longueur et de fréquence est encore présent en CE2. Si on regroupe les verbalisations de mots, qu'elles soient intentionnelles ou non, alors celles-ci représentent maintenant 43% des erreurs, ce qui reste encore inférieur aux erreurs de décodage (52%). Toutefois, comme pour les niveaux précédents, ce constat doit se confirmer par une analyse des mots courts fréquents.

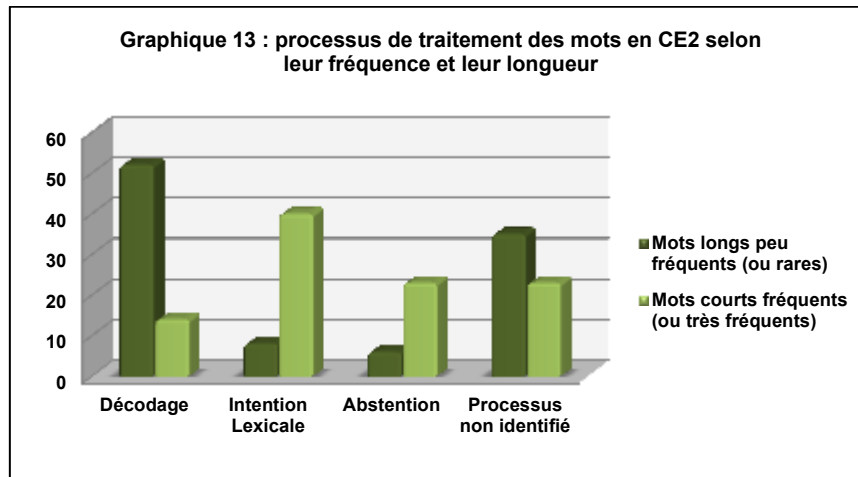
Si on regarde les différents types d'erreurs commis sur les mots courts, on remarque que les erreurs de décodage sont moins importantes que sur les mots longs rares (14% contre 52%) et qu'elles ne sont pas le type d'erreur principal.

En effet, comme au niveau précédent, ce sont les intentions lexicales qui sont le premier type d'erreur puisqu'elles représentent 40% des erreurs commises sur les mots courts fréquents. Si on ajoute les autres verbalisations de mots alors cette catégorie passe à 63%.

Les abstentions représentent 23% des erreurs.

Et les erreurs de décodage sur les mots courts fréquents ne sont plus que de 14%.

En **résumé**, il apparaît qu'en CE2, les mots longs rares entraînent encore essentiellement des erreurs de décodage, ce qui laisse penser que la longueur et la fréquence ont encore une influence lors de la lecture de mots. Ce constat semble se confirmer par le fait que les erreurs de ce type sont moins importantes lors de la lecture de mots courts fréquents.



D'autre part, la présence de verbalisations de mots, qu'elles soient intentionnelles ou non, nous conforte dans l'idée qu'il y aurait une **mise en place progressive de la procédure orthographique**.

Les élèves de CM1 (N = 188)

En CM1, le type d'erreur principalement commis lors de la lecture de mots longs peu fréquents change et concerne les **verbalisations de mots dont nous ignorons le processus**. Cette catégorie représente 46% des erreurs. On retrouve le mot « poison » lu [pwasɔ̃] dont la lecture a certainement pu être là aussi influencée par le co-texte sémantique.

Les erreurs de décodage **représentent encore 40% des erreurs** et la plupart concernent encore le mot « féeries » lu majoritairement [fɛri].

Les **intentions lexicales sont de 9%**. On remarque qu'elles ne portent plus sur le mot « hirondeau » mais sur le mot « décochée ». Ce dernier a été prononcé [dekɔʃe]. Il se pourrait qu'ici, comme pour le mot « poisons », ce soit un effet du co-texte sémantique dans lequel ce mot apparaît puisque la phrase dans lequel ce mot apparaît invite à lire « décochée ».

Enfin, **les abstentions ne représentent plus que 5%** des erreurs commises lors de la lecture de mots longs peu fréquents.

Ainsi, il apparaît qu'en CM1, les erreurs de décodage sont encore plus nombreuses lors de la lecture de mots longs rares que de mots courts fréquents. Toutefois, on constate que ce type d'erreur n'est plus celui qui est principalement commis lors de la lecture de mots longs

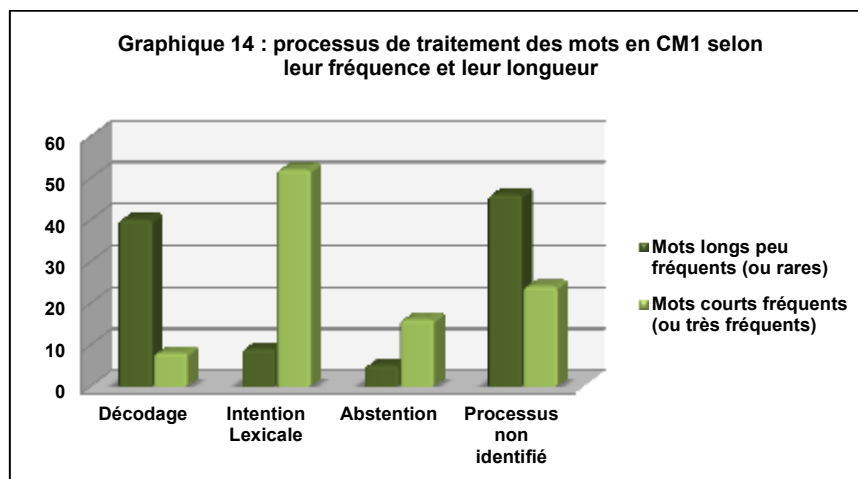
peu fréquents. En effet, les élèves ont en moyenne verbalisé plus de mots que de non mots. Il semble donc que la longueur et la fréquence d'un mot n'auraient plus d'influence, ou du moins, que cet effet s'amointrit. Ce constat doit toutefois se confirmer par une analyse des mots courts fréquents.

Lorsqu'on regarde du côté des mots courts, on remarque que les erreurs de décodage ne représentent plus que 8% des erreurs alors qu'elles sont de 40% pour les mots longs rares.

Ce sont les intentions lexicales qui sont les plus nombreuses puisqu'elles sont de 52%. Les verbalisations de mots dont nous ignorons le processus représentent 24% des erreurs commises sur les mots courts fréquents. Cela signifie que si on totalise les verbalisations de mots, alors cette catégorie s'élève à 76%.

Enfin, on constate que 16% des erreurs sur les mots courts fréquents sont des non-lectures.

En **résumé**, alors que jusqu'à présent lorsqu'on incluait les erreurs issues de processus non identifiés, les erreurs de décodage restaient les plus nombreuses, il apparaît qu'en CM1, ce n'est plus le cas. En effet, si on considère ces erreurs comme étant des intentions lexicales alors cette catégorie devient le processus de lecture le plus important lors de la lecture de mots longs rares. Ce constat laisserait donc penser que l'effet de longueur tend à diminuer.



D'autre part, la prédominance de verbalisations de mots semble confirmer l'établissement d'une procédure orthographique.

Les élèves de CM2 (N = 182)

Enfin, en CM2 comme en CM1, ce sont **les verbalisations issues de processus non identifiés qui sont les plus nombreuses** : elles représentent 44% des erreurs. On retrouve le mot « poison » qui a été lu par 86% des élèves [pwasɔ̃].

Les erreurs de décodage représentent 35% des erreurs commises sur les mots longs rares. On retrouve encore essentiellement le mot « féeries » (32% des élèves) lu dans 70% des cas [fɛri].

Les intentions lexicales représentent 17% des erreurs et portent toujours essentiellement sur le mot « décochée » lu [dekɔʃe].

Les **non-lectures**, bien qu'encore présentes ne représentent plus que **4% des erreurs** commises sur les mots longs rares.

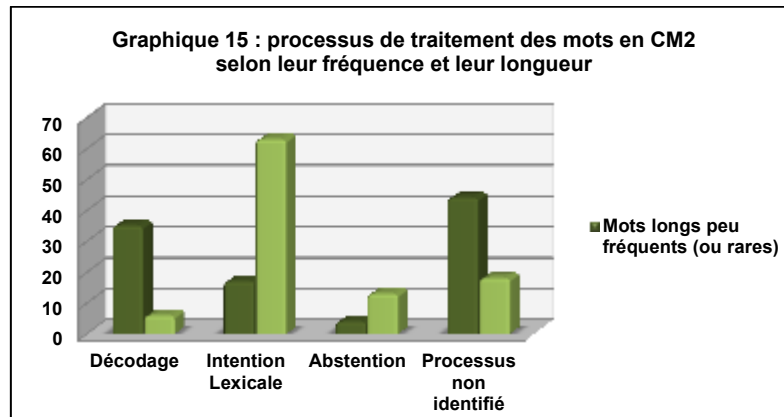
Ainsi, comme en CM1, on constate d'une part que les erreurs de décodage restent plus nombreuses lors de la lecture de mots longs rares que de mots courts fréquents et d'autre part, qu'elles ne sont pas le type d'erreur principal : ce sont les erreurs issues de processus non identifiés. Par conséquent, les élèves verbalisent en moyenne plus de mots que de non-mots ce qui signifierait que la longueur et la fréquence d'un mot tendent à influencer de moins en moins les élèves. Ce constat doit toutefois se confirmer par une analyse des erreurs sur les mots courts fréquents.

Si l'on compare ces taux d'erreurs avec ceux des mots courts, on constate que les erreurs de décodage sont de 6% contre 35% pour les mots longs rares et comme en CM1, ce ne sont plus le type d'erreur principal.

En effet, ce sont les intentions lexicales qui sont les plus importantes (63%) et au total, les verbalisations de mots s'élèvent à 81% des erreurs commises sur les mots courts fréquents.

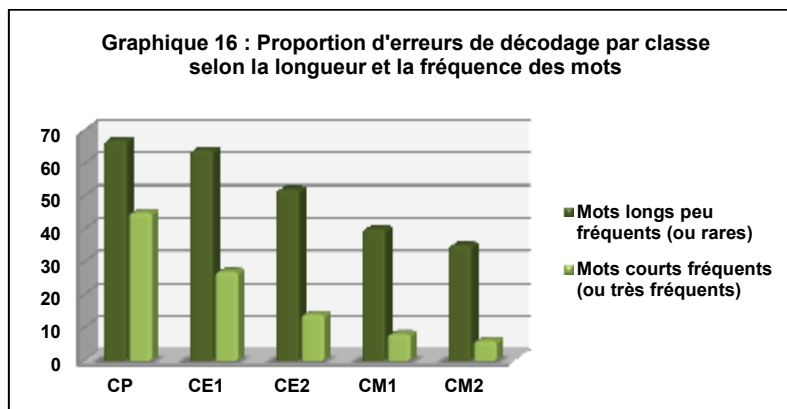
Les abstentions représentent encore 13% des erreurs.

En **résumé**, il apparaît qu'en CM2, comme au niveau précédent, la longueur et la fréquence influencent encore la lecture mais dans de moindres proportions.



Par ailleurs, la prédominance de verbalisations de mots semble donc confirmer **l'établissement d'une procédure orthographique.**

En **conclusion**, d'après les résultats de cette étude, il semblerait que nous puissions confirmer notre hypothèse selon laquelle **face à un mot long rare (ou peu fréquent), les apprentis-lecteurs commettraient davantage d'erreurs dues à un décodage.** En effet, le nombre d'erreurs entraînées par la lecture de mots longs et peu fréquents est systématiquement supérieur à celui des mots courts fréquents (*cf.* graphique 16).



Cette analyse nous a aussi permis de constater que la lecture de mots longs peu fréquents a entraîné des abstentions de lecture et des verbalisations de mots. Ce dernier résultat laisse supposer une existence d'une procédure lexicale lors de la lecture de certains mots. En effet, que ce soit de façon intentionnelle ou pas, **les élèves ont à chaque niveau scolaire, verbalisé des mots porteurs de sens.** Ainsi, il semble donc que **les enfants ne**

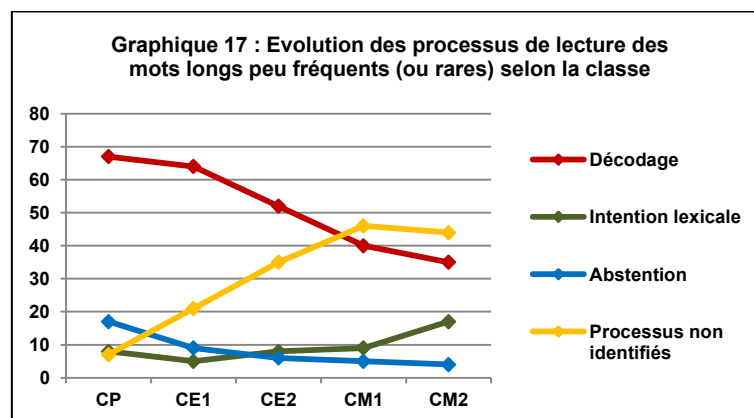
passent pas tous par le même stade en même temps et qu'ils aient recours à plusieurs stratégies pour traiter les mots longs. Le phénomène de discontinuité ne semble donc pas être absolu. Une analyse portant sur l'évolution de ces stratégies employées lors de la lecture des mots devrait nous permettre d'en savoir un peu plus.

III. Evolution des processus de lecture

S'il existe un effet de longueur et de fréquence alors nous nous attendons à ce que les erreurs de décodage sur les mots longs rares diminuent avec la maîtrise de la procédure alphabétique. La mise en place d'un stade orthographique devrait faire apparaître une augmentation des erreurs lexicales. Ainsi, nous supposons que le nombre d'erreurs de décodage diminue et que le nombre d'erreurs lexicales augmente au fur et à mesure de l'avancée dans la scolarisation.

Evolution des processus de lecture des mots longs peu fréquents (ou rares)

S'il existe un effet de longueur et de fréquence lors de l'apprentissage de la lecture, alors on s'attend à ce que les erreurs de décodage diminuent et les erreurs lexicales augmentent d'année en année.



D'une manière générale, on remarque que le taux d'**erreurs dues à un processus de décodage** varie selon le niveau de scolarisation ($\chi^2(4) = 78.337 ; p < .01$) (cf. graphique 17). En effet, alors qu'en CP ce type d'erreur est de 67%, il diminue significativement de 3 points au niveau suivant ($\chi^2(1) = -2.847, p < .05$). En CE2, ce taux continue de diminuer et

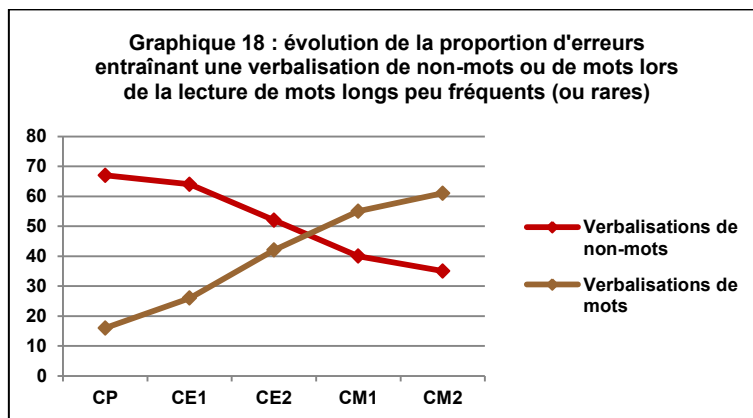
représente 52% des erreurs ($\chi^2(1) = 3.048, p < .05$). Au niveau suivant, il est de 40% ($\chi^2(1) = 3.519, p < .01$). Enfin on remarque qu'en fin d'école primaire, cette proportion baisse légèrement mais de façon non significative puisqu'elle représente 35% des erreurs commises lors de la lecture de mots longs ($\chi^2(1) = 1.304, NS$). Ce constat semble donc confirmer le fait que les mots longs peu fréquents sont de moins en moins traités par décodage au fur et à mesure de l'avancée dans la scolarisation, et par conséquent, de la maîtrise de la lecture.

En ce qui concerne les **intentions lexicales**, on constate qu'elles varient de façon significative entre les classes ($\chi^2(4) = 61.838, p < .01$) mais généralement pas d'une classe à l'autre. Elles sont de 8% en CP. En CE1, ce taux varie peu puisqu'il est de 5% ($\chi^2(1) = -.880, NS$). En CE2, la variation de +3 points est là aussi non significative ($\chi^2(1) = -2.002, NS$). En CM1, c'est à peu près la même tendance puisque ce type d'erreur est de 9% ($\chi^2(1) = -.836, NS$). En revanche en CM2, on remarque que les intentions lexicales sont significativement plus nombreuses puisqu'elles sont de 17% ($\chi^2(1) = -3.631, p < .01$). Ce constat semble donc aller avec l'idée d'une mise en place progressive d'une procédure orthographique puisque les erreurs de ce type sont plus nombreuses en fin d'école primaire. Par ailleurs, nous avons vu que pour certaines verbalisations de mots, il nous est difficile d'identifier le processus de départ. Si on regroupe ces deux catégories comme étant des **verbalisations de mots** alors on constate que la proportion d'erreurs de ce type augmente d'année en année ($\chi^2(1) = 240.956, p < .01$). En effet, alors qu'en CP 16% des erreurs commises sur les mots longs rares sont des verbalisations de mots, elles sont 26% en CE1 ($\chi^2(1) = -3.993, p < .01$). En CE2, cette proportion passe à 42% ($\chi^2(1) = -4.895, p < .01$). En CM1, ce sont plus de la moitié (55%) des erreurs qui sont des verbalisations de mots ($\chi^2(1) = -3.319, p < .01$). Enfin, en CM2, les verbalisations de mots augmentent (+6 points) mais de façon non significative ($\chi^2(1) = -1.371, NS$). Ces nouveaux résultats semblent donc confirmer notre constat, à savoir qu'il existerait une procédure orthographique qui se mettrait en place progressivement entre le CP et le CM2.

Enfin, pour ce qui est des **abstentions**, on ne constate pas de différence significative excepté entre les CE1 et CM2 ($\chi^2(1) = -2.926, p < .05$) en ce qui concerne les mots longs peu fréquents.

En **résumé**, si on considère les verbalisations de mots, qu'elles soient intentionnelles ou pas, alors ces résultats semblent aller dans le sens de nos suppositions puisqu'on assiste à une

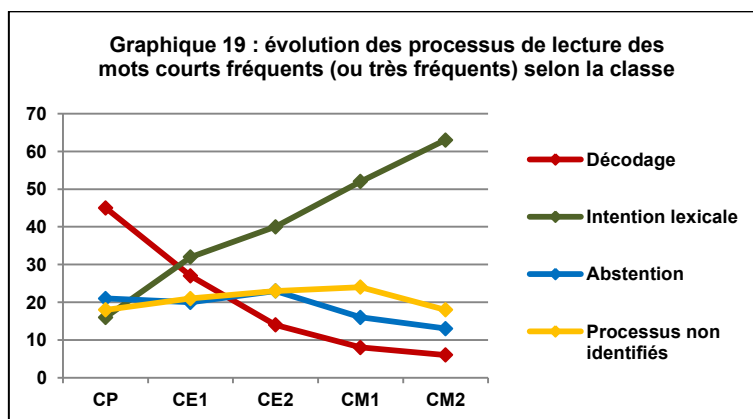
diminution des erreurs dues à un processus de décodage et une augmentation des verbalisations de mots lors de la lecture de mots longs rares (cf. graphique 18).



Comme pour les analyses précédentes, celle-ci doit se confirmer par une analyse des processus de lecture des mots courts fréquents.

Evolution des processus de lecture des mots courts fréquents (ou très fréquents)

En effet, s'il existe un effet de la longueur et de la fréquence d'un mot, alors on devrait certes, voir apparaître une diminution des erreurs de décodage et une augmentation des erreurs lexicales, mais dans de moindres proportions que pour les mots longs peu fréquents (ou rares). Ceci correspondant simplement à la maîtrise du code et à la mise en place d'une voie lexicale.



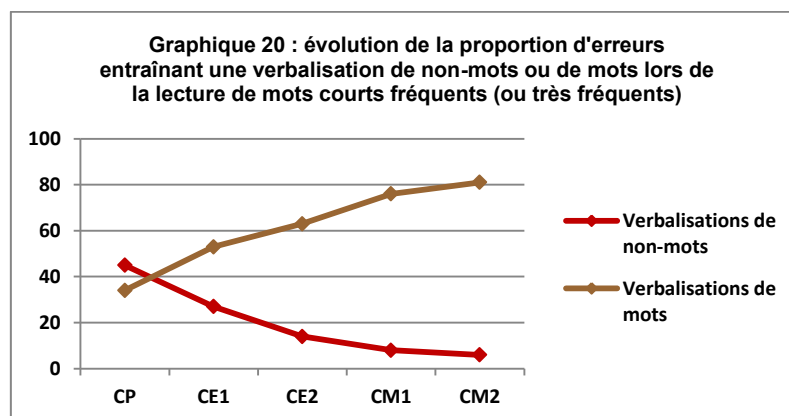
Ainsi, d'une façon générale, on observe que les **erreurs de décodage** diminuent avec le niveau scolaire ($\chi^2(4) = 188.070 ; p < .01$). Alors que la proportion d'erreurs de ce type est

de 45% en CP, elle passe à 27% en CE1 ($\chi^2(1) = 4.090 ; p < .01$). Au niveau d'après, les erreurs de décodage lors de la lecture de mots courts fréquents diminuent de 13 points et sont de 14% ($\chi^2(1) = 4.691 ; p < .01$). En CM1, cette proportion diminue de 6 points et passe à 8%. Toutefois, cette baisse n'est pas significative ($\chi^2(1) = 1.993 ; NS$), tout comme en CM2 où ce type d'erreur ne représente plus que 6% ($\chi^2(1) = 1.385 ; NS$).

Si on regarde du côté des **intentions lexicales**, on remarque que leur nombre augmente d'année en année ($\chi^2(4) = 205.448, p < .01$). Alors qu'elles sont de 16% en CP, elles doublent en CE1 ($\chi^2(1) = -4.598 ; p < .01$). En CE2, cette proportion augmente de 8 points : elles représentent 40% des erreurs commises sur les mots courts fréquents. Néanmoins, cette hausse n'est pas significative ($\chi^2(1) = -2.253 ; NS$). En revanche, en CM1, l'augmentation est plus forte (+12 points) et ce type d'erreur représente 52% ($\chi^2(1) = -3.708 ; p < .01$). Enfin, en CM2, les erreurs lexicales augmentent encore de 10 points et s'élèvent à 63% ($\chi^2(1) = -3.007 ; p < .05$).

Enfin, les **abstentions** ne varient pas de façon significative d'une classe à l'autre ($\chi^2(4) = 8.475 ; NS$).

En **résumé**, ces résultats semblent aller dans le sens de notre hypothèse puisque le nombre d'erreurs de décodage sur les mots longs fréquents est moins important que sur les mots longs rares et qu'il diminue entre le CP et le CM2, ce qui correspondrait à la mise en place progressive du stade alphabétique. Parallèlement, on remarque que le nombre d'erreur lexicales, qui est plus important pour cette catégorie de mots, augmente, ce qui semble confirmer l'existence d'un stade orthographique. Si on considère les verbalisations de mots, qu'elles soient intentionnelles ou non, alors on remarque que ce type d'erreur augmente entre le CP et le CM2 (*cf.* graphique 20).



CONCLUSION

L'objectif de cette recherche était de savoir s'il existe un effet de longueur et de fréquence des mots lors de l'apprentissage de la lecture et de quelle façon il nous renseigne sur le déroulement de cet apprentissage.

A cet effet, nous sommes partie de notre hypothèse théorique développementale selon laquelle l'apprentissage de la lecture se déroulerait de façon discontinue, c'est-à-dire que les enfants traverseraient tous un stade logographique, alphabétique puis orthographique. Ainsi nous avons supposé que la longueur et la fréquence ont un effet au début de l'apprentissage de la lecture. La maîtrise du code devrait faire diminuer cet effet d'année en année. Autrement dit, si l'apprentissage de la lecture se déroule bien de façon discontinue alors nous pensons que le nombre d'erreurs sur les mots longs peu fréquents est plus important que sur les mots fréquents. Cet effet devait s'amoinrir dans le temps.

En effet, il apparaît que d'un point de vue quantitatif, les mots contenant plus de lettres et étant moins fréquents, entraînent davantage d'erreurs de lecture que ceux en contenant moins. Autrement dit, les élèves qu'ils soient scolarisés en cycle 2 ou cycle 3, lisent mieux les mots courts fréquents que les mots longs rares. Ces résultats recourent donc ceux trouvés par Mousty P., Leybeart J. et Alegria J, lors de leur recherche menée en 1999.

Cet effet de longueur semble confirmer la mise en place d'un stade alphabétique au cours duquel les mots seraient traités par décodage. Ainsi, lorsqu'un élève utilise la stratégie d'assemblage, l'information phonologique est maintenue dans une sorte de mémoire tampon (appelée « *buffer* » par les psycholinguistes), le temps que les traitements se réalisent. Nous supposons par conséquent, que le caractère transitoire de ce stockage entraîne une dégradation rapide de l'information en fonction du temps. Plus un mot est long et peu fréquent et plus l'implication de la mémoire de travail doit être grande. En d'autres termes, plus un mot contient de lettres, plus il est long à traiter ce qui entraîne davantage de risques de commettre une erreur de lecture et notamment d'assemblage. Notre hypothèse se confirme aussi dans la mesure où le nombre d'erreurs diminue d'année en année.

Par conséquent s'il existe un effet de longueur et de fréquence, nous nous attendions à ce que les mots longs peu fréquents entraînent davantage d'erreurs de décodage au début de l'apprentissage de la lecture. De plus, si les élèves traversent tous un stade orthographique, nous nous attendions à voir augmenter les erreurs lexicales.

L'analyse qualitative menée nous permet de confirmer un effet de longueur et de fréquence puisque que systématiquement, les erreurs de décodage sont plus nombreuses lors de la lecture de mots longs peu fréquents que de mots courts fréquents. Ce type d'erreur diminuant d'année en année laisse penser que la maîtrise du code fait diminuer cet effet.

Par ailleurs, nous avons constaté que dès le début du CP, les élèves commettent des erreurs lexicales lors de la lecture de certains mots longs et peu fréquents. L'analyse de l'évolution des erreurs nous permet de voir que, pendant que les erreurs de décodage diminuent, celles du type lexical augmentent. Cela signifie que durant l'automatisation de la procédure par médiation phonologique, une procédure lexicale se développe. Dans les modèles développementaux, l'enfant passe d'un stade à un autre si et seulement si les compétences propres à l'étape antérieure sont maîtrisées. Ainsi, d'après l'analyse des erreurs commises sur les mots longs peu fréquents nos élèves seraient dans le stade alphabétique et orthographique. Situation impossible dans le cadre des modèles génétiques. Par conséquent, l'enfant ne traverserait pas d'étapes mais mettrait plutôt en place plusieurs procédures de traitement (ici, alphabétiques et orthographiques).

En outre, parmi les erreurs lexicales, nous avons remarqué que la plupart d'entre elles se rapportent au contexte et au co-texte (images) et ce dès le CP. Il semble donc que l'enfant qui apprend à lire cherche aussi à donner du sens. Or les modèles développementaux n'abordent pas le versant compréhension en lecture. Nous développerons ce point dans les chapitres suivants.

Enfin, nous avons aussi remarqué que certains mots longs peu fréquents avaient été souvent sautés au début de l'apprentissage de la lecture et que certains mots courts fréquents n'avaient pas été lus à un niveau plus intermédiaire de l'apprentissage. Nous pensons que ce phénomène peut être expliqué de deux façons. D'une part, nous supposons qu'au début de l'apprentissage de la lecture, l'enfant saute des mots parce qu'il ne maîtrise pas suffisamment les correspondances graphèmes-phonèmes lui permettant de lire (Malmquist, 1973). D'autre part, l'enfant plus avancé dans l'apprentissage de la lecture sauterait des mots non plus à cause d'une incapacité à décoder mais plutôt à cause d'une lecture plus fluide et par conséquent plus rapide, qui ferait que certains mots (les plus courts notamment) ne seraient pas directement lus (Foucambert, 1996).

Pour conclure, nous pouvons dire qu'il existe un effet de longueur et de fréquence qui semble aller dans le sens d'un apprentissage discontinu de la lecture. Néanmoins, nos résultats font apparaître que cette discontinuité ne serait pas absolue et que les élèves auraient recours à plusieurs stratégies pour lire les mots.

D'autre part, il convient de préciser que l'effet de longueur et de fréquence d'un mot en lecture est très difficile à étudier car il est fortement corrélé avec d'autres variables telle que la structure graphémique. En effet, les mots complexes phonologiquement tendent à être plus rares. C'est pourquoi dans le chapitre qui suit nous allons étudier l'influence de la structure graphémique.

Chapitre 10



L'INFLUENCE DE LA STRUCTURE GRAPHÉMIQUE D'UN MOT

Plan du chapitre

INTRODUCTION	196
PREDICTIONS	197
METHODOLOGIE DE TRAITEMENT DES DONNEES.....	198
I. Est-ce que la structure graphémique d'un mot fait difficulté ?	200
<i>Le nombre d'erreurs varie-t-il avec la structure graphémique d'un mot ?.....</i>	<i>200</i>
<i>L'effet de la structure graphémique diminue-t-il avec l'avancée dans la scolarisation ?.....</i>	<i>201</i>
II. Les processus de lecture selon la structure graphémique d'un mot.....	204
<i>Les élèves de CP (N=143).....</i>	<i>204</i>
<i>Les élèves de CE1 (N=180).....</i>	<i>207</i>
<i>Les élèves de CE2 (N=174).....</i>	<i>210</i>
<i>Les élèves de CM1 (N=188).....</i>	<i>212</i>
<i>Les élèves de CM2 (N=176).....</i>	<i>214</i>
III.Evolution des processus de lecture selon la structure graphémique d'un mot	217
<i>Evolution des processus de lecture des mots complexes</i>	<i>217</i>
<i>Evolution des processus de lecture des mots simples</i>	<i>219</i>
CONCLUSION	221

INTRODUCTION

D'une manière générale, les modèles d'acquisition de la lecture sont issus de recherches anglophones. Or, bien que l'anglais soit une langue alphabétique comme le français¹, cette dernière se différencie sur plusieurs points. En effet, bien que ces deux langues utilisent le même système d'écriture, elles possèdent chacune leur propre orthographe et leurs propres conventions graphophonologiques.

Ainsi, alors que pour la première langue à chaque graphème correspond un phonème, celle de la seconde est plutôt opaque. En effet, le français n'est pas biunivoque (un graphème peut correspondre à plusieurs sons et un même phonème peut être transcrit de plusieurs façons). En effet, si on considère que l'unité de traitement est le graphème, alors les correspondances graphèmes-phonèmes sont plutôt régulières et prédictibles. Par exemple, les graphèmes « eau », « au », « o » se prononcent toujours [o]. A l'inverse, les correspondances phonèmes-graphèmes le sont moins. Ainsi, le son [o] peut être retranscrit de différentes manières. D'où l'asymétrie du graphisme en français.

Gak (1976)² distingue quatre types de signes graphiques en français. D'abord, les lettres simples, les lettres à signe diacritique (c'est-à-dire celles avec un accent aigu, grave ou circonflexe), les ligatures (telles que « œ » et « ae ») et les groupes de lettres (digrammes et trigrammes). La difficulté de l'apprentissage de la lecture en français réside dans le fait qu'il existe des distorsions entre la graphie et le son. Ainsi, la prononciation de certains graphèmes est régie par des lois contextuelles (ex. « g » va se lire différemment dans « garçon » et « gitan » ou le graphème « -ill » va se lire différemment dans « fille » ou « ville »). Certaines lettres simples peuvent correspondre à un groupe de phonèmes (par exemple la lettre « x » du mot « axe » qui se lit [ks]). Enfin, certaines lettres sont aussi parfois muettes. C'est le cas du *e* muet, du *h* et aussi de certaines consonnes finales qui sont parfois muettes devant certains mots mais qui vont se prononcer devant d'autres (ex. « les yeux » et « les yaourts » ; « les hommes » et « les haricots »).

Toutes ces caractéristiques font que l'apprentissage de la lecture en français entraîne des difficultés de traitement pour lire certains mots.

¹ Un système est alphabétique lorsqu'il permet, à partir d'un nombre déterminé de lettres, de représenter les composants phonologiques de la langue.

² L'orthographe du français. Essai de description théorique et pratique. (1976).

Dans le cadre de l'acquisition de la lecture, des études ont été menées afin d'évaluer l'incidence de facteurs tels que la régularité et la complexité des correspondances graphèmes-phonèmes dans la mise en place de la procédure de médiation phonologique.

On trouve notamment la recherche menée par Sprenger-Charolles (1996)³ sur l'acquisition de la lecture/écriture en français qui porte sur 57 élèves de CP et 48 de CE1. Dans une étude, elle suppose que si l'unité de traitement utilisée par les enfants au début de l'apprentissage de la lecture est la lettre, alors elle devrait trouver un effet de la complexité graphémique. Les résultats font apparaître que, de façon générale, ce facteur n'a aucune incidence sur la lecture de mots.

PREDICTIONS

Nous nous demandons donc dans quelle mesure les erreurs dues à la complexité graphémique d'un mot peuvent être considérées comme résultant d'un apprentissage discontinu ou continu de la lecture.

Nous supposons donc qu'au début de l'apprentissage de la lecture, c'est-à-dire lors de la mise en place du principe alphabétique, la complexité graphémique d'un mot a un effet lors de sa lecture. Si cela est vrai, alors les mots contenant un graphème complexe seraient plus difficiles à lire que les mots contenant des graphèmes simples (une lettre pour un phonème).

Ainsi, d'un point de vue quantitatif, nous nous attendons à ce que de façon générale, le nombre d'erreurs sur les mots contenant au moins un graphème complexe soit supérieur au nombre d'erreurs sur les mots contenant des graphèmes simples. Nous nous attendons à observer une rupture avec la maîtrise du code qui devrait faire diminuer ce type d'erreurs dans le temps.

D'un point de vue qualitatif, nous nous attendons à ce que les mots complexes engendrent davantage d'erreurs de décodage au début de l'apprentissage de la lecture. La maîtrise du traitement alphabétique devrait faire diminuer ce type d'erreur dans le temps et nous nous attendons à voir apparaître chez les élèves plus âgés des erreurs de type lexical.

³ « Lecture et écriture : acquisition et troubles du développement »

METHODOLOGIE DE TRAITEMENT DES DONNEES


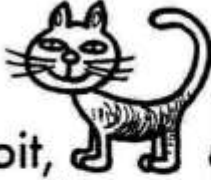

Pour cette étude, nous avons catégorisé les mots en deux groupes : les mots simples et les mots complexes. Ainsi, nous avons considéré qu'un mot est simple lorsqu'à chaque lettre correspond un phonème (« joli »). Par opposition, nous dirons qu'un mot est complexe, lorsque un phonème est représenté par un digramme (deux lettres pour un son : ex. « au ») ou un trigramme (trois lettres pour un son : ex. « eau »). Nous avons aussi inclus les mots qui comportent une lettre dont la lecture dépend du contexte orthographique (comme le *g* dans les mots « gai » et « geai ») et les lettres muettes (« temps ») (*cf. CD annexes*).


Nous avons répertorié 65 mots ne contenant que des graphèmes simples et 200 mots contenant au moins un graphème complexe.


Les élèves n'ayant pas tous lu le même nombre de mots, nous avons travaillé sur le rapport entre le nombre d'erreurs commises sur les mots simples/complexes et le nombre de mots simples/complexes lus.




Le traitement statistique des données a été effectué avec des tests non paramétriques étant donné que notre distribution ne suit pas une loi normale. Ainsi, pour les comparaisons entre les mots simples et les mots complexes à même niveau, nous avons eu recours au test de Wilcoxon (deux échantillons liés : les mots simples et les mots complexes). Puis, afin de comparer les différences des taux d'erreurs sur les mots simples/complexes selon les différents niveaux de scolarisation, nous avons utilisé dans un premier temps, le test de Mann-Whitney (pour comparer 2 échantillons indépendants, soit le cycle 2 et le cycle 3) puis, dans un second temps, le test de Kruskal-Wallis (pour *k* échantillons indépendants, c'est-à-dire la comparaison des classes) (*cf. CD annexes*).



Cartographie 3 : mots simples et complexes dans le texte de *L'alouette*


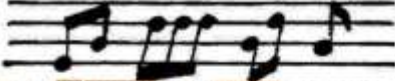
 **L'alouette.**  

Sous **la** mousse ou **sur** **le** toit,
dans les haies vives ou **le** chêne fourchu,
le printemps **a** mis ses nids.
Le printemps **a** nids au bois. 

Annie amie, **du** renouveau, c'est **le** doux temps.
Amie Annie, au bois **joli** gamine **le** pinson.
Dans les buis, gîte **une** biche, au bois chantant.
Annie, Annie! au doigt **joli**, **une** églantine laisse **du** sang :
au bout **du** temps des féeries viendra l'ennui. 

 L'alouette fait ses jeux; alouette fait un nœud **avec** un rien **de** paille.
L'hirondeau piaille sous **la** pente des bardeaux et, **vif** et gai, **le** geai
sur l'écaille argentée **du** bouleau, promène **un** brin d'osier.
 Au verger, dans **le** soleil **matinal** goutte **une** pompe dégelée.
On voit un **bec** luisant qui trille éperdument des notes claires
et, dans les pampres **d'or** que suspend **la** grille antique,
on surprend des rixes **de** moineaux.
Au potager s'alignent les cordeaux; **l'if** est **triste** **à** l'horizon
et lourd et lent l'envol des corbeaux. 

 Un **lac** étire ses calmes rives et, quand **le** soir descend,
le miroir **de** ses eaux **reflète** les poisons des brignoles perfides.
Et, quand descend **le** soir, quand joue **la** pourpre **du** couchant,
le ciel rougit ses eaux.
Dans **la** moire **de** l'eau danse l'ombre d'un écueil.
Tout est cris! Tout est bruits! 

 Une **amarre** est décochée... **une** barque est arrimée... des matelots
jettent leurs cassettes **sur** **le** rivage...
Tout est cris! Tout est bruits!
Au clair **de** **la** lune mon ami Pierrot... 
Au clair **de** lune mon amie annie...
Au clair **de** **la** lune mon ami Pierrot, **prête**-moi **la** plume pour **écrire** un mot.

o u e i a
le la les un dans des do ti pu mi

Mots simples
Mots complexes

I. Est-ce que la structure graphémique d'un mot fait difficulté ?

Le nombre d'erreurs varie-t-il avec la structure graphémique d'un mot ?

S'il existe un effet de la structure graphémique alors nous nous attendons à observer davantage d'erreurs sur les mots complexes que sur les mots simples.

De façon générale, on constate que les mots complexes posent plus de difficultés de lecture que les mots simples puisqu'en moyenne 14% des mots complexes ont été mal lus contre 6% des mots simples (cf. tableau 18).

Tableau 18 : pourcentages d'erreurs commises selon la complexité des mots et le cycle de scolarisation

Cycle	% d'erreur sur les mots simples lus	% d'erreur sur les mots complexes lus	Valeur du Z de Wilcoxon	Seuil de significativité p.
Cycle 2 (N=323)	7	23	14.951	.000
Cycle 3 (N=538)	6	8	13.315	.000
Ensemble	6	14	20.590	.000

Lecture : 23% des mots complexes qui ont été lus en cycle 2, n'ont pas été bien lus.

Si on regarde de plus près les résultats, on constate que la tendance est la même en cycle 2 et en cycle 3. En effet, en cycle 2, 23% des mots complexes ont été mal lus contre 7% des mots simples. Il en est de même en cycle 3 puisque 8% des mots complexes ont été mal lus contre 6% des mots simples.

Ces premiers résultats semblent donc aller dans le sens de notre supposition, à savoir que les mots contenant un phonème représenté par plusieurs lettres est plus difficile à lire qu'un mot dans lequel à chaque lettre correspond un son.

Affinant encore notre analyse, on regarde maintenant les taux d'erreurs par classe. On remarque que les résultats sont les mêmes quel que soit le niveau scolaire considéré : les mots complexes engendrent davantage d'erreurs de lecture que les mots simples (cf. tableau 19).

Tableau 19 : pourcentages d'erreurs commises selon la complexité des mots et la classe

Classe	% d'erreur sur les mots simples lus	% d'erreur sur les mots complexes lus	Valeur du Z de Wilcoxon	Seuil de significativité p.
CP (N=143)	6	32	10.279	.000
CE1 (N=180)	8	17	10.710	.000
CE2 (N=174)	6	11	9.465	.000
CM1 (N=188)	6	8	7.482	.000
CM2 (N=176)	5	6	5.359	.000

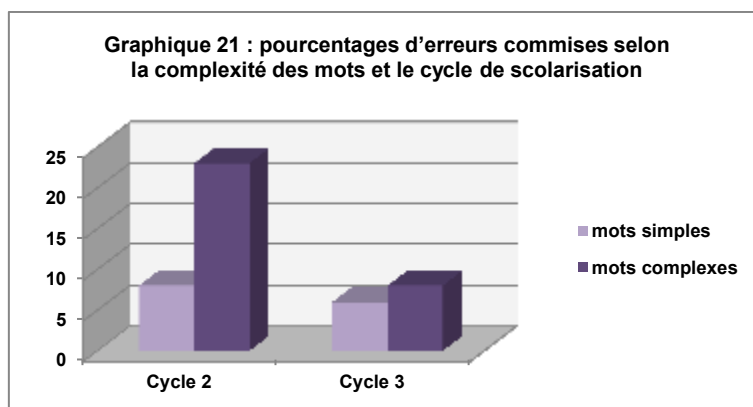
Lecture : en CP, 32% des mots complexes qui ont été lus, n'ont pas été bien lus.

En effet, en CP, 32% des mots complexes ont été mal lus contre 6% pour les mots simples. En CE1, 17% d'erreurs sont sur les mots complexes et 8% sur les mots simples. En CE2, bien que l'écart entre les erreurs commises sur les mots complexes et simples se réduise (3 points d'écart), on constate que ce sont 11% des mots du premier groupe qui ont été mal lus contre 6% des mots du second groupe. En CM1, ces taux passent à 8% pour les premiers et 6% pour les seconds. Enfin, en CM2, ce sont 6% des mots complexes qui ont été mal lus contre 5% des mots simples.

En résumé, il semble donc qu'il existe un effet de la structure graphémique lors de la lecture de mots puisque les mots complexes posent plus de difficulté de lecture que les mots simples, et ce quelle que soit la classe fréquentée. Par ailleurs, il semble que cet effet soit davantage présent au début de l'apprentissage de la lecture. Nous supposons donc qu'avec la maîtrise du code, cet effet devrait diminuer de façon significative avec le temps.

L'effet de la structure graphémique diminue-t-il avec l'avancée dans la scolarisation ?

D'une façon générale, il apparaît que l'effet de complexité d'un mot diminue d'un cycle de scolarisation à l'autre (cf. graphique 21).



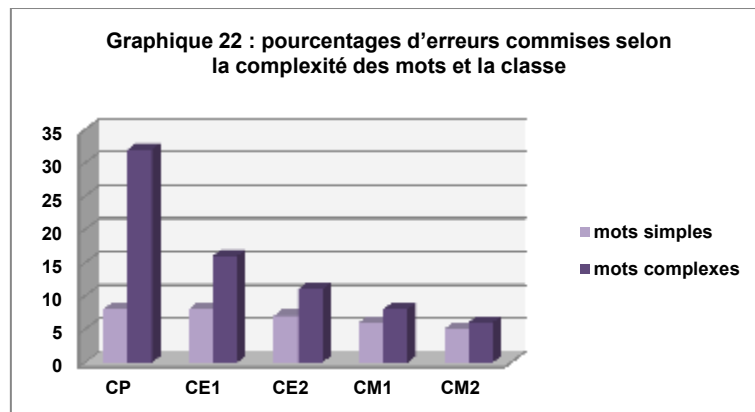
En effet, on remarque une baisse significative d'un moment à l'autre de l'apprentissage de la lecture. Alors qu'en cycle 2, 23% des mots complexes ont été mal lus, ils ne sont plus que 8% en cycle 3, ce qui correspond à une baisse de 15 points ($z = -16.281, p < .01$). Cette tendance n'est pas la même pour les mots simples puisqu'on n'observe pas de différence entre

les cycles. En effet en cycle 2, ce sont 7% des mots de ce type qui sont mal lus et 6% en cycle 3 (-1 point ; $z = .347$, *NS.*).

En **résumé**, le nombre d'erreurs dues à la complexité d'un mot diminue donc de façon significative entre le cycle 2 et le cycle 3. Cette baisse va dans le sens de notre supposition puisque les mots complexes semblent poser moins de difficultés de lecture au fur et à mesure de l'avancement dans la scolarité. Toutefois, il semblerait qu'on observe une certaine continuité étant donné que la progression des élèves ne semble pas être marquée par une rupture entre les deux cycles. Néanmoins, le fait que nous n'observons pas de rupture entre ces deux moments du cursus scolaire ne signifie pas pour autant qu'elle n'existe pas.

Une analyse plus détaillée des taux d'erreurs selon la classe nous permettra alors de confirmer ou d'infirmer le phénomène de continuité observé.

Dans le détail, on remarque que cette baisse est significative d'une classe à l'autre.



Ainsi, en ce qui concerne les erreurs commises sur les mots complexes, on remarque entre le CP et le CE1, une baisse de 15 points ($\chi^2(1) = 6.723$, $p < .01$) (cf. graphique 22). A partir du CE1, cette diminution apparaît dans de moindres proportions. En effet, on constate une baisse de 6 points entre le CE1 et le CE2 ($\chi^2(1) = 3.974$, $p < .01$), une baisse de 3 points entre le CE2 et le CM1 ($\chi^2(1) = 4.038$, $p < .01$) et de 2 points entre le CM1 et CM2. Toutefois, cette dernière est moins significative qu'aux niveaux précédents ($\chi^2(1) = 2.993$, $p < .05$).

En revanche, les résultats relatifs aux erreurs commises sur les mots simples ne suivent pas la même tendance : il y a une différence significative des pourcentages d'erreurs commis sur les mots simples seulement entre le CP et le CE1. En effet, cette proportion augmente

entre les deux classes puisqu'elle est de 6% en CP et de 8% en CE1 ($\chi^2(1)=-3.450, p < .01$). En CE2, elle diminue de deux points et passe à 6% ($\chi^2(1)=.742, NS$). En CM1, elle est de 6% ($\chi^2(1)=.728, NS$) et de 5% en CM2 ($\chi^2(1)=.466, NS$).

En **résumé**, bien que cette diminution de l'effet de la structure graphémique soit constante entre le CP et le CM2, elle n'est pas pour autant régulière : elle est plus importante entre le CP et le CE1. Il semble donc qu'il existe une rupture dans la progression des élèves. Rupture qui semble confirmer la mise en place d'un stade alphabétique au cours duquel l'élève aurait recours à la médiation phonologique. L'idée d'un apprentissage discontinu de la lecture, comme le suggèrent les modèles développementaux, semble donc ici se vérifier.

En **conclusion**, ces résultats semblent donc confirmer nos prédictions de départ. D'une part, il apparaît que la structure graphémique d'un mot a une influence sur sa lecture. En effet, on observe un **nombre d'erreurs plus important lors de la lecture de mots complexes** que de mots simples et ce quel que soit le niveau de scolarisation. D'autre part, cet effet diminue de façon constante avec le niveau d'avancement dans l'apprentissage de la lecture : **le nombre d'erreurs commises diminue systématiquement d'une classe à l'autre**. Cependant, il apparaît une certaine irrégularité. En effet, cette baisse est plus marquée entre le CP et le CE1. Ce constat laisse donc penser qu'il existerait une rupture entre les élèves de CP et les autres. Cette rupture semble donc être le reflet d'un apprentissage discontinu de la lecture comme le suggèrent les modèles développementaux. Il est alors possible que cette rupture soit le signe d'un changement de stratégie pour lire les mots complexes, stratégie qui correspondrait au moment où les correspondances graphèmes-phonèmes sont maîtrisées et où la voie orthographique se met en place.

Afin de voir si cette rupture correspond à un changement de stratégie, nous avons mené une seconde analyse, d'ordre qualitatif. A partir des différents types d'erreurs commises, nous avons cherché à savoir comment l'élève procède face à des mots contenant des graphèmes complexes.

II. Les processus de lecture selon la structure graphémique d'un mot

Rappelons que nous avons identifié trois types de processus : le décodage, l'intention lexicale, l'abstention et les processus non identifiés (*cf.* Chapitre 7 « *Types de réponses erronées et processus de lecture* »). En effet, si on regarde les différentes verbalisations des élèves, on distingue les logatomes des mots. Le premier cas correspond à des erreurs de décodage conduisant l'élève à prononcer un non-mot. Lorsque que la verbalisation est un mot, il est possible que l'élève ait eu une intention lexicale. Cependant, pour certains mots il nous est plus difficile d'en déduire le processus de lecture, c'est pourquoi, ce type d'erreur figure dans la catégorie « processus non identifiés ». Enfin, dans d'autres cas, l'élève n'a rien prononcé.

S'il existe un effet de la complexité graphémique, alors nous nous attendons à ce que les mots contenant des graphèmes complexes entraînent davantage d'erreurs de décodage au début de l'apprentissage de la lecture. En effet, si les élèves passent par une étape dite « alphabétique », on s'attend alors à observer davantage d'erreurs de décodage sur les mots complexes que sur les mots simples, dès les premières années de l'enseignement institutionnel de la lecture. La maîtrise des correspondances graphèmes-phonèmes devrait faire diminuer les écarts de ce type d'erreur entre les deux catégories de mots. Si on considère que les élèves passent ensuite par une étape « orthographique », alors nous nous attendons à ce que les erreurs lexicales soient plus nombreuses chez les élèves plus âgés.

Les élèves de CP (N=143)

En CP, il apparaît que parmi les types d'erreurs les plus fréquemment commis lors de la lecture de mots complexes, on trouve d'abord les erreurs dues à un mauvais décodage, puis les abstentions, les verbalisations de mots issues de processus non identifiés et en dernier les intentions lexicales.

En effet, les élèves qui ont commis des erreurs lors de la lecture de mots complexes ont commis en moyenne **56% d'erreurs de décodage**. Ces erreurs se traduisent essentiellement, par un décodage séquentiel (lettre par lettre ou graphème par graphème) la plupart du temps incomplet. Par exemple, le mot « printemps », mot sur lequel 41% des élèves de CP ont commis des erreurs de décodage, a été lu de façon partielle. En effet, il a été lu par exemple

[pɾitə], [pɾɛ̃tə] ou [pɾi]. Il semblerait que ce soit le morphème « -temps » qui soit difficile à prononcer. Le phonème [ã] qui se compose de 4 lettres (une voyelle et 3 consonnes dont 2 finales muettes) semble poser des difficultés de lecture aux élèves et face à cette difficulté de traitement, certains élèves n'ont pas lu la dernière partie du mot. Ce constat semble se confirmer avec le mot « temps » puisque 37% des élèves de CP ont commis une erreur de décodage séquentiel et ont prononcé [tə], [tãp] ou encore [təmə]. Autre exemple, le mot « renouveau », sur lequel 40% des élèves ont commis une erreur de lecture. Il apparaît que pour ce mot, la lecture s'est effectuée deux éléments par deux éléments entraînant une resyllabation du mot : [ɾənɯvəo], [ɾãnuvo] ou encore [ɾãuvo].

Parmi les erreurs les plus fréquentes commises sur les mots complexes, viennent ensuite les **abstentions** qui **représentent en moyenne 21%** des erreurs. On remarque que ce type d'erreur concerne essentiellement les mots « haies », « printemps » et « chêne » (11%). Pour le premier, la difficulté vient du fait qu'ici cinq lettres représentent un son (« haies » pour [ɛ]). Il semblerait que face à cette difficulté, 35% des élèves aient préféré ne rien prononcer au risque de se tromper. Les particularités graphémiques du mot « printemps » énoncées précédemment font que 13% des élèves de CP ne l'ont pas lu. Quant au mot « chêne », il se pourrait que ce soit le graphème « ch » et le « ê » qui aient posé des difficultés de lecture.

Ensuite, on remarque qu'en moyenne **13% des erreurs sont des verbalisations de mots issues de processus non identifiés**. Cela signifie que nous ne pouvons dire si les élèves ont eu une intention lexicale ou pas. On trouve dans cette catégorie les mots « bois » lu [dwa] (15%), « amie » lu [ani] (13%) et « doux » lu [bu] (9%). En effet, il paraît difficile de dire si le mot prononcé est le résultat d'une erreur de prononciation due à une ressemblance graphique et phonémique entre les lettres b/d et n/m ou d'une intention lexicale. En effet, la phrase précédente commence par « Annie amie » et les élèves devaient lire « Amie Annie ». Il est donc possible que cette similitude les ait amenés à vouloir commencer la phrase de la même façon que la précédente (*cf.* Chapitre 12 « *L'influence des similitudes lors de la lecture du texte de L'alouette* »).

Enfin, les élèves ont commis en moyenne **11% d'erreurs lexicales** lors de la lecture de mots complexes. Elles portent essentiellement sur les mots « haies », « au » et « pinson » lu respectivement [ʃa], [dy] et [ani]. En ce qui concerne le premier mot, sa structure grapho-phonémique a pu amener les élèves à effectuer un décodage partiel, les premières lettres « ha- », puis à prononcer un mot qui leur est plus familier « chat ». De plus, précisons qu'il existe aussi un possible effet du dessin pour ce type d'erreur. En effet, non loin du mot « haies » se

situe le dessin d'un chat (cf. Chapitre 11 « *L'influence des dessins lors de la lecture du texte de L'alouette* »). En ce qui concerne le mot « au », qui présente peu de ressemblances graphique et phonémique avec le mot « du », il se pourrait que, comme pour le mot « amie », les élèves aient été influencés par la phrase précédente dont le début est similaire : « Annie amie, du renouveau [...] » et « Amie annie, au bois joli [...] ». Il est donc possible que cette similitude les ait amenés à vouloir commencer la phrase de la même façon (cf. chapitre 12 « *L'influence des similitudes lors de la lecture du texte de L'alouette* »).

Ainsi, cette analyse semble aller dans le sens de notre supposition, à savoir qu'en CP les mots complexes entraînent davantage d'erreurs de décodage. Cependant, il convient de rappeler que nous avons la catégorie « processus non identifiés » dans laquelle les verbalisations sont des mots mais nous ignorons si l'élève a effectivement voulu produire un mot. De ce fait, si on considère que pour ces mots il y a eu une intention lexicale, alors cette catégorie (« intention lexicale ») passe de 11% à 23%. On remarque donc que, même en tenant compte de ce résultat, ce sont toujours les erreurs de décodage qui sont les plus nombreuses puisqu'elles s'élèvent à 56%. La complexité d'un mot entraînerait donc plus d'erreurs de décodage au début de l'apprentissage de la lecture. Toutefois, cette supposition ne saurait se vérifier sans une analyse des mots simple : s'il existe un effet de la complexité, alors nous devrions trouver moins d'erreurs de décodage sur les mots simples.

Il apparaît en effet que pour les mots simples le type d'erreur le plus fréquent reste les erreurs de décodage mais dans de moindres proportions : alors que les mots complexes engendrent 56% d'erreurs de décodage, on constate que 28% des erreurs de ce type ont lieu sur les mots simples.

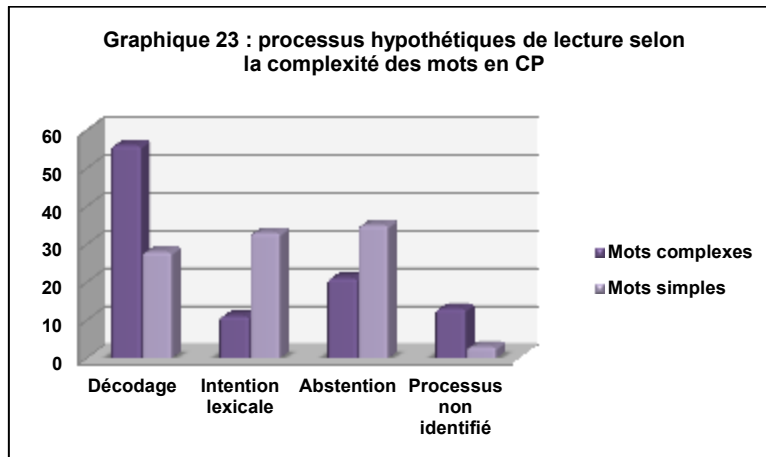
On remarque que ce sont les abstentions les plus nombreuses puisqu'elles représentent en moyenne 35% des erreurs commises sur les mots simples.

Les intentions lexicales, quant à elle, représentent 33% pour les mots simples contre 11% des mots complexes.

Enfin, les verbalisations issues de processus non identifiés sur les mots simples sont de seulement 3%.

En **résumé**, cette première analyse qualitative montre qu'en CP les mots complexes entraînent essentiellement des erreurs de décodage séquentiel, que ce soit graphème par

graphème ou lettre par lettre (*cf.* graphique 23). Le fait que ces erreurs soient plus nombreuses sur les mots complexes que sur les mots courts semble aller dans le sens d'un effet de la structure graphémique lors de l'apprentissage de la lecture.



Par ailleurs, la présence d'erreurs de type lexical à ce niveau scolaire, laisse penser qu'au début de l'apprentissage de la lecture, les élèves ne feraient pas que décoder mais chercheraient aussi à donner du sens à leur lecture.

Les élèves de CE1 (N=180)

En CE1, comme au niveau précédent, le type d'erreur le plus fréquent sur les mots complexes concerne le décodage. En effet, en moyenne **48% des erreurs sont dues à un mauvais décodage**. Les mots « féeries », « piaille » et « d'osier » sont ceux qui ont été mal lus par au moins la moitié des élèves. Il semble qu'à ce niveau scolaire les erreurs de décodage ne soient plus essentiellement des erreurs de décodage séquentiel, mais plutôt des erreurs de décodage deux éléments par deux éléments ou d'attribution de son. Le mot « féeries » a une prononciation particulière puisqu'il se prononce [fœəri]. 72% des élèves de CE1 l'ont prononcé [fəri]. Il est possible que les élèves aient reconnus le morphogramme « fée » et l'aient prononcé comme tel (*cf.* chapitre 8 « *Etude des mots qui ont posé le plus de difficulté de lecture* »). En ce qui concerne le mot « piaille » (50% d'erreurs), il a été lu [pɛjə], [piɛjə] ou [pjɛlə]. Ici, les élèves semblent avoir lu ce mot en effectuant un décodage deux à deux : pi_ai_lle, pai_lle. Enfin, quant au mot « d'osier » (48% d'erreurs), il a été lu [dosje].

Dans ce cas, il semblerait que ce soit davantage une méconnaissance des règles de prononciation du graphème « s » qui entre deux voyelles se prononce [z].

Ensuite, on trouve **les verbalisations de mots issues de processus non identifiés qui représentent en moyenne 20% des erreurs commises sur les mots complexes**. Dans cette catégorie, on trouve les mots « geai », « pompe » et « rixes » lus respectivement [ge], [pɔm] et [risk]. En ce qui concerne le mot « geai », il semble difficile de dire si le mot prononcé est dû à une mauvaise prononciation de la lettre « g » ou si cette erreur a été influencée par une similitude avec le mot précédent « gai ». Il en est de même pour le mot « pompe » dont nous ne pouvons expliquer la verbalisation : est-ce une erreur due à une ressemblance phonémique [m]/[p] ou est-ce une intention lexicale due à la présence du dessin de la pomme situé juste à côté ? Enfin, concernant le mot « rixes », l'incertitude repose sur les sons [ks] : est-ce que l'élève lorsqu'il a lu ce mot a eu l'intention de dire [risk] ou est-ce sa confusion de son avec les sons [sk] qui l'a fait prononcer [risk] ?

Les abstentions sont le troisième type d'erreur puisqu'on compte qu'en moyenne **17% des mots complexes n'ont pas été lus**. On retrouve le mot « haies » qui n'a pas été lu par 16% des élèves de CE1. Ensuite, on trouve les mots « l'écaille » et « un » de la phrase « sur l'écaille argentée du bouleau promène un brin d'osier ». Pour le premier mot, 22% des élèves ne l'ont pas lu. Il se pourrait que pour ce mot, la difficulté réside dans la fusion des graphèmes « a » et « -ille » qui donne les sons [aj]. En ce qui concerne le mot « un », nous supposons que cette erreur fait partie du saut de ligne. En effet, plusieurs lignes du texte ont été sautées et celle-ci en fait partie.

Enfin, **les erreurs lexicales représentent en moyenne 15% des erreurs sur les mots complexes en CE1**. Ce sont essentiellement, les mots « l'ennui », « nœud » et « l'hirondeau » qui sont concernés. « L'ennui » (32% d'erreurs) a été majoritairement lu [lɛnuji] ou [lanuji]. Il se pourrait que pour ce mot, ce soit l'article élide « l' » qui les aient amenés à effectuer un mauvais découpage du mot et à prononcer un mot qui leur est plus familier. Précisons que là aussi, nous ne pouvons écarter l'hypothèse de l'influence du dessin, puisque non loin de ce mot, se situe le dessin de la nuit. Il en est de même pour le mot « nœud » lu essentiellement [œf]. Les élèves ont certainement décodé les graphèmes du milieu du mot « œu » et, à partir de ces indices, ont ensuite lexicalisé un mot qui leur est plus familier. Dans ce cas là aussi, il est possible que le dessin du nid avec les œufs les ait amenés à cette lexicalisation. Enfin, le mot « l'hirondeau » a été lu [lirɔdɛl]. On remarque que ces deux mots se ressemblent d'un point de vue graphique et phonémique : même graphèmes du début et du milieu « hirond ». Il

se pourrait que cette partie décodée ait amené les élèves à prononcer un mot qui leur est plus familier. Là aussi, le dessin de l'hirondelle a dû les influencer.

Ainsi, il semble donc qu'en CE1, l'essentiel des erreurs commises sur les mots complexes concerne encore les erreurs de décodage, ce qui semble donc aller dans le sens d'un effet de la structure graphémique. Si on considère que les processus non identifiés sont en réalité des intentions lexicales, alors, cette dernière catégorie passe de 15% à 35%. On constate que cette proportion reste inférieure à celles des erreurs dues à un mauvais décodage (48%). Toutefois, comme précédemment, nous devons confirmer ce constat par une analyse des mots simples. Ainsi, si la structure graphémique a une influence lors de l'apprentissage de la lecture alors nous devrions trouver moins d'erreurs de décodage sur les mots simples.

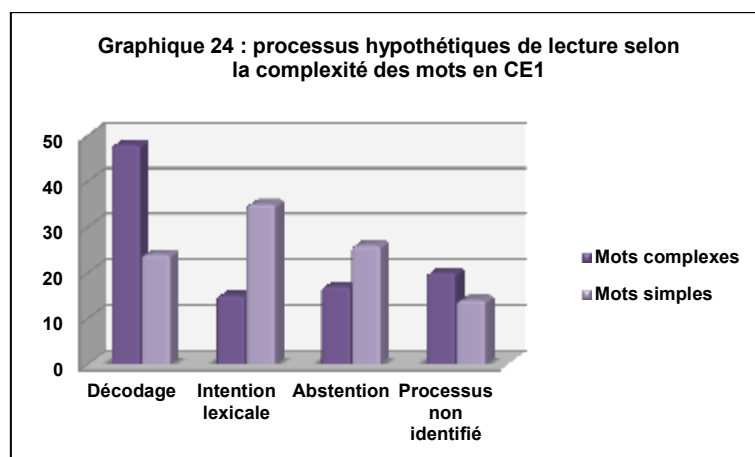
En effet, les erreurs de décodage sont deux fois moins nombreuses sur les mots simples que sur les mots complexes : 24% contre 48%.

Il apparaît que le type d'erreur le plus fréquent lors de la lecture des mots simples n'est pas le décodage mais l'intention lexicale puisque ce type d'erreur représente en moyenne 35% des erreurs commises sur les mots simples.

L'abstention représente 26% des erreurs.

Et les verbalisations issues de processus non identifiables sont de 14%.

En résumé, il apparaît qu'en CE1, les erreurs de décodage sur les mots complexes sont encore, à ce niveau, le type d'erreur principal et sont plus nombreuses que sur les mots simples. Ce constat va donc dans le sens d'une influence de la structure graphémique lors de l'apprentissage de la lecture. On observe qu'à ce niveau le décodage ne s'effectue plus essentiellement de manière séquentielle mais plutôt deux éléments par deux éléments.



D'autre part, la présence d'erreurs de type lexical semble confirmer le fait observé au CP, c'est-à-dire qu'au début de l'apprentissage de la lecture, les élèves ne feraient pas que décoder mais chercheraient aussi à construire du sens.

Les élèves de CE2 (N=174)

En CE2, **le type d'erreur principalement commis sur les mots complexes est encore le décodage** puisque 39% des erreurs sur les mots complexes sont dues à un mauvais décodage. On retrouve le mot « féeries » lu essentiellement [fɛri], le mot « piaille » lu [pajə] ou [piɛjə] et « les haies » lu [lɛzɛ]. Concernant ce dernier, l'erreur est plutôt supra-lexicale puisqu'il s'agit d'une liaison abusive. Il semblerait qu'à ce niveau, les élèves n'aient pas totalement acquis le principe de la liaison puisqu'ici la liaison ne doit pas se faire car c'est un « h » aspiré.

Les verbalisations de mots issues de processus non identifiés représentent 28% des erreurs commises sur les mots complexes. Elles concernent surtout les mots « poisons », « geai » et « cordeaux » lu respectivement [pwasɔ̃], [gɛ] et [kɔ̃rbo]. On se demande donc si, pour les deux premiers mots, cette verbalisation est le résultat d'une intention lexicale (décodage partiel puis lexicalisation) ou d'une erreur de prononciation des graphèmes contextuels « s » et « g ». Quant au troisième mot (« cordeaux »), on se demande si la verbalisation est le résultat d'une intention lexicale ou d'une erreur due à la ressemblance graphique entre les lettres d et b.

Les intentions lexicales représentent 18% et portent sur les mêmes mots qu'en CE1 : « l'hirondeau », « l'ennui » et « nœud ». Comme dans la classe précédente, ces mots ont été lus respectivement [lirɔ̃dɛl], [lanɥi]/[lenɥi] et [œf].

Enfin, **les abstentions représentent 15%** des erreurs commises sur les mots complexes et ont eu lieu sur les mots « un », « brin », « d'osier » de la phrase « sur l'écaille argentée du bouleau promène un brin d'osier ». Ici, comme pour les mots « écaille » et « un », il semble que la raison de la non lecture soit davantage le saut de ligne plutôt que la complexité.

Ainsi, la supériorité des erreurs de décodage commises sur les mots complexes nous amène à penser à un effet de la structure graphémique. Cependant, si on considère les processus non identifiés comme des intentions lexicales, alors cette catégorie totalise 46% des erreurs commises sur les mots complexes. Cette proportion devient alors le type d'erreur

principalement commis lors de la lecture de mots complexes. Par conséquent, même s'il nous est difficile d'identifier le processus de départ pour certaines verbalisations de mots, on peut toutefois conclure qu'en CE2, les élèves ont tendance à verbaliser plus de mots que de non mots. Cela signifierait donc que la procédure alphabétique serait sur le point d'être maîtrisée et par conséquent, la structure graphémique d'un mot aurait une moindre influence par rapport aux niveaux précédents. Ce constat nous amène donc à penser que la procédure orthographique s'établirait à ce niveau scolaire. Toutefois, ces suppositions doivent se confirmer ici aussi par une analyse des processus sur les mots simples. Ainsi, si la structure graphémique a une moindre influence alors nous devrions observer encore quelques erreurs de décodage mais dans de moindres proportions et une augmentation des erreurs lexicales.

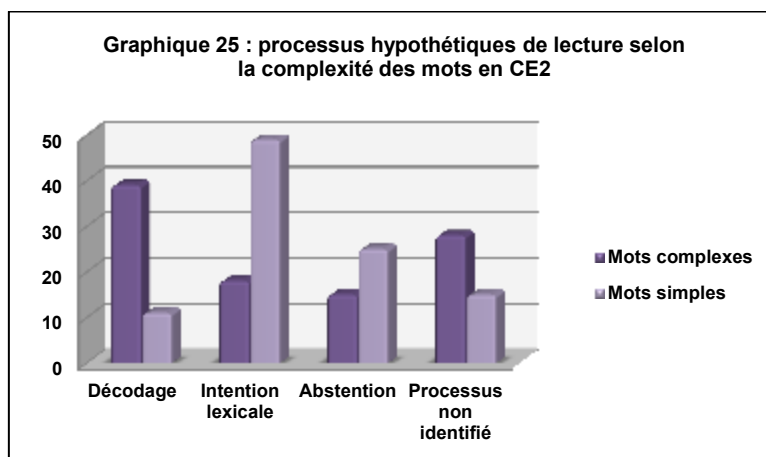
En effet, on remarque que les erreurs de décodage sont moins nombreuses sur les mots simples que sur les mots complexes : 11% contre 39%.

Comme au niveau précédent, ce sont les erreurs lexicales qui sont les plus nombreuses puisqu'elles représentent 49% des erreurs.

A ce niveau scolaire, les abstentions sur les mots simples s'élèvent encore à 25%.

Enfin, les verbalisations issues de processus non identifiés sont de 15%.

En **résumé**, il apparaît qu'en CE2 les mots complexes entraînent encore essentiellement des erreurs de décodage et en entraînent davantage que sur les mots simples (*cf.* graphique 25). Toutefois, si on tient compte des verbalisations de mots, qu'elles soient intentionnelles ou non, alors on remarque que ce type d'erreur devient celui qui est essentiellement commis lors de la lecture de mots complexes.



De plus, les verbalisations de mots sont plus nombreuses que les verbalisations de non mots, quelle que soit la structure graphémique des mots. Ces résultats semblent donc indiquer d'une part, un effet moindre de la structure graphémique, probablement dû à la maîtrise des correspondances grapho-phonémiques et d'autre part, la mise en place d'une procédure orthographique. Ce changement de processus semble donc indiquer qu'à ce niveau scolaire, une rupture s'opère. Rupture synonyme donc d'un apprentissage discontinu de la lecture. Si ce constat est vrai, alors les erreurs de décodage ne devraient plus être le type d'erreur principalement commis dans les classes suivantes.

Les élèves de CM1 (N=188)

On remarque qu'en CM1, les erreurs de décodage sont quasiment aussi nombreuses que celles issues de processus non identifiés.

En effet, **les verbalisations issues de processus non identifiables représentent 34%** des erreurs. On retrouve dans cette catégorie les mêmes mots qu'au niveau précédent, c'est-à-dire « poisons » lu [pwasõ], « geai » lu [ge] et « cordeaux » lu [kõRbo]. Comme expliqué précédemment, on se demande ici aussi si ces verbalisations sont le résultat d'une intention lexicale ou d'une erreur due à une méconnaissance de certaines règles de prononciation concernant les graphèmes contextuels « s » et « g » ou d'une ressemblance graphique des lettres b et d.

Les erreurs de décodage sont quasiment aussi nombreuses puisqu'elles sont de 33%. Comme au niveau précédent, on retrouve les mots « féeries » (49% d'erreurs de décodage) et « haies » (32%) lus respectivement [feri] et [leze]. En revanche, un nouveau mot apparaît : il s'agit de « pourpre » (31% d'erreurs de décodage) lu [pupRø]. Pour ce dernier, il est possible que ce soit la répétition des sons [p] et [R] qui ait posé des difficultés de lecture. Les élèves auraient donc omis de prononcer le premier [R] peut-être pour leur faciliter la verbalisation.

Les intentions lexicales représentent 21% des erreurs commises sur ce type de mot. Elles concernent les mots « écueil », « décochée » et « alouette » lu respectivement [ekyRøe], [dekrofe] et [laluæt]. Il est probable que pour le mot « écueil », les élèves aient cherché à prononcer un mot qui leur soit plus familier en ajoutant un « r » au mot qui est écrit. Pour cette erreur aussi, nous ne pouvons écarter l'idée d'une possible influence du dessin car, juste à côté se trouve le dessin de l'écureuil. En ce qui concerne le mot « décochée », les élèves ont

lu [dekroʃe] certainement à cause du co-texte sémantique dans lequel le mot se situe (*cf.* Chapitre 13 « *L'influence du co-texte sémantique lors de la lecture du texte de L'alouette* »). En effet, la phrase est la suivante : « Une amarre est décochée... ». Enfin, en ce qui concerne le mot « alouette » lu [laluet], il est fort probable là aussi, que ce soit le test de Lefavrais qui les ait amenés à cette verbalisation. En effet, la phrase juste avant commence par « l'alouette ». Comme pour les mots Annie et amie, il se pourrait que ce soit la similitude avec la phrase précédente qui les ait conduits à commettre cette erreur (*cf.* Chapitre 12 « *L'influence des similitudes lors de la lecture du texte de L'alouette* »).

Enfin, **11% des mots complexes n'ont pas été lus**. Les abstentions ont eu lieu essentiellement sur les mots « « écaille », « argentée », « bouleau ». On remarque que, comme au niveau précédent, ces trois mots font partie de la même phrase : « sur l'écaille argentée du bouleau promène un brin d'osier ». Ainsi, il semblerait que ce ne soit pas la complexité graphémique de ces mots qui ait amené les élèves à ne pas les lire.

Ainsi, il apparaît qu'en CM1, les erreurs de décodage sont quasiment aussi nombreuses que celles issues de processus non identifiés. Toutefois, si on considère ces dernières comme étant en réalité des intentions lexicales alors cette catégorie devient, comme en CE2, le processus le plus important lors de la lecture de mots complexes (56%). Nous pouvons donc émettre le même constat qu'au niveau précédent à savoir qu'en CM1 les élèves verbalisent davantage de mots que de non mots.

Comme aux niveaux précédents, cette analyse doit se compléter par une étude des processus lors de la lecture des mots simples.

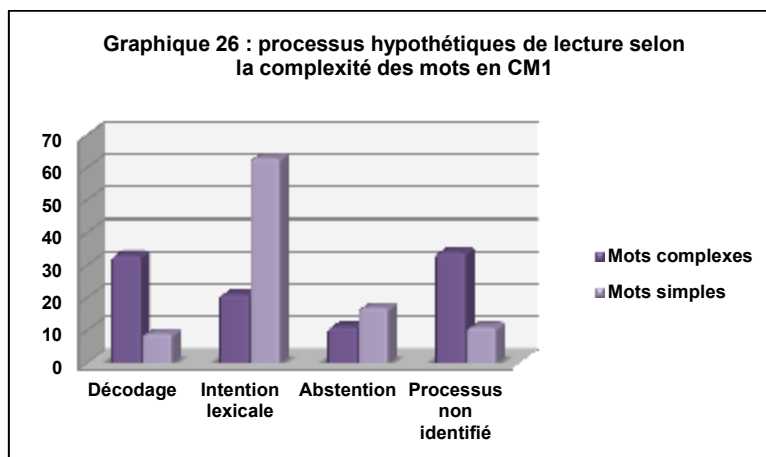
Du côté des mots simples, on constate que ce sont toujours les intentions lexicales qui sont le plus fréquemment commises puisqu'elles représentent 63% des erreurs.

Les abstentions sont de 17% pour les mots simples.

Les verbalisations de mots aux processus non identifiés sont de 11%.

Et les erreurs de décodage lors de la lecture de mots simples ne sont plus que de 9% (contre 33% pour les mots complexes)

En **résumé**, il semblerait qu'en CM1 la structure graphémique pose encore des difficultés de lecture puisqu'on constate toujours que les erreurs de décodage restent plus nombreuses sur les mots complexes que sur les mots simples (33% contre 9%).



Toutefois, cet effet est moindre puisque nous avons remarqué que les erreurs de décodage sur les mots complexes sont aussi nombreuses que les verbalisations de mots issues de processus non identifiés. Cela signifie que si on considère l'ensemble des verbalisations de mots, qu'elles soient intentionnelles ou non, alors cette catégorie devient, comme en CE2, le processus de lecture le plus important pour lire les mots complexes. Ces résultats semblent donc aller dans le sens du constat effectué en CE2, à savoir qu'il existe une rupture dans l'apprentissage de la lecture correspondant à un changement de stratégie.

Les élèves de CM2 (N=176)

Les résultats en CM2 semblent aller dans le sens de ceux effectués précédemment.

En effet, le type d'erreur principalement commis lors de la lecture de mots complexes ne semble plus être les erreurs de décodage mais les verbalisations de mots dont nous ne pouvons identifier le processus de départ. Cette catégorie représente **32% des erreurs**. Comme pour les élèves de CM1, elles concernent les mots « poisons », « geai » et « cordeaux » lus [pwasɔ̃], [ge] et [kɔ̃rbo].

Les intentions lexicales représentent 30% et concernent, ici aussi, les mêmes mots qu'au niveau précédent : « décochée », « écueil » et « alouette » lus [dekɔʃe], [ekyɾœj] et [laluet].

Les erreurs de décodage lors de la lecture de mots complexes sont le troisième type d'erreur commis puisqu'elles **représentent 28%**. Les mots les plus touchés par ce type d'erreur sont les mêmes qu'en CM1 : « féeries », « haies » et « pourpre » lus [feri], [leze] et [puprɛ].

Enfin, **les abstentions sont de 10%** et ont eu lieu sur les mots « quand », « et », « quand ». Ici, aussi, il semblerait que ce ne soit pas essentiellement la complexité graphémique de ces mots qui ait amené les élèves à ne pas les lire. En effet, ces mots se situent sur la même ligne « Et, quand descend le soir, quand joue la pourpre du couchant ».

Ainsi, à ce niveau aussi, ce sont les verbalisations de mots qui sont les plus nombreuses puisque au total cette catégorie représente 62% des erreurs contre 28% d'erreurs de décodage. Cela semble donc aller dans le sens d'une diminution de l'effet de la structure graphémique et de la mise en place de la procédure orthographique. Comme aux niveaux précédents, ces constats doivent se confirmer par une analyse des mots simples.

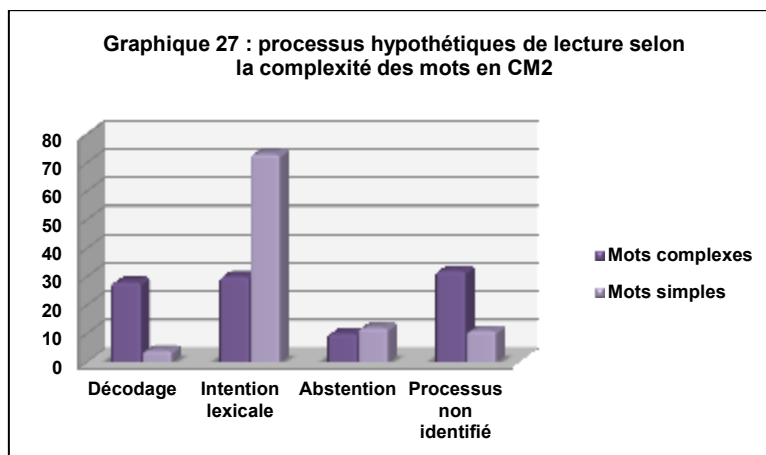
En ce qui concerne les mots simples, les erreurs les plus nombreuses sont les intentions lexicales qui se chiffrent à 73%.

Ensuite, viennent les erreurs qui découlent de processus non identifiés (11%).

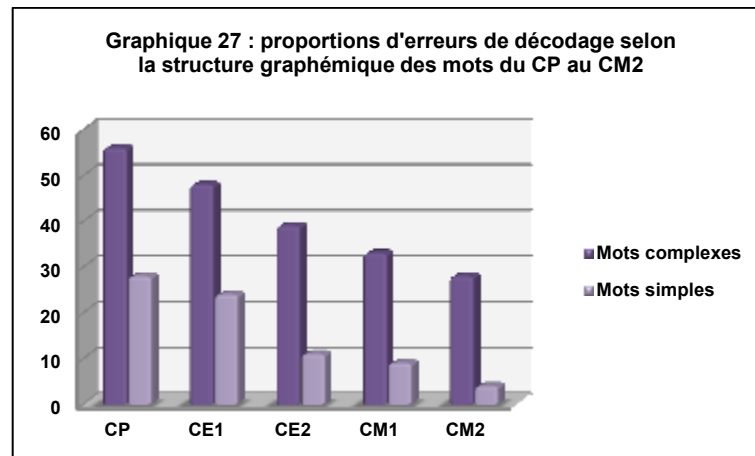
Les abstentions représentent encore 11% des erreurs commises sur les mots simples.

Enfin, les erreurs de décodage ne se chiffrent plus qu'à 4%.

En **résumé**, il apparaît qu'en CM2, bien que les erreurs de décodage ne soient plus le type d'erreur principal sur les mots complexes, elles sont toutefois plus nombreuses que sur les mots simples (*cf.* graphique 27). Ce constat semble donc confirmer encore un effet de la structure graphémique mais dans de moindres proportions. A ce niveau, les élèves commettent essentiellement des erreurs de verbalisations de mots, intentionnelles ou non, quelle que soit la difficulté du mot (62% pour les mots complexes et 84% pour les mots simples). Cela signifie donc que les élèves ont davantage recours à une procédure orthographique pour traiter les mots.



En **conclusion**, il apparaît que **les erreurs de décodage lors de la lecture de mots complexes, en comparaison avec les mots simples, sont à chaque fois plus nombreuses quel que soit le niveau scolaire considéré**. Ce résultat semble donc aller dans le sens de nos suppositions de départ, à savoir qu'il existe un effet de la structure graphémique.



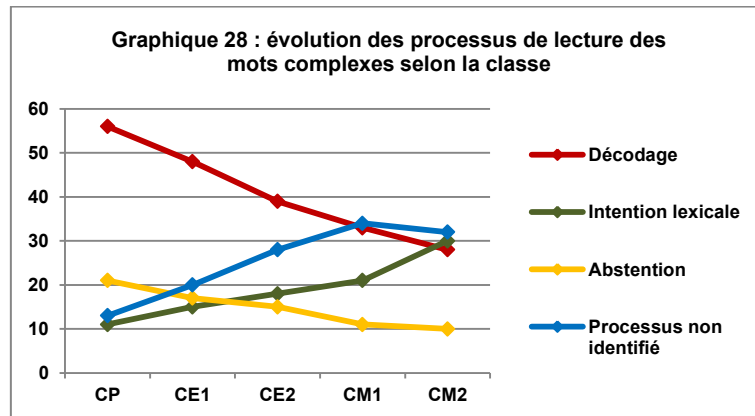
L'analyse des différents types de processus nous permet de constater par ailleurs que dès les premières années de l'enseignement institutionnel de la lecture, les élèves verbalisent des mots, que ce soit de façon intentionnelle ou non. Ce constat signifie que les élèves ne feraient pas que décoder mais chercheraient aussi à donner du sens à leur lecture. Il semble donc que les élèves ne passent pas tous par le même stade en même temps, contrairement à ce que supposent les modèles développementaux : les élèves semblent avoir recours à plusieurs stratégies pour traiter les mots complexes. L'analyse de l'évolution des processus va nous permettre d'en savoir un peu plus. Lorsqu'on observe ces verbalisations, on remarque que la plupart d'entre elles sont en lien avec le test de *L'alouette* (amie/annie, pompe/pomme, geai/gai, écueil/écureuil, cordeaux/corbeaux, l'hirondeau/l'hirondelle, décochée/décrochée, ... etc). Certaines peuvent avoir été influencées par les dessins, d'autres par les similitudes ou encore par le co-texte sémantique. Ainsi, selon que l'on considère les mots résultant d'un processus non identifié comme intention lexicale ou non, alors la conclusion n'est pas la même. En effet, dans le premier cas, on observe une rupture dans la progression des élèves à partir du CM1, puisqu'à partir de ce niveau le type d'erreur dominant est lexical. Dans le second cas, l'apprentissage de la lecture s'effectuerait de manière continue, puisque les erreurs de décodage resteraient dominantes du CP au CM2.

III. Evolution des processus de lecture selon la structure graphémique d'un mot

S'il existe un effet de la complexité graphémique, nous nous attendons à ce que les erreurs de décodage observées précédemment lors de la lecture des mots complexes, diminuent dans le temps avec la maîtrise des correspondances graphèmes-phonèmes. La mise en place d'un stade orthographique devrait faire apparaître une augmentation des erreurs lexicales. Ainsi, nous supposons que le nombre d'erreurs de décodage sur les mots contenant au moins un graphème complexe diminue et que les intentions lexicales augmentent au fur et à mesure de l'avancée dans la scolarisation.

Evolution des processus de lecture des mots complexes

S'il existe un effet de la structure graphémique lors de l'apprentissage de la lecture, alors nous devrions voir diminuer les erreurs de décodage et augmenter les erreurs lexicales lors de la lecture de mots complexes.



De façon générale, il apparaît que la proportion d'**erreurs de décodage** est différente selon le niveau de scolarisation ($\chi^2(1)= 157.047, p < .001$) (cf. graphique 28). Plus précisément, alors qu'en CP, 56% des erreurs sur les mots complexes, sont des erreurs de décodage, elles diminuent de 8 points en CE1, puisqu'elles sont de 48% ($\chi^2(1)= 2.959, p < .05$). Au niveau suivant, en CE2, ce taux baisse de 9 points (39%) ($\chi^2(1)= 3.390, p < .01$). En CM1, la proportion d'erreurs de décodage lors de la lecture de mots complexes diminue de 6 points mais de façon non significative et représente 33% des erreurs ($\chi^2(1)= 2.784, NS.$). En

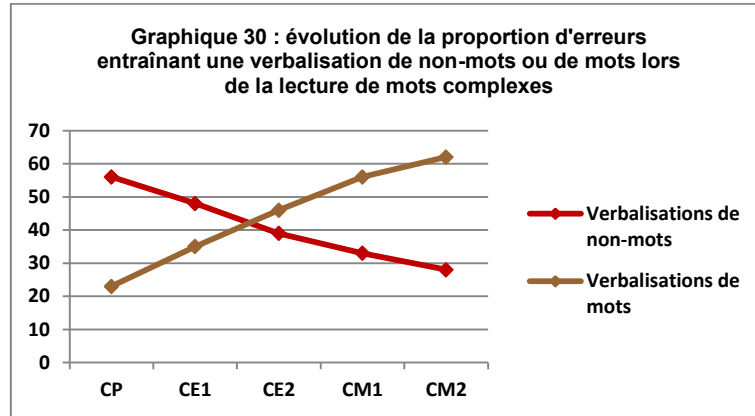
fin d'école primaire, là aussi on observe une baisse (-5 points) mais qui n'est pas non plus significative ($\chi^2(1) = 2.258$, *NS.*). Ces résultats confirment notre supposition selon laquelle les erreurs de décodage devaient diminuer avec le niveau d'avancement en lecture, puisqu'on observe une baisse significative entre les élèves de CP, CE1 et CE2 et les élèves de CE2 et les CM1/CM2.

Nous avons ensuite supposé que les **intentions lexicales** devaient augmenter au fur et à mesure de l'avancée dans la scolarisation. On constate qu'entre le CP et le CE1, les erreurs dues à une intention lexicale augmentent de façon significative : elles passent de 11% à 15% ($\chi^2(1) = -3.276$, $p < .05$). Aux niveaux suivants, bien que cette proportion augmente, on ne constate pas de différence significative. En effet, en CE2, les intentions lexicales représentent 18% des erreurs sur les mots complexes ($\chi^2(1) = -1.193$, *NS.*) et 21% en CM1 ($\chi^2(1) = -2.037$, *NS.*). En CM2, elles augmentent à nouveau de façon significative et passent à 30% ($\chi^2(1) = -3.322$, $p < .01$).

Toutefois, comme pour les analyses précédentes, si on considère que les processus non identifiés sont en réalité des intentions lexicales, alors on remarque que cette catégorie augmenterait au fur et à mesure de l'avancée dans la scolarisation. En effet, alors qu'en CP, ce type d'erreur représenterait 23%, il serait de 35% en CE1 ($\chi^2(1) = -4.444$; $p < .01$). En CE2, cette proportion continuerait d'augmenter et représenterait 46% ($\chi^2(1) = -4.151$; $p < .01$). Au niveau suivant, en CM1, elle serait de 56% ($\chi^2(1) = -3.841$; $p < .01$). En CM2, cette proportion se stabiliserait puisqu'elle est de 62% ($\chi^2(1) = -2.270$; *NS.*).

Les **abstentions**, quant à elles, sont moins nombreuses en CP qu'en CM2 ($\chi^2(1) = 5.503$, $p < .01$) mais ne varient pas de façon significative d'une classe à l'autre. En effet, Alors qu'elles représentent 21% des erreurs sur les mots complexes en CP, elles sont de 17% en CE1 ($\chi^2(1) = 2.077$, *NS.*), 15% en CE2 ($\chi^2(1) = .867$, *NS.*), 11% en CM1 ($\chi^2(1) = 2.071$, *NS.*) et 10% en CM2 ($\chi^2(1) = .732$, *NS.*).

En **résumé**, ces résultats semblent donc aller dans le sens de nos suppositions puisqu'on observe une **diminution des erreurs de décodage** et une **augmentation des intentions lexicales**, augmentation plus prononcée si on y inclut les verbalisations de mots issues de processus non identifiés. En effet, si on ne tient compte que de la verbalisation produite (mot/non mot), alors notre supposition se confirme d'autant plus puisque la proportion d'erreurs de cette catégorie augmente d'année en année (*cf.* graphique 30).

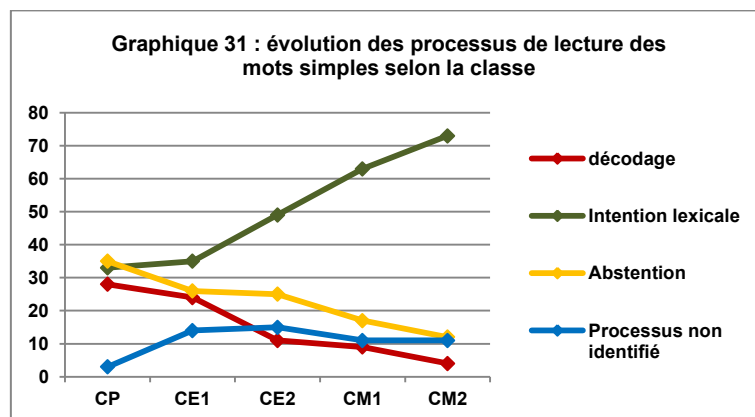


Ces résultats semblent donc aller avec l'idée qu'au début de l'apprentissage de la lecture, face à un mot complexe, les élèves commettent davantage d'erreurs de décodage dues à une maîtrise partielle des correspondances graphèmes-phonèmes. L'augmentation des erreurs lexicales semble donc confirmer la mise en place progressive d'une voie orthographique.

Comme dans l'analyse précédente, cette étude doit se compléter par une analyse des processus de lecture sur les mots simples.

Evolution des processus de lecture des mots simples

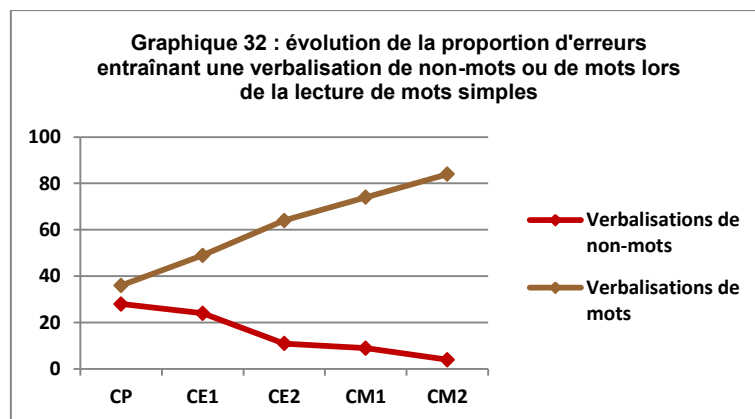
En effet, s'il existe un effet de la complexité graphémique, alors on devrait certes voir apparaître une diminution des erreurs de décodage et une augmentation des erreurs lexicales, mais dans de moindres proportions que pour les mots complexes. Ceci correspondant simplement à la maîtrise du code et à la mise en place d'une voie lexicale.



On remarque que pour les mots simples, la proportion d'**erreurs de décodage** suit la même tendance que pour les mots complexes. Alors qu'en CP, ce type d'erreur représente 28%, il diminue de 4 point en CE1 (24%) ($\chi^2(1) = -3.475, p < .05$). En CE2, cette proportion tombe à 11% ($\chi^2(1) = 3.278, p < .05$) et reste stable en CM1 (9%) ($\chi^2(1) = .437, NS.$) et en CM2 (4%) ($\chi^2(1) = 2.052, NS.$).

Les **erreurs lexicales**, quant à elles varient plus fortement. En effet, au début de l'école primaire, ce type d'erreur représente déjà 33% des erreurs commises sur les mots simples en CP et 35% en CE1 ($\chi^2(1) = -2.807, p = .05$). En CE2, on constate une forte augmentation puisqu'elles représentent 49% ($\chi^2(1) = -3.142, p < .05$). En CM1, 63% des erreurs sont des intentions lexicales ($\chi^2(1) = -3.761, p < .01$) et elles sont 73% en CM2 ($\chi^2(1) = -3.093, p < .05$).

En **résumé**, cette analyse des processus de lecture des mots simples semble confirmer un effet de la complexité graphémique en lecture. En effet, on constate que le taux d'erreur de décodage sur les mots complexes diminue au fur et à mesure de l'avancée dans la scolarisation. Certes cette diminution apparaît aussi lors de la lecture des mots simples, mais dans de moindres proportions. Parallèlement on observe une augmentation des erreurs lexicales que ce soit lors de la lecture de mots simples ou complexes.



Pour les mots simples aussi, on constate que si on considère l'ensemble des verbalisations de mots, alors elles augmentent d'année et année contrairement aux verbalisations de non-mots qui diminuent (*cf.* graphique 32)

CONCLUSION

L'objectif de ce chapitre était de voir s'il existe un effet de la structure graphémique en lecture. Autrement dit, nous nous sommes demandée si les mots contenant au moins un graphème complexe étaient plus difficiles à lire que les mots contenant des graphèmes simples. Les modèles dominants en lecture postulent que, lors de cet apprentissage, les élèves passeraient par une série d'étapes définies selon les différents aspects de la langue. De cette façon, les élèves traverseraient un stade dit « alphabétique » caractérisé par l'apprentissage du code. Ce n'est qu'au stade suivant que les élèves se centrerait sur la construction de sens, lors de la mise en place d'une voie orthographique. Ainsi, selon cette conception, l'apprentissage de la lecture se déroulerait de façon discontinue. Nous nous sommes donc demandée, **dans quelle mesure les erreurs liées à la structure graphémique d'un mot peuvent-elles rendre compte d'un apprentissage continu ou discontinu de la lecture.**

Nous avons d'abord supposé que le nombre d'erreurs commises sur les mots complexes serait plus important que celui sur les mots simples. Nous nous attendions à ce que la maîtrise du code fasse diminuer ce type d'erreur dans le temps.

Notre première analyse semble confirmer la première supposition. Ainsi, il apparaît que **le nombre d'erreurs est plus important lors de la lecture de mots complexes que de mots simples,** et ce quel que soit le niveau de scolarisation. De plus, on remarque que cet effet s'amointrit dans le temps puisque le nombre d'erreurs commises sur les mots complexes diminue de façon systématique d'une classe à l'autre. Toutefois, cette baisse constante n'est pas pour autant régulière. En effet, on constate qu'elle est plus marquée lors du passage en CE1. Cette rupture semble donc être le reflet d'un apprentissage discontinu de la lecture correspondant à un changement de stratégie pour lire les mots complexes.

Afin de préciser ce constat, nous avons par la suite mené une étude plus qualitative sur les processus de lecture des élèves. Ainsi, à partir des différents types d'erreurs, nous avons cherché à savoir **comment l'élève procède face à des mots contenant au moins un graphème complexe.** Nous avons répertorié quatre types de processus : le décodage conduisant les élèves à prononcer un non-mot, l'intention lexicale qui correspond à la verbalisation d'un mot, l'abstention et les processus non identifiés. Dans cette dernière catégorie, la verbalisation est un mot mais nous ne pouvons dire clairement si cette verbalisation provient d'une intention lexicale ou non. Nous avons supposé que, s'il existe un effet de la complexité graphémique, alors nous devrions trouver davantage d'erreurs de

décodage lors de la lecture de mots complexes que de mots simples. La maîtrise du code et la mise en place d'une voie orthographique devraient faire diminuer ce type d'erreur et faire augmenter le nombre d'erreurs lexicales dans le temps.

Cette seconde analyse nous permet de constater qu'effectivement, **les erreurs dues à un mauvais décodage sont plus nombreuses lors de la lecture de mots complexes que de mots simples quelle que soit la classe fréquentée.** Il semble ici, que la constance de ce type d'erreur laisse penser à un apprentissage continu de la lecture. Toutefois, si on considère les verbalisations issues de processus non identifiés comme étant des intentions lexicales, alors on remarque que cette catégorie devient le type d'erreur dominant à partir du CE2. Cette rupture laisserait donc penser à un apprentissage plutôt discontinu de la lecture.

La troisième analyse portant sur l'évolution des processus de lecture, nous permet de constater que les erreurs liées à un mauvais décodage commises sur les mots complexes diminuent entre le CP et le CM2. Ceci confirme donc le fait que la maîtrise du code tend à faire diminuer le nombre d'erreurs de décodage et réduit ainsi l'effet de complexité. Parallèlement, l'augmentation des intentions lexicales confirme, là aussi, la mise en place progressive d'une voie orthographique.

Ainsi, face à des mots complexes, les élèves, au début de l'apprentissage de la lecture ont davantage recours au décodage mais leur maîtrise partielle des correspondances graphèmes-phonèmes ne leur permet pas de lire sans erreur les mots contenant des graphèmes complexes. A partir du CE2, il semble que cette procédure soit maîtrisée puisqu'ils ont davantage tendance à verbaliser des mots. D'autre part, nous avons aussi constaté que face à un mot complexe, certains élèves s'abstiennent de lire. En effet, ces derniers préfèrent peut-être ne rien prononcer plutôt que de commettre une erreur de lecture. Ce constat est valable pour les élèves au début de l'apprentissage de la lecture puisque ce type d'erreur est celui qui est le plus important après les erreurs de décodage. Enfin, nous avons constaté qu'au début de l'apprentissage de la lecture, les erreurs lexicales, bien qu'elles soient moins nombreuses que les erreurs de décodage ou les abstentions, sont toutefois présentes. Cela signifie donc qu'**au-delà du décodage, les apprentis lecteurs cherchent à donner du sens à leur lecture.** Ce constat va à l'encontre des modèles développementaux qui postulent que la construction de sens n'interviendrait que plus tard dans cet apprentissage.

Nous devons par ailleurs préciser que cette étude a été menée dans des conditions textuelles de lecture. Par conséquent, d'autres variables peuvent entrer en jeu. En effet, il est possible que la fréquence ou la longueur des mots jouent aussi un rôle (*cf.* Chapitre 9

« *L'influence de la longueur et de la fréquence d'un mot* »). Pour pallier ces influences, il faudrait faire une expérimentation dans laquelle ces variables seraient contrôlées. Par exemple, il faudrait faire lire à des élèves un texte dans lequel on aurait auparavant contrôlé les variables longueur et fréquence. Autrement dit, faire lire des mots de longueur et de fréquence égales mais de complexité différente.

De plus, le test de *L'alouette* comporte un certain nombre d'*artefacts* (dessins, similitudes, co-texte sémantique) qui amèneraient les élèves à commettre des erreurs. Lors de notre analyse, nous avons constaté que la plupart des intentions lexicales et de mots issus des processus non identifiés peuvent être dus à un effet du test de Lefavrais. Ainsi, on se demande si face à un mot complexe, les élèves ont eu recours à ces faux indices ou si ces derniers n'ont fait que confirmer leur lecture.

Néanmoins, même si nous ne pouvons pas expliquer clairement les verbalisations de mots, nous pouvons toutefois conclure qu'au début de l'apprentissage de la lecture, les prononciations de logatomes sont plus nombreuses sur les mots complexes. En revanche, à partir du CM1, les élèves verbalisent davantage de mots que de non-mots que ce soit de manière intentionnelle ou non.

L'influence des variables linguistiques lors de l'apprentissage de la lecture

CONCLUSION

Dans la plupart des modèles reconnus, l'apprentissage de la lecture, considéré à travers les processus d'identification du mot, repose sur une progression prenant en compte les différents aspects de la langue. Cette progression est envisagée de manière discontinue dans la mesure où les enfants traverseraient une série d'étapes durant cet apprentissage. Lors de la première étape, l'identification du mot ne se ferait qu'à partir de sa forme globale. Ce n'est que dans la seconde étape, dite alphabétique, que les enfants se centreraient sur l'apprentissage du code. Et ce n'est que lors de l'étape orthographique que la recherche de sens interviendrait. Malgré la diversité des modèles, tous s'accordent sur l'importance de l'étape alphabétique.

Dans cette conception, le test de Lefavrais nous a semblé particulièrement adapté puisqu'il mesure essentiellement le niveau de décodage. Ainsi, **nous avons donc cherché à savoir dans quelle mesure le test de *L'alouette* peut nous renseigner sur l'apprentissage de la lecture et plus particulièrement de quelle façon il nous permet de discuter de la discontinuité de cet apprentissage.**

Nous avons supposé que, lors de la mise en place de la procédure alphabétique, certaines variables linguistiques, comme la longueur, la fréquence ou encore la structure graphémique d'un mot entraînent des difficultés.

Nous avons donc cherché à savoir dans quelle mesure les erreurs liées à ces variables linguistiques peuvent être considérées comme résultant d'un apprentissage continu ou discontinu de la lecture.

A cet effet, nous avons procédé de la même façon pour chaque variable. D'abord nous avons voulu savoir quelle est l'influence de ces variables lors de la lecture. Puis, nous avons cherché à savoir quels sont les processus hypothétiques de lecture lorsque l'enfant rencontre

un mot présentant les caractéristiques de ces variables. Enfin, nous avons regardé de quelle façon évoluent ces processus d'une classe à l'autre.

Notre étude sur la longueur et la fréquence d'un mot fait ressortir que les mots longs et peu fréquents (voire rares) entraînent plus d'erreurs que les mots courts fréquents (ou peu fréquents). Ainsi, à fréquence équivalente, les mots courts sont mieux lus que les mots longs quel que soit le cycle de scolarisation et, à longueur équivalente, les mots fréquents sont mieux lus que les mots rares. Ceci explique donc pourquoi les mots courts fréquents sont mieux lus que les mots longs rares. On remarque que les mots longs peu fréquents (ou rares) entraînent systématiquement plus d'erreurs que les mots courts fréquents (ou très fréquents) et ce quelle que soit la classe fréquentée. **La longueur et la fréquence d'un mot ont donc une influence lors de la lecture de mots.** De plus, conformément à nos attentes, **cet effet diminue entre le cycle 2 et cycle 3.** Toutefois, bien que cette baisse soit significative, elle ne se retrouve pas lors du passage d'une classe à une autre. Ce n'est qu'entre le CE2 et le CM1 que cette diminution est significative. **On constate donc une rupture** entre ces deux classes. Ce constat laisse penser que les élèves changeraient alors de stratégie pour lire ces mots. C'est ce que nous avons cherché à savoir en procédant à une analyse qualitative.

Nous avons supposé que s'il existe un effet de la fréquence et de la longueur d'un mot, alors nous devrions observer un plus grand nombre d'erreurs de décodage lors de la lecture de mots longs peu fréquents que de mots courts fréquents. Cet effet devant diminuer avec la maîtrise de la procédure alphabétique et les erreurs lexicales devant augmenter avec la mise en place de la procédure orthographique.

En effet, il apparaît que **les erreurs de décodage sont systématiquement plus nombreuses sur les mots longs peu fréquents que courts fréquents** que ce soit en CP ou en CM2. L'effet de longueur et de fréquence d'un mot lors de la lecture se confirme donc. En outre, il apparaît que, jusqu'en CM1, ce processus est celui qui est le plus utilisé par les élèves. Toutefois, si on compte l'ensemble des verbalisations de mots, qu'elles soient intentionnelles ou non, alors il s'avère que **dès le CE2, les élèves verbalisent plus de mots que de non mots lors de la lecture de mots longs peu fréquents.**

Les résultats sont à peu près les mêmes en ce qui concerne l'effet de la structure graphémique. Ainsi, on constate que **les mots contenant au moins un graphème complexe**

entraînent plus d'erreurs, que les élèves soient en cycle 2 ou 3 et plus précisément quelle que soit la classe fréquentée.

Comme pour l'effet de la longueur et de la fréquence, on remarque que **cet effet diminue d'un cycle à l'autre mais aussi d'année en année**. On observe donc une baisse constante mais qui n'est pas non plus ici aussi régulière. Il apparaît que l'effet de la structure graphémique diminue plus fortement entre le CP et le CE1. Cette forte diminution laisse envisager (comme pour la longueur et la fréquence d'un mot) une rupture synonyme d'un changement de stratégie pour traiter les mots contenant au moins un graphème complexe.

Ainsi, l'analyse qualitative menée sur les différents types de processus nous permet de confirmer un effet de la structure graphémique puisqu'à chaque niveau scolaire, **les erreurs sur les mots complexes sont plus nombreuses que celles sur les mots simples**. Bien que cet effet soit observable jusqu'en CM2, il s'avère que ce type de processus n'est pas systématiquement celui qui est le plus employé. En effet, on remarque que jusqu'en CM1, les mots complexes sont essentiellement traités par décodage. Toutefois si on compte les verbalisations de mots, alors **à partir du CE2, les élèves verbalisent plus de mots que de non mots**.

Ces deux analyses nous permettent de confirmer l'**existence d'un effet de longueur, fréquence et structure graphémique** lors de l'apprentissage de la lecture. Conformément à nos suppositions, cet effet diminue entre le CP et le CM2, ce qui correspondrait à la mise en place de la procédure alphabétique. Toutefois, nous avons remarqué que cette baisse n'est pas pour autant régulière et qu'à un certain moment, ce type d'erreur diminue plus fortement. Cette rupture dans la progression des élèves serait alors synonyme d'un changement de stratégie pour lire ces mots. En effet, on remarque que pour ces deux variables, les élèves effectuent jusqu'en CM1, essentiellement une procédure de décodage pour lire les mots les plus difficiles. Toutefois, si on tient compte de toutes les verbalisations de mots, qu'elles soient intentionnelles ou non, alors dès le CE2 les élèves verbalisent plus de mots.

Par conséquent, en accord avec les modèles développementaux, on constate une **discontinuité dans l'apprentissage** de la lecture qui est le reflet d'un changement de stratégie. Cependant, le fait que les erreurs liées à ces variables soient encore présentes en fin d'école primaire, nous permet de relativiser cette discontinuité. En effet, selon ces modèles l'enfant ne passerait à l'étape suivante que si les compétences de l'étape antérieure sont maîtrisées. Or il apparaît que ce n'est pas le cas puisqu'on voit que les erreurs de décodage

perdurent alors que les erreurs lexicales apparaissent. **La discontinuité envisagée par les modèles développementaux, n'est donc pas absolue.**

Par ailleurs, lorsqu'on observe les verbalisations de mots, on remarque que la plupart d'entre elles sont en lien avec les spécificités du test de *L'alouette* (amie/annie, pompe/pomme, geai/gai, écueil/écureuil, cordeaux/corbeaux, l'hirondeau/l'hirondelle, décochée/décrochée, ... etc). Certaines peuvent avoir été influencées par les dessins, d'autres par les similitudes ou encore par le co-texte sémantique. Il semblerait donc que face à des mots difficiles à lire, les élèves ont recours à d'autres indices, plus accessibles que le mot. C'est ce que nous allons étudier dans une troisième partie.

L'influence des variables liées au test de L'alouette

Plan

INTRODUCTION	229
SPECIFICITES DU TEST DE <i>L'ALOUETTE</i>	231
METHODOLOGIE D'ANALYSE	232
Chapitre 11 : L'influence des dessins lors de la lecture du texte de <i>L'alouette</i>	235
Chapitre 12 : L'influence des similitudes lors de la lecture du texte de <i>L'alouette</i>	259
Chapitre 13 : L'influence du co-texte sémantique lors de la lecture du texte de <i>L'alouette</i>	284
Chapitre 14 : L'influence des mots surdéterminés lors de la lecture du texte de <i>L'alouette</i>	303
CONCLUSION	320

INTRODUCTION

Dans les modèles de lecture les plus reconnus, l'acquisition de la lecture repose sur une progression prenant en compte les différents aspects de la langue. Cette progression est alors envisagée de manière discontinue puisque l'enfant traverse un certain nombre d'étapes lors de cet apprentissage. Par exemple, dans le modèle de Frith (1985), une fois passée l'étape logographique, l'enfant passerait par une étape caractérisée par l'apprentissage du code (étape alphabétique) et ce n'est qu'après cette étape que l'enfant se centrerait sur la construction de sens. Ainsi, dans cette conception de la lecture, cet apprentissage se déroulerait de manière discontinue.

L'objectif de cette partie est donc de savoir de quelle façon les variables liées au test de Lefavrais nous renseignent sur l'apprentissage de la lecture et plus précisément dans quelle mesure elles nous permettent de rendre compte d'un apprentissage continu ou discontinu de la lecture.


Lefavrais a construit le test de *L'alouette* sur une conception classique de la lecture, c'est-à-dire qu'il mesure essentiellement le décodage ; ce qui correspond dans les modèles développementaux à la phase alphabétique du modèle de Frith (1985). De ce point de vue, certaines erreurs apparaissent donc comme une défaillance dans les capacités de décodage.

L'avantage de l'utilisation du test de *L'alouette*, par rapport aux recherches menées sur l'apprentissage de la lecture, est qu'il présente les mots en contexte, ce qui rendrait disponible d'autres types d'indices pour identifier les mots.

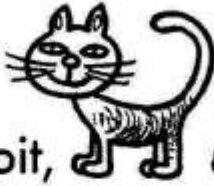
Nous supposons donc, qu'au début de l'apprentissage de la lecture, l'élève face à des mots qui présentent des difficultés de lecture, aura recours à d'autres indices plus accessibles que l'écrit.

Rappelons que le test de *L'alouette* se compose d'un texte de 265 mots, de caractères de taille décroissante, entouré de dessins dans les marges de gauche et droite. Les phrases vont d'une demi-ligne à deux lignes et demi. Dans le détail, on remarque que le test de *L'alouette* contient un certain nombre de spécificités dont nous essaierons de voir le rôle.

Cartographie 4 : test de L'alouette (Lefavrais, 1965)









L'alouette.




Sous la mousse ou sur le toit,
 dans les haies vives ou le chêne fourchu,
 le printemps a mis ses nids.
 Le printemps a nids au bois.

Annie amie, du renouveau, c'est le doux temps.
 Amie Annie, au bois joli gamine le pinson.
 Dans les buis, gîte une biche, au bois chantant.
 Annie, Annie! au doigt joli, une églantine laisse du sang :
 au bout du temps des féeries viendra l'ennui.






L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
 L'hirondeau piaille sous la pente des bardeaux et, vif et gai, le geai
 sur l'écaïlle argentée du bouleau, promène un brin d'osier.
 Au verger, dans le soleil matinal, goutte une pompe dégelée.
 On voit un bec luisant qui trille éperdument des notes claires
 et, dans les pampres d'or que suspend la grille antique,
 on surprend des rixes de moineaux.
 Au potager s'alignent les cordeaux; l'if est triste à l'horizon
 et lourd et lent l'envol des corbeaux.

Un lac étire ses calmes rives et, quand le soir descend,
 le miroir de ses eaux reflète les poisons des brignoles perfides.
 Et, quand descend le soir, quand joue la pourpre du couchant,
 le ciel rougit ses eaux.
 Dans la moire de l'eau danse l'ombre d'un écueil.
 Tout est cris! Tout est bruits!

Une amarre est décochée... une barque est arrimée... des matelots
 jettent leurs cassettes sur le rivage...
 Tout est cris! Tout est bruits!
 Au clair de la lune mon ami Pierrot...
 Au clair de lune mon amie annie...
 Au clair de la lune mon ami Pierrot, prête-moi la plume pour écrire un mot.







o u e i a

le la les un dans des do ti pu mi

SPECIFICITES DU TEST DE L'ALOUETTE

Bien que le texte ait l'aspect d'un texte réel, il contient des mots qui, mis les uns à côté des autres, forment des phrases qui sont, pour la plupart d'entre elles, asémantiques (« *alouette fait un nœud* », etc.). Lefavrais l'associe d'ailleurs à « *la structure d'un rêve et de ses incohérences* » (1967). Il ajoute que ce texte est comme « *un puzzle d'idées plaisantes ou déplaisantes, de rythmes, de consonances, d'images...* ».

En effet, il décrit deux idées principales. La première est celle de la nature au printemps, le temps s'écoule du matin jusqu'au soir et l'activité diminue : dans l'ordre du texte, on parle d'abord de l'arrivée du printemps (« *le printemps a mis ses nids* », « *du renouveau* »), puis du matin (« *Au verger, dans le soleil matinal* ») et de la nuit vers la fin du texte (« *et, quand le soir descend* »). La seconde idée est celle de la vie, de la naissance et de la joie de vivre (« *au bois chantant* », « *féeries* », « *gai* », etc.). L'auteur explique que cette dernière est parfois mise en opposition avec des peines : « *au doigt joli une églantine laisse du sang* », « *au bout du temps des féeries viendra l'ennui* ». Outre ces deux idées principales, on y trouve aussi plusieurs champs lexicaux. Il y a celui des oiseaux : « *pinson, alouette, hirondeau, geai, bec, moineaux, corbeaux* ». Puis celui des couleurs et des lumières « *argentée, soleil, luisant, claires, or, reflète, pourpre, rougit* ». Et, vers la fin du texte, le champ lexical de l'eau : « *lac, rives, eaux, amarre, barque, matelots, rivage* ».

Par ailleurs, comme le dit Lefavrais (*id.*), la présence de rythmes, de consonances et d'images dans ce texte, nous amène à penser à un poème. Ainsi, comme dans toute poésie, on retrouve une certaine musicalité due à la présence d'allitérations¹ (« *chêne fourchu* », « *et lourd et lent l'envol* ») et d'assonances² (« *sous la mousse ou sur* », « *au, hirondeau, bouleau, bardeaux* »). On trouve aussi des figures de style jouant sur le sens des mots à travers des images (« *le miroir de ses eaux* », « *quand joue la pourpre du couchant* », etc.) et des personnifications (« *une églantine laisse du sang* »)

D'autre part, des leurres (iconiques et sémantiques) ont été introduits afin de perturber le lecteur. Ainsi, à certains endroits, le mot à reconnaître n'est pas celui qu'on pourrait attendre. Par exemple, le dessin d'un corbeau est situé à proximité du mot « *cordeaux* ». Le champ sémantique de ce paragraphe laisserait d'ailleurs attendre le mot « *corbeaux* » plutôt que « *cordeaux* ». Autre exemple, le mot « *poisons* » situé juste après le mot « *lac* ». Dans ce

¹ Répétition d'une même sonorité consonantique

² Répétition d'une même sonorité vocalique

contexte, on s'attend plus facilement à lire « poissons » que « poisons ». A la fin du texte, on trouve aussi une inférence qui concerne la chanson « Au clair de la lune ». Les dessins du Pierrot et des notes de musique suggèrent au lecteur de lire « au clair de la lune » et non « au clair de lune » ou encore « prête-moi ta plume » au lieu de « prête-moi la plume ». Enfin, le texte contient aussi des similitudes : ce qui a l'air d'être pareil ne l'est pas forcément. Par exemple : « *Amie Annie, au bois joli* » et « *Annie, Annie ! au doigt joli* ».

Toutes ces caractéristiques réunies font que la lecture de ce texte présente un certain nombre de difficultés.

METHODOLOGIE D'ANALYSE

Pour répondre à notre objectif, nous avons procédé de la même façon que pour l'analyse de l'influence des variables linguistiques. Autrement dit, nous avons étudié la progression des élèves.

Ainsi, d'un point de vue quantitatif, nous parlerons de continuité lorsque la progression des élèves est régulière (la performance augmente progressivement et/ou les erreurs diminuent progressivement). Toute rupture dans cette progression sera alors considérée comme un signe de discontinuité.

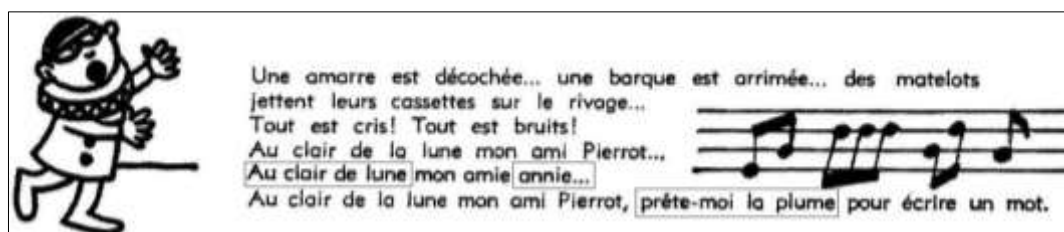
D'un point de vue qualitatif, nous dirons qu'il y a continuité lorsque les types d'erreurs restent les mêmes d'un moment à l'autre du cursus scolaire. Toute différence dans les types d'erreurs sera alors considérée comme un critère de discontinuité.

L'analyse quantitative se déroule en deux temps. Dans un premier temps, nous avons cherché à savoir ***quels sont les élèves les plus influencés et dans quelle proportion ont-ils été influencés ?*** Nous précisons que, pour cette analyse, nous avons utilisé comme indicateur le nombre d'élèves qui a commis au moins une erreur et non le nombre d'erreurs. En effet, il nous était impossible de travailler avec les taux d'erreurs étant donné que le nombre d'erreurs possibles n'est pas défini. Autrement, il peut y avoir ces erreurs sur n'importe quel mot du texte. Cette analyse a d'abord été effectuée pour chaque cycle (cycle 2 et cycle 3) puis de façon détaillée pour chaque niveau scolaire qui composent les cycles. Néanmoins, nous nous

sommes rendue compte qu'une telle analyse comportait un biais. En effet, les élèves disposaient de trois minutes pour lire le texte. De ce fait, le nombre de mots lus n'est pas équivalent pour tous : il est plus important chez les élèves plus âgés. Par conséquent, au risque de perdre de l'information, nous avons repris cette analyse en comparant ce qui est comparable, c'est-à-dire la partie du texte que tous les élèves ont lue en considérant le nombre moyen de mots lus des élèves les plus jeunes et donc les moins avancés en lecture : les élèves de CP. Ces derniers ont lu en moyenne 55 mots (sur 265 au total), ce qui correspond à 21% du texte. Nous avons donc repris cette analyse en ne tenant compte que du début du texte. Toutefois, cette étude restait incomplète puisque nous ne disposions pas de toute l'information (seulement le début du texte). C'est pourquoi, nous avons poursuivi cette étude en ne tenant compte, cette fois-ci, que de l'autre partie du texte, c'est-à-dire les 89% restant. Par conséquent, bien que l'analyse par cycle présente un biais, nous avons fait le choix de la garder pour avoir une vue d'ensemble. De ce fait, l'analyse quantitative présente d'abord l'influence de la variable étudiée pour l'ensemble des élèves en cycle 2 et cycle 3. Puis, l'influence cette variable dans le début du texte et le reste du texte.

La seconde analyse, d'ordre plus qualitatif, a été effectuée afin de voir **quels sont les types d'erreurs les plus induits par la variable du test étudiée**. Ainsi, à l'aide du nombre moyen de mots lus pour chaque niveau scolaire³, nous avons d'abord regardé quels sont les mots sur lesquels le plus d'élèves se sont trompés (c'est-à-dire la moitié ou plus) et quels sont les types d'erreurs les plus fréquemment commis sur ces mots.

Nous avons remarqué que certaines erreurs peuvent être dues à de multiples influences. C'est notamment le cas des trois dernières lignes du texte qui font fortement référence à la chanson « Au clair de la lune ». La présence des dessins du Pierrot et des notes de musique ne faisant qu'influencer davantage les enfants à lire « au clair de la lune » → *au clair de la lune*, « mon amie annie » → *mon ami Pierrot* et « prête-moi la plume » → *prête-moi ta plume*.



³ (CP : 55, CE1 : 154, CE2 : 197, CM1 : 234 et CM2 : 253)

Cette surdétermination constitue donc une variable parasite. C'est la raison pour laquelle nous avons repris ces analyses en ne tenant compte que des erreurs qui peuvent être dues à la variable du test étudiée. C'est donc ces analyses que nous présentons dans les chapitres à venir.

Chapitre 11



L'INFLUENCE DES DESSINS LORS DE LA LECTURE DU TEXTE DE *L'ALOUETTE*

Plan du chapitre

INTRODUCTION	236
PREDICTIONS	236
METHODOLOGIE D'ANALYSE	237
I. L'influence des dessins chez les élèves	239
<i>Quels sont les élèves les plus influencés par les dessins ?</i>	239
<i>Dans quelle proportion ont-ils été influencés ?</i>	239
1. L'influence des dessins en cycle 2	240
<i>Quels sont les élèves les plus influencés par les dessins en cycle 2?</i>	240
<i>Dans quelle proportion ont-ils été influencés ?</i>	240
2. L'influence des dessins en cycle 3	241
<i>Quels sont les élèves les plus influencés par les dessins en cycle 3 ?</i>	241
<i>Dans quelle proportion ont-ils été influencés ?</i>	242
II. L'influence des dessins dans le texte	243
1. L'influence des dessins dans le début du texte	243
<i>Quels sont les élèves les plus influencés par les dessins dans le début du texte?</i>	243
<i>Dans quelle proportion ont-ils été influencés ?</i>	244
2. L'influence des dessins dans le reste du texte	245
<i>Quels sont les élèves les plus influencés par les dessins dans le reste du texte?</i>	245
<i>Dans quelle proportion ont-ils été influencés ?</i>	246
III. Quels sont les types d'erreurs les plus induits par les dessins ?	247
<i>Le type d'erreur des lecteurs débutants</i>	248
<i>Les types d'erreurs des lecteurs plus avancés</i>	249
CONCLUSION	255

INTRODUCTION

Le test de *L'alouette* se compose d'un texte entouré de dessins. Leur disposition est telle qu'au moment de lire un mot l'élève peut être tenté de lire le dessin qui se situe à côté (ex. le dessin d'une hirondelle placé à côté du mot *hirondeau*). Généralement, les textes pour enfants sont illustrés d'images et la relation entre les deux peut prendre différentes formes. En effet, au-delà de leurs fonctions décoratives, les images peuvent jouer aussi un rôle¹. Dans la plupart des cas, elles sont en adéquation avec le texte et apportent souvent des informations complémentaires. Les écarts entre le texte et les images sont plus rares². Contrairement à la plupart des livres de jeunesse dans lesquels les images sont en adéquation avec le texte, celles qui entourent le texte de *L'alouette* ne le sont pas réellement. En effet, non seulement elles n'enrichissent pas le texte mais elles semblent être là pour détourner le lecteur.

Ainsi, nous nous demandons dans quelle mesure les dessins présents autour du texte de *L'alouette* influencent la lecture des élèves et dans quelle mesure ils nous renseignent sur l'apprentissage de la lecture.

PREDICTIONS

On se demande alors plus précisément, **dans quelle mesure les erreurs dues aux dessins présents autour du texte de *L'alouette* font apparaître un caractère discontinu ou continu de l'apprentissage de la lecture.**

Dans la conception sous-jacente au test de Lefavrais, les erreurs semblent être le reflet d'une insuffisance du traitement alphabétique. Dans ce contexte, le recours aux images pourrait apparaître comme une stratégie compensatoire.

Par conséquent, d'un point de vue quantitatif, nous supposons que l'apprentissage du code devrait faire diminuer ce type d'erreur dans le temps. Du point de vue qualitatif, nous nous attendons à ce que la nature des erreurs rende compte d'une évolution dans le rapport code/sens : erreurs de type lexical chez les plus jeunes et erreurs de type grapho-sémantique chez les plus âgés.

¹ Cf. modèle hypothétique de classification des relations texte-images de Barthes (1969) et Bourron (1980) : redondance, concurrence, complémentarité, inflation et imposture.

² « L'image, complice de l'apprenti lecteur? », H. Giguère (1991)

METHODOLOGIE D'ANALYSE




Le texte de *L'alouette* est illustré de 13 dessins représentant successivement une fourchette, un chat, une fleur, un nid rempli d'œufs, des corbeaux, une hirondelle, une pomme, une bobine de fil, un corbeau, la nuit, un Pierrot, un écureuil et des notes de musique.

Nous avons considéré que lorsqu'un élève a interprété un dessin au lieu de lire un mot, il a pu être influencé par celui-ci (ex. le mot « fourchu » lu *fourchette*).



Nous rappelons que la lecture de certains mots (*cf.* cartographie 5) peut être influencée de plusieurs façons. C'est pourquoi, nous avons gardé seulement les mots qui ne peuvent avoir subi qu'une seule influence.

Cartographie 5 : erreurs dues aux dessins présents autour du texte de L'alouette



L'alouette.





Sous la mousse ou sur le toit,
dans les haies vives ou le chêne fourchu,
le printemps a mis ses nids.
Le printemps a nids au bois.





Annie amie, du renouveau, c'est le doux temps.
Amie Annie au bois joli gamine le pinson.
Dans les buis gîte une biche, au bois chantant
Annie Annie! au doigt joli, une églantine laisse du sang :
au bout du temps des féeries viendra l'ennui





L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau piaille sous la pente des bordeaux et, vif et gai, le geai
sur l'écaille argentée du bouleau, promène un brin d'osier.
Au verger, dans le soleil matinal, goutte une pompe dégelée.
On voit un bec luisant qui trille éperdument des notes claires
et, dans les pampres d'or que suspend la grille antique,
on surprend des rixes de moineaux.
Au potager s'alignent les cordeaux. l'if est triste à l'horizon
et lourd et lent l'envol des corbeaux.



Un lac étire ses calmes rives et, quand le soir descend,
le miroir de ses eaux reflète les poisons des brignoles perfides.
Et, quand descend le soir, quand joue la pourpre du couchant,
le ciel rougit ses eaux.
Dans la moire de l'eau danse l'ombre d'un écueil
Tout est cris! Tout est bruits!



Une amarre est décochée... une barque est arrimée... des matelots
jettent leurs cassettes sur le rivage...
Tout est cris! Tout est bruits!
Au clair de la lune mon ami Pierrot...
Au clair de lune mon amie annie...
Au clair de la lune mon ami Pierrot, prête-moi la plume pour écrire un mat.



o u e i a
le la les un dans des do ti pu mi

Plusieurs influences possibles

I. L'influence des dessins chez les élèves

Quels sont les élèves les plus influencés par les dessins ?

Dans l'ensemble, il apparaît que 57% (soit 489 élèves sur 861) des élèves de notre échantillon semblent avoir été influencés au moins une fois par un dessin durant la lecture du texte (cf. tableau 20).

Tableau 20 : pourcentage et nombre d'élèves ayant été influencés ou non par un dessin selon le cycle de scolarisation

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
Cycle 2	57% (183)	43% (140)	100% (323)		
Cycle 3	35% (189)	65% (349)	100% (538)	37.241	.000
Ensemble	43% (372)	57% (489)	100% (861)		

Lecture : 57 % des élèves scolarisés en cycle 2 (soit 183 sur 323) n'ont commis aucune erreur due à la présence des dessins et 43 % en ont commis au moins une (soit 140 sur 323).

Mieux encore, il semble qu'il y ait davantage d'élèves en cycle 3 qu'en cycle 2 qui aient été influencés. En effet, alors qu'ils sont une majorité en cycle 2 à ne pas avoir commis d'erreur due aux dessins (57% des élèves), ils sont 65% en cycle 3. Ainsi, il semble donc que ce type d'erreur augmente avec le niveau scolaire.

Dans quelle proportion ont-ils été influencés ?

Rappelons que, sur les 861 élèves, 489 semblent avoir été influencés pendant leur lecture par les dessins présents autour du texte. On remarque que, parmi ces derniers, plus de la moitié, soit 60% (cf. tableau 21), a commis une seule erreur de ce type.

Tableau 21 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par un dessin selon le cycle de scolarisation

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
Cycle 2	66% (92)	34% (48)	100% (140)		
Cycle 3	58% (202)	42% (147)	100% (349)	2.558	.110
Ensemble	60% (294)	40% (195)	100% (489)		

Lecture : 66% des élèves scolarisés en cycle 2 (soit 92 sur 140) n'ont commis qu'une seule erreur due à la présence des dessins et 34% en ont commis plusieurs (soit 48 sur 140).

Ainsi, que les élèves soient scolarisés en cycle 2 ou cycle 3, ils ont tendance à ne commettre qu'une seule erreur de ce type (66% en cycle 2 et 58% en cycle 3).

Par conséquent, bien que les élèves en cycle 3 soient plus nombreux à avoir commis des erreurs de ce type, ils n'en font pas pour autant davantage que ceux du cycle 2.

Voyons maintenant dans le détail ce qu'il en est pour les élèves en cycle 2 et en cycle 3. Les tendances sont-elles les mêmes d'une classe à l'autre?

1. L'influence des dessins en cycle 2

Parmi les 861 élèves de notre échantillon, 323 sont scolarisés en cycle 2 dont 143 sont en CP et 180 en CE1.

Quels sont les élèves les plus influencés par les dessins en cycle 2?

Bien qu'en cycle 2, les élèves qui n'ont pas été influencés soient plus nombreux que ceux qui l'ont été (57% contre 43%), on constate que cette tendance ne se retrouve pas lorsqu'on regarde les résultats par classe (cf. tableau 22).

Tableau 22 : pourcentage et nombre d'élèves ayant été influencés ou non par un dessin selon la classe en cycle 2

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CP	73% (104)	27% (39)	100% (143)		
CE1	44% (79)	56% (101)	100% (180)	25.827	.000
Ensemble	57% (183)	43% (140)	100% (323)		

Lecture : 73 % des élèves scolarisés en CP (soit 104 sur 143) n'ont commis aucune erreur due à la présence des dessins et 27 % en ont commis au moins une (soit 39 sur 143).

En effet, alors que les CP ne sont majoritairement pas influencés (73%), il apparaît que c'est l'inverse pour les CE1 (56% d'entre eux ont commis au moins une erreur). Ce constat semble donc faire apparaître une rupture dans la progression des élèves, puisque ce sont les élèves plus âgés qui sont les plus nombreux à avoir commis ce type d'erreur.

Dans quelle proportion ont-ils été influencés ?

On remarque que, dans l'ensemble, les élèves en cycle 2 sont une majorité (66%) à avoir été influencés une seule fois par un dessin (cf. tableau 23).

Tableau 23 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par un dessin selon la classe en cycle 2

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CP	69% (27)	31% (12)	100% (39)	.120	.586
CE1	64% (65)	36% (36)	100% (101)		
Ensemble	66% (92)	34% (48)	100% (140)		

Lecture : 69% des élèves scolarisés en CP (soit 27 sur 39) n'ont commis qu'une seule erreur due à la présence des dessins et 31% en ont commis plusieurs (soit 12 sur 39).

En effet, les élèves qui ont donné ce type de réponse ont essentiellement été influencés une seule fois par les dessins quelle que soit leur classe de scolarisation : ils sont 69% en CP et 64% en CE1.

En conséquence, bien que, d'une manière générale, les élèves en cycle 2 qui n'ont pas été influencés par les dessins soient plus nombreux que ceux qui ne l'ont pas été, on constate que cette tendance ne se retrouve pas au sein de chaque classe. Ainsi, lorsqu'on compare les élèves de CP et de CE1, on s'aperçoit que ce sont les plus jeunes, et donc les moins avancés en lecture, qui semblent être le moins influencés par la présence des dessins autour du texte. Ce constat nous amène à penser que les dessins n'influenceraient pas essentiellement les apprentis lecteurs puisque ce type d'erreur concerne davantage les élèves en CE1. Suivant cette logique, l'étude de l'influence des dessins en cycle 3 devrait alors faire apparaître que ce sont les élèves les plus âgés qui sont le plus influencés.

2. L'influence des dessins en cycle 3

Quels sont les élèves les plus influencés par les dessins en cycle 3 ?

Alors que les élèves en cycle 2 sont 43% à avoir été influencés au moins une fois par les dessins, ils sont 65% en cycle 3 (cf. tableau 24).

Tableau 24 : pourcentage et nombre d'élèves ayant été influencés ou non par un dessin selon la classe en cycle 3

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CE2	33% (57)	67% (117)	100% (174)	.701	.704
CM1	36% (67)	64% (121)	100% (188)		
CM2	37% (65)	63% (111)	100% (176)		
Ensemble	35% (189)	65% (349)	100% (538)		

Lecture : 33 % des élèves scolarisés en CE2 (soit 57 sur 174) n'ont commis aucune erreur due à la présence des dessins et 67 % en ont commis au moins une (soit 117 sur 174).

Dans le détail, on remarque que leur nombre reste stable entre le CE2 et le CM2 : ils sont 67% en CE2, 64% en CM1 et 63% en CM2. Ainsi, au-delà du fait que l'utilisation des dessins ne serait pas une caractéristique des lecteurs débutants, ces résultats font apparaître une amplification de ce phénomène jusqu'en fin d'école primaire.

Dans quelle proportion ont-ils été influencés ?

On remarque que, comme en cycle 2, les élèves qui se sont trompés sont à chaque fois une majorité à avoir commis une seule erreur de ce type (cf. tableau 25).

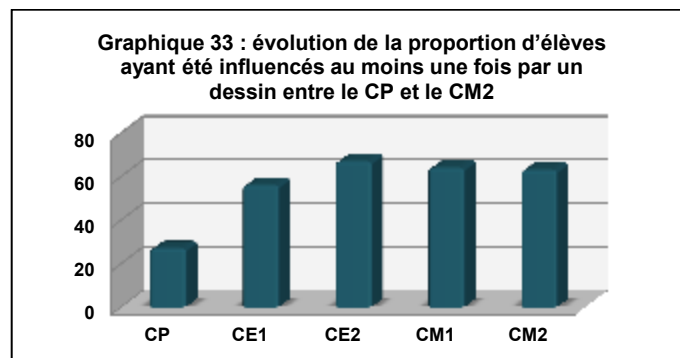
Tableau 25 : pourcentage (et nombre) d'élèves ayant été influencés une ou plusieurs fois par un dessin selon la classe en cycle 3

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CE2	60% (70)	40% (47)	100% (117)	.276	.871
CM1	57% (69)	43% (52)	100% (121)		
CM2	57% (63)	43% (48)	100% (111)		
Ensemble	58% (202)	42% (147)	100% (349)		

Lecture : 60 % des élèves scolarisés CE2 qui ont commis une erreur (soit 70 sur 117) n'ont commis qu'une seule erreur due à la présence des dessins et 40 % en ont commis plusieurs (soit 47 sur 117).

En effet, on remarque qu'en CE2, ils sont 60% à avoir commis une seule erreur. En CM1, leur nombre varie peu puisqu'ils sont 57% comme en CM2.

En **conclusion**, bien que les dessins du test de *L'alouette* aient influencé la lecture des élèves du CP au CM2, il apparaît que ce phénomène varie selon la classe (cf. graphique 33). En effet, alors que les élèves en CP à avoir eu ce comportement relatif aux images ne sont pas majoritaires (un peu plus du quart), ceux des autres niveaux sont plus de la moitié (cf. graphique 33).



On observe donc bien ici une discontinuité quantitative (rupture de progression entre les CP et les autres élèves) liée à l'influence des dessins. Discontinuité certes, mais qui ne va pas dans le sens de nos suppositions de départ. En effet, partant d'une conception traditionnelle de l'acquisition de la lecture, nous nous attendions à ce que ce soient les élèves les moins avancés en lecture qui utilisent les dessins. Or, cette analyse nous montre que c'est l'inverse : ce phénomène concerne davantage les élèves qui ont un niveau plus avancé en lecture. Plus intéressant encore, l'utilisation des dessins augmente avec le niveau de scolarisation : les élèves en cycle 3 commettent davantage d'erreur de ce type que les élèves en cycle 2. Il semble donc que, les erreurs liées aux images ne traduiraient pas seulement une défaillance du décodage puisque ce ne sont pas les apprentis lecteurs qui utilisent le plus les dessins mais au contraire, les lecteurs les plus avancés.

Cependant, nous devons rappeler que les élèves n'ont pas tous lu le même nombre de mots (*cf. Méthodologie*). C'est pourquoi, pour plus de rigueur, nous avons repris cette analyse en ne tenant compte que de la partie du texte lue par tous les élèves. Pour ce faire, nous avons considéré le nombre moyen de mots lus des élèves les plus jeunes et donc les moins avancés en lecture : les élèves de CP. Ces derniers ont lu en moyenne 55 mots (sur 265 au total), ce qui correspond à 21% du texte, soit les deux premiers paragraphes du texte. Nous nous sommes donc demandée dans quelle mesure les dessins ont influencé les élèves du CP au CM2 lors de la lecture du début du texte de *L'alouette*.

II. L'influence des dessins dans le texte

1. L'influence des dessins dans le début du texte

Quels sont les élèves les plus influencés par les dessins dans le début du texte?

Lorsqu'on analyse la partie du texte lue par la majorité des élèves, les résultats ne sont plus tout à fait les mêmes. En effet, de façon générale, 97% des élèves ne semblent pas avoir eu recours aux dessins lors de la lecture du début du texte (*cf. tableau 26*).

Tableau 26 : pourcentage et nombre d'élèves ayant été influencés ou non par un dessin dans le début du texte selon la classe

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CP (N=143)	82% (117)	18% (26)	100% (143)	122.015	.000
CE1 (N=180)	99% (178)	1% (2)	100% (180)		
CE2 (N=174)	100% (174)	0% (0)	100% (174)		
CM1 (N=188)	100% (188)	0% (0)	100% (188)		
CM2 (N=176)	100% (176)	0% (0)	100% (176)		
Ensemble (N=861)	97% (833)	3% (28)	100% (861)		

Lecture : 82 % des élèves scolarisés CP (soit 17 sur 143) n'ont commis aucune erreur due à la présence des dessins et 18 % en ont commis au moins une (soit 26 sur 143).

Néanmoins, bien que les élèves qui n'ont pas commis d'erreurs lors de la lecture du début du texte soient, dans l'ensemble, majoritaires, on remarque que parmi ceux qui se sont trompés, ce sont les élèves de CP qui sont les plus nombreux (18%). On retrouve ici, la rupture entre les élèves de CP et les autres mais qui se traduit de façon inverse : ce sont les élèves les plus jeunes qui sont les plus nombreux à avoir commis ce type d'erreur dans le début du texte.

Dans quelle proportion ont-ils été influencés ?

Sur les 861 élèves, 28 d'entre eux ont commis au moins une erreur due à un dessin. Parmi eux, 75% (soit 20) ont utilisé une seule fois les dessins et 25% (soit 8) les ont utilisés plusieurs fois (*cf.* tableau 27).

Tableau 27 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par un dessin dans le début du texte selon la classe

	Une erreur	Plusieurs erreurs	Total	χ^2	p.
CP (N=143)	77% (20)	23% (6)	100% (26)	.718	.397
CE1 (N=180)	50% (1)	50% (1)	100% (2)		
Ensemble (N=861)	75% (21)	25% (7)	100% (28)		

Lecture : 77 % des élèves scolarisés en CP qui ont commis une erreur dans le début du texte (soit 20 sur 26) n'ont commis qu'une seule erreur due à la présence des dessins et 23 % en ont commis plusieurs (soit 6 sur 26).

Rappelons que ce sont les élèves de CP qui ont le plus utilisé les dessins pour lire cette partie du texte. Nous ne discuterons pas ici des résultats, compte tenu des faibles effectifs en CE1 (N=2) et CM2 (N=1).

En **conclusion**, on remarque que lorsqu'on ne tient compte que de la partie du texte lue par tous les élèves, les résultats changent. En effet, il apparaît que ce sont les élèves les plus jeunes, et donc les moins avancés en lecture qui sont les plus nombreux à commettre une erreur en lien avec les dessins. Toutefois, cette analyse reste incomplète puisque nous ne

disposons pas de toute l'information (seulement le début du texte). C'est pourquoi, nous avons poursuivi cette étude en ne tenant compte cette fois-ci que de l'autre partie du texte, c'est-à-dire les 88% restant.

2. L'influence des dessins dans le reste du texte

Parmi les 861 élèves, 193 ont lu le texte en entier. On constate qu'aucun CP n'a lu tous les mots du texte. Cette analyse porte donc sur les élèves plus avancés en lecture, c'est-à-dire ceux scolarisés du CE1 au CM2.

Quels sont les élèves les plus influencés par les dessins dans le reste du texte?

Alors que dans le début du texte, seul 18% des CP semblent avoir commis au moins une erreur en lien avec les dessins, il apparaît que lorsqu'on analyse l'autre partie du texte, les élèves entre le CE1 et le CM2 sont 62% à avoir eu recours au moins une fois aux dessins (cf. tableau 28).

Tableau 28 : pourcentage et nombre d'élèves ayant été influencés ou non par un dessin dans le reste du texte selon la classe

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CE1 (N=180)	17% (1)	84% (5)	100% (6)	1.507	.681
CE2 (N=174)	37% (9)	63% (15)	100% (24)		
CM1 (N=188)	41% (28)	59% (40)	100% (68)		
CM2 (N=176)	37% (35)	63% (60)	100% (95)		
Ensemble (N=861)	38% (73)	62% (120)	100% (193)		

Lecture : 17 % des élèves scolarisés en CE1 (soit 1 sur 6) n'ont commis aucune erreur due à la présence des dessins et 84 % en ont commis au moins une (soit 5 sur 6).

On observe que, malgré des variations de pourcentages, il n'existe pas de différence significative entre les classes. En effet, alors que 84% des élèves de CE1 qui ont lu le texte en entier ont fourni au moins une fois ce type de réponse, ils sont 63% en CE2, 59% en CM1 et 69% en CM2. Bien que les élèves de CE1 aient l'air d'être plus nombreux à avoir commis au moins une erreur, ils ne sont pas significativement différents des autres. Cette absence de relation s'explique sans doute par le fait qu'ils soient peu nombreux (N=6).

Par conséquent, si on ne tient pas compte des CP, il apparaît donc une certaine régularité entre le CE1 et le CM2 puisqu'ils sont une majorité à avoir commis au moins une erreur de ce type du CE1 au CM2.

Dans quelle proportion ont-ils été influencés ?

Il apparaît que les lecteurs plus avancés, comme les lecteurs débutants, sont une majorité à donner une seule fois ce type de réponse.

Tableau 29 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par un dessin dans le reste du texte selon la classe

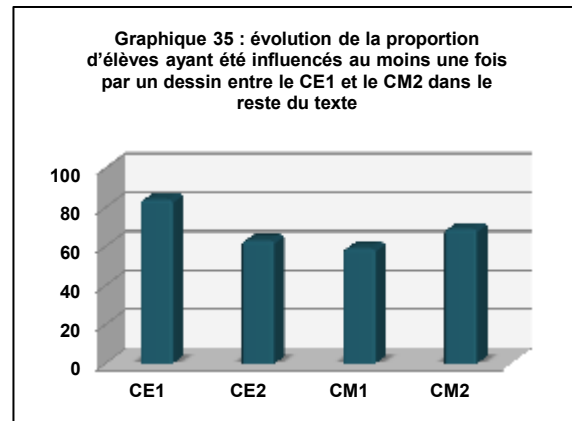
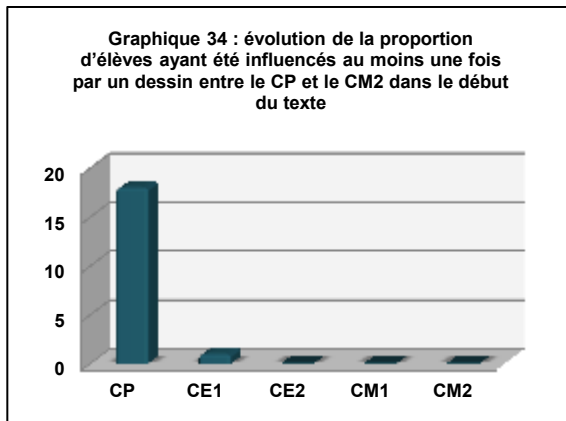
	Une erreur	Plusieurs erreurs	Total	χ^2	<i>p.</i>
CE1 (N=180)	80% (4)	20% (1)	100% (5)	.725	.867
CE2 (N=174)	60% (9)	40% (6)	100% (15)		
CM1 (N=188)	62% (25)	38% (15)	100% (40)		
CM2 (N=176)	65% (39)	35% (21)	100% (60)		
Ensemble (N=861)	64% (77)	36% (43)	100% (120)		

Lecture : 80 % des élèves scolarisés en CE1 qui ont commis une erreur (soit 4 sur 5) n'ont commis qu'une seule erreur due à la présence des dessins et 20% en ont commis au moins une (soit 1 sur 5).

En effet, 80% des élèves de CE1 qui se sont trompés ont commis une seule erreur. Ils sont 60% en CE2, 62% en CM1 et 65% en CM2. Il semble donc qu'à un niveau plus avancé en lecture, les élèves ne sont pas plus nombreux à utiliser plusieurs dessins.

En **conclusion**, ces différentes analyses nous permettent de constater qu'il existe une discontinuité quantitative dans l'utilisation des dessins entre le CP et le CM2. En effet, contrairement à ce que nous avons supposé au départ, ce phénomène ne concerne pas essentiellement les élèves les plus jeunes mais plutôt ceux qui ont un niveau avancé en lecture. Fait d'autant plus intéressant que ce type de réponse concerne essentiellement les élèves scolarisés entre le CE1 et CM2.

En effet, il apparaît que les résultats varient selon que l'on considère le début du texte ou le reste du texte. Ainsi, dans le début du texte, bien que les élèves qui ont commis ce type d'erreur ne soient pas majoritaires en CP, ils restent toutefois les plus nombreux à avoir été influencés par les dessins par rapport aux élèves des autres classes (*cf.* graphique 34). En revanche, dans le reste du texte, on constate que les élèves plus âgés ont quasiment tous été influencés par les dessins lors de leur lecture (*cf.* graphique 35).



Cette seconde analyse semble donc confirmer l'idée d'une rupture entre les élèves de CP et les autres en ce qui concerne leur progression dans l'apprentissage de la lecture. Ici, il semble que les erreurs liées aux dessins soient le reflet d'une certaine discontinuité dans l'utilisation des dessins lors de la lecture du texte de *L'alouette* qui serait non pas une caractéristique des apprentis lecteurs mais plutôt une spécificité des « bons lecteurs ». Autrement dit, il semblerait que le recours aux dessins ne soit pas seulement une stratégie compensatoire des lecteurs débutants. Il se pourrait que les lecteurs plus experts mettent en place une autre stratégie. C'est ce que nous nous proposons d'étudier dans une troisième analyse : à partir d'une approche qualitative des types d'erreurs commises, nous essaierons de déterminer les stratégies de lecture pour chaque niveau scolaire.

III. Quels sont les types d'erreurs les plus induits par les dessins ?

Si les dessins ont une influence lors de la lecture du texte de *L'alouette*, alors nous nous attendons à ce que les élèves les plus jeunes, et par conséquent les moins avancés en lecture, commettent des erreurs lexicales et les plus âgés des erreurs grapho-sémantiques.


Ainsi, nous avons regardé, en fonction du nombre moyen de mots lus³, quelles sont les erreurs les plus fréquemment commises, c'est-à-dire celles où plus de 10% des élèves se sont trompés. Par exemple, les élèves de CP ont lu en moyenne 55 mots et il semble qu'ils aient utilisé un dessin pour la lecture de l'un d'entre eux (au moins 10% des élèves ont commis une erreur imputable aux dessins présents autour du texte). Cette analyse a été reprise à chaque niveau scolaire. Suite à cet inventaire, nous avons remarqué que la lecture de six mots semble

³ Rappel du nombre moyen de mots lus selon la classe : CP : 55 mots, CE1 : 154, CE2 : 197, CM1 : 234 et CM2 : 253.

avoir été particulièrement influencée par les dessins présents autour du texte de *L'alouette* (cf. tableau 30).

Tableau 30 : mots sur lesquels le plus grand pourcentage d'élèves (nombre) a commis une erreur due aux dessins selon le niveau de scolarisation

N° mots	Début du texte	Reste du texte				
	10	70	84	114	147	206
Mots	haies	l'ennui	L'hirondeau	pompe	cordeaux	écueil
CP (N=143)	13% (12)	16% (7)	19% (5)	64% (9)	0%	0%
CE1 (N=180)	0%	31% (51)	12% (19)	30% (38)	16% (14)	34% (13)
CE2 (N=174)	0%	18% (32)	18% (31)	30% (46)	30% (43)	28% (22)
CM1 (N=188)	0%	16% (31)	14% (26)	22% (39)	31% (55)	21% (30)
CM2 (N=176)	0%	12% (21)	16% (28)	20% (34)	33% (58)	20% (32)
Ensemble (N=861)	2% (12)	19% (142)	15% (109)	26% (166)	29% (170)	23% (97)

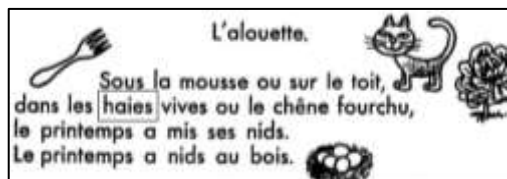
 Mots lus par moins de la moitié des élèves

Lecture : En CP, 13% des élèves qui ont lu le mot « haies » ont commis une erreur en lien avec un dessin.

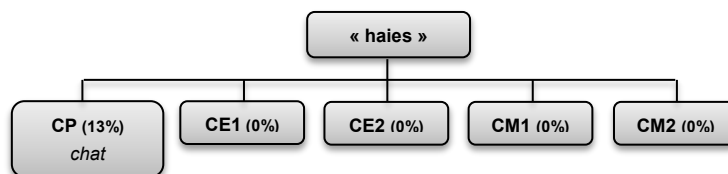
On remarque que les mots dont la lecture erronée est liée aux dessins ne sont pas les mêmes selon le niveau de scolarisation des élèves. En effet, alors que pour les lecteurs débutants (ceux qui n'ont lu que le début du texte) la lecture d'un mot (« haies ») semble avoir été influencée par les dessins, il apparaît qu'à des niveaux plus avancés en lecture, ce sont d'autres mots qui semblent avoir posé des difficultés.

Le type d'erreur des lecteurs débutants

Rappelons que les lecteurs débutants, c'est-à-dire les élèves de CP, ont lu en moyenne 55 mots. Ainsi, il apparaît que dans cette partie du texte, la lecture d'un mot semble avoir été particulièrement influencée par les dessins. Il s'agit du mot « haies » situé dans la seconde phrase du texte.



Parmi les 91 élèves de CP qui ont lu le mot « haies », 12 d'entre eux, soit 13%, ont prononcé *chat*. On constate que ce type d'erreur n'apparaît qu'à ce niveau scolaire.



Le mot « haies » présente un certain nombre de difficultés de lecture puisque ce monosyllabique se compose de cinq lettres qui ne forment qu'un son [ɛ]. La première difficulté vient de l'attaque (« h ») qui ne se prononce pas. Ensuite, le noyau du mot est formé par trois voyelles (« aie »). Enfin, le « s » qui marque le pluriel à la fin du mot peut aussi être une source de difficulté.

Il semble que les apprentis-lecteurs aient produit leur réponse à partir de quelques indices graphiques, notamment le début du mot (*-ha* que l'on retrouve dans *chat*) et se soient aidés du dessin pour produire un mot qui leur est plus familier : *chat*. L'utilisation du dessin semble apparaître alors ici comme une stratégie compensatoire dans l'identification du mot.

Par ailleurs, on remarque que le mot prononcé est syntaxiquement acceptable mais sans rapport avec le contexte sémantique. Ainsi, au-delà de l'identification du mot, ce constat nous laisse penser que les lecteurs débutants chercheraient à donner du sens à leur lecture mais de façon locale, mot par mot, sans tenir compte du co-texte sémantique de la phrase.

Qu'en est-il pour les lecteurs plus avancés? Quels sont les mots dont la lecture semble avoir été influencée par les dessins?

Les types d'erreurs des lecteurs plus avancés

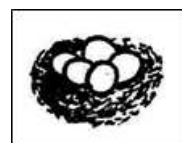
En ce qui concerne les lecteurs plus avancés, on remarque que la lecture de sept mots semble avoir été influencée par les dessins : « l'ennui, pompe, cordeaux, l'hirondeau et écueil ». Toutefois, on constate que ce type d'erreur ne touche pas la même proportion d'élèves selon leur niveau en lecture. Ainsi, alors que pour certains mots, le recours aux dessins est décroissant, pour d'autres, il est plutôt croissant.

➤ Les mots entraînant une utilisation décroissante des dessins : l'ennui et pompe

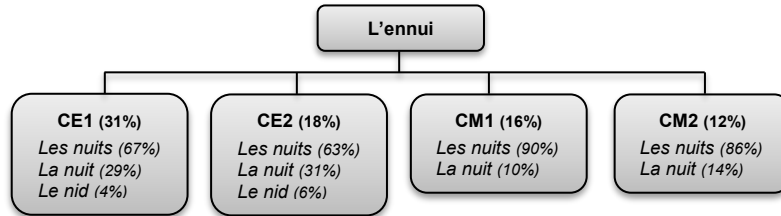
Dans l'ordre du texte, le premier mot concerné est « **l'ennui** » situé à la fin du second paragraphe du texte. De façon générale 19% des élèves (soit 142 sur 861) qui ont lu ce mot semblent avoir été influencés par les dessins.



Annie amie, du renouveau, c'est le doux temps.
Amie Annie, au bois joli gamine le pinson.
Dans les buis, gîte une biche, au bois chantant.
Annie, Annie! au doigt joli, une églantine laisse du sang :
au bout du temps des féeries viendra l'ennui.



On observe que 31% des élèves en CE1 semblent avoir été sensibles aux dessins (cf. schéma ci-dessous). En CE2, leur nombre diminue quasiment de moitié puisqu'ils sont 18%. En CM1, ils sont 16% puis 12% en CM2.

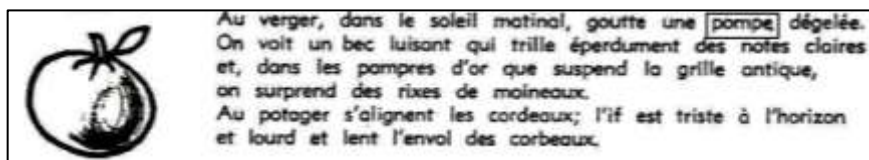


Bien que le dessin de la nuit soit situé deux paragraphes plus bas, ce mot a été lu essentiellement *les nuits* ou *la nuit*. En effet, lors des passations, nous avons laissé les élèves regarder le texte avant de les faire lire (cf. *Méthodologie*). Ce temps de latence a ainsi pu leur permettre d'assimiler ce dessin et par conséquent, les induire en erreur. Ils ont aussi lu parfois *le nid*. Deux dessins ont donc été utilisés pour la lecture de ce mot : celle de la nuit et celle du nid. Les difficultés de ce mot sont l'article élide *l'*, la double consonne *n* et la fin du mot composée de deux voyelles faisant le son [uj].

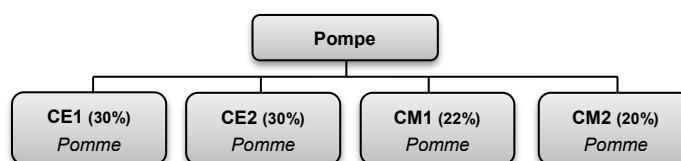
Lorsqu'on observe les différents types d'erreurs pour chaque niveau scolaire, on s'aperçoit qu'ils sont plus diversifiés pour les élèves de cours élémentaire que pour ceux des cours moyens. En effet, l'essentiel d'entre eux ont utilisé le dessin de la nuit, puisqu'ils sont plus de la moitié à avoir lu *les nuits* (67% pour les CE1 et 63% pour les CE2) et presque un tiers à avoir lu *la nuit* (29% en CE1 et 31% en CE2). Quelques élèves ont utilisé le dessin du nid puisqu'ils ont prononcé *le nid* (4% en CE1 et 6% en CE2). Néanmoins, la prononciation *les nuits* reste majoritaire du CE1 au CM2. Il se pourrait que l'article élide *l'* et le double *n* aient amené les élèves à effectuer un mauvais découpage du mot entraînant ainsi une resyllabation du mot. Toutefois nous pensons que les causes de cette erreur ne sont pas les mêmes pour les lecteurs plus avancés en lecture. Il est possible que les éléments qui précèdent (« des fêtes ») les aient influencés. Il se pourrait que ces élèves, parce qu'ils ont un meilleur niveau de lecture, lisent plus vite. Ainsi pris dans leur lecture rapide, ils ont peut-être effectué une rime en mettant « l'ennui » au pluriel comme l'antécédent : « [...] des fêtes viendra *les nuits* ».

Par ailleurs, contrairement aux élèves de CP, on remarque ici que les erreurs les plus commises sur ce mot sont, non seulement cohérentes avec la syntaxe, mais elles sont aussi cohérentes d'un point de vue sémantique (« [...] *au bout du temps* [...] viendra *les nuits/la nuit* »).

Parmi les erreurs qui diminuent, on trouve le mot « **pompe** » dont la lecture semble avoir été influencée par une image pour un grand nombre d'élèves (26%, soit 166 sur 861), quelle que soit la classe fréquentée.



Ainsi, en CE1 et CE2, ils sont 30% à avoir lu *pomme*. A partir du CM1, leur nombre diminue mais ils sont encore 22% à avoir commis cette erreur, puis 20% en CM2.



Hormis le son [ɔ̃], ce mot ne semble pas présenter de difficultés particulières. Pourtant, les élèves sont un tiers en cours élémentaire et un peu moins d'un quart en cours moyen à avoir commis ce type d'erreur. On constate que le mot prononcé est proche graphiquement du mot du texte : seule une lettre est différente. Il est possible que cette ressemblance grapho-phonémique et la présence du dessin les aient amenés à prononcer un mot qui leur est plus familier.

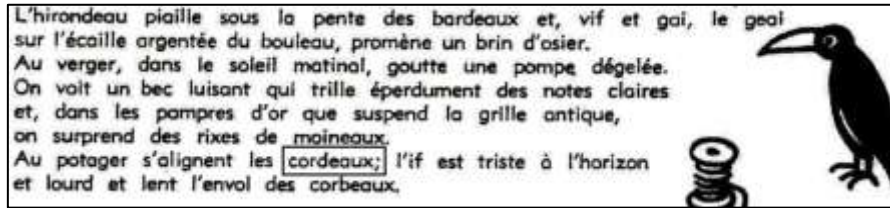
Comme pour les erreurs précédentes, on remarque que le mot prononcé respecte la syntaxe.

En dehors du fait que ces deux types d'erreurs semblent diminuer au fur et à mesure de l'apprentissage de la lecture, on constate qu'ils restent encore particulièrement présents en fin d'école primaire (12% pour « l'ennui » en CM2, 20% pour « pompe » en CM2).

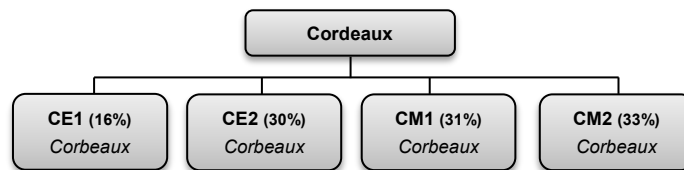
Alors que, pour ces mots, la proportion d'élèves qui ont été influencés par une image tend à diminuer, on remarque que pour un autre, elle augmente.

➤ Le mot entraînant une utilisation croissante d'un dessin : cordeaux

Le mot « **cordeaux** » est situé au début de la seconde moitié du texte. Dans l'ensemble, 29% des élèves qui l'ont lu semble avoir été influencés par un dessin.



Il a été lu *corbeaux* par 16% des élèves de CE1. En CE2, leur nombre augmente puisqu'ils représentent 30% des élèves, tout comme en CM1 (31%). En CM2, cette proportion varie légèrement puisqu'ils sont 33%. Autrement dit, il semblerait que les élèves aient eu recours au dessin, qu'ils soient à un niveau intermédiaire de l'apprentissage de la lecture ou à un niveau plus avancé.



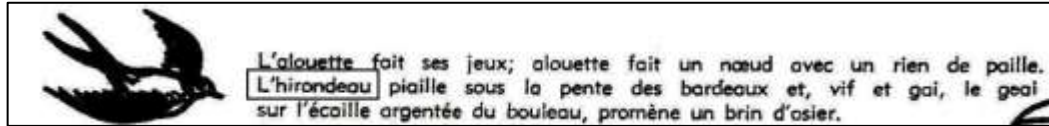
Il est possible que les causes de cette erreur ne soient pas les mêmes pour les CE1 et les autres. En effet, il se pourrait que les premiers aient confondus la lettre « d » avec la lettre « b ». En revanche, pour les élèves en cycle 3, il est peu probable que cette erreur puisse s'expliquer de la sorte. Comme pour le mot « pompe » on remarque que le mot prononcé présente une forte ressemblance grapho-phonémique. Ceci nous amène donc à penser que, là aussi, la rapidité de lecture associée à la prise d'indices graphiques les ait amenés à prononcer un mot qu'ils connaissent : *corbeaux*.

De plus, comme pour les erreurs précédentes, le mot prononcé est en cohérence avec la syntaxe et peut aussi s'inscrire dans le contexte sémantique.

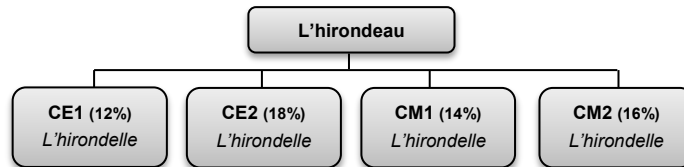
Au-delà des erreurs qui diminuent ou augmentent avec le niveau scolaire, d'autres, restent plutôt constantes entre le CE1 et le CE2.

➤ Les mots entraînant une utilisation constante des dessins : l'hirondeau et écueil

Le premier est le mot « **hirondeau** » situé à la moitié du texte. Dans l'ensemble, 15% des élèves (109 sur 861) semblent avoir été influencés par une image.



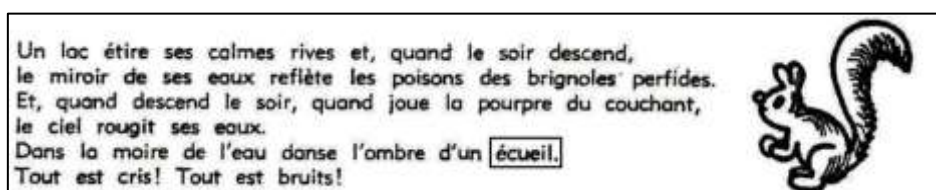
Cette proportion reste à peu près stable entre le CE1 et le CM2. En effet, en CE1, ils sont 12% à avoir lu *hirondelle*. Les élèves de CE2 sont légèrement plus nombreux : 18%. Entre le CM1 et le CM2, ce taux varie entre 14% et 16%.



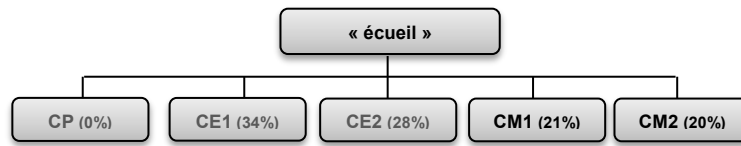
On retrouve ici, comme pour le mot « haies », le *h* muet en début de mot et en fin de mot, un phonème représenté par trois lettres (-eau). Néanmoins, il ne semble pas ici, que ce soit le *h* qui ait posé des difficultés de lecture. Nous ne pensons pas non plus que ce soit le phonème [o]. En réalité, nous supposons, que les élèves, lorsqu'ils ont lu le début du mot (*hironde-*), aient prononcé le mot *hirondelle*, mot qui leur est plus familier, sans même lire la fin du mot. L'identification du début et du milieu du mot a activé la représentation lexicale du mot *hirondelle*.

On remarque que, ici aussi, le mot prononcé est en cohérence avec le contexte syntaxique et sémantique.

Le mot « *écueil* », quant à lui, apparaît vers la fin du texte. De ce fait, les élèves qui ont lu jusque-là sont peu nombreux et sont essentiellement scolarisés en CM1 et CM2. En effet, même si cette erreur apparaît dès les premières années de l'apprentissage de la lecture (dès le CE1), nous n'en avons pas tenu compte puisque ce mot est le 206e du texte et que les élèves de CP ont lu en moyenne 55 mots, ceux de CE1 154, et 197 pour les CE2. Les élèves de CM1 ayant lu en moyenne 234 mots et les élèves de CM2 ayant lu en moyenne 253 mots, nous avons donc effectué notre analyse pour seulement ces deux niveaux scolaires



Ainsi, en CM1, 21% des élèves qui ont lu ce mot, ont prononcé *écureuil*. En CM2, leur nombre ne varie guère puisqu'ils sont 20%.



La difficulté de ce mot vient sans doute de la fin du mot avec les sons [œ] et [j]. Comme pour le mot « hirondeau », il est possible que l'identification du début et de la fin du mot ait provoqué l'activation de la représentation écreuil.

Là aussi, on remarque que le mot produit est proche graphiquement et phonétiquement du mot du texte et qu'il a un sens dans la phrase.

En **conclusion**, cette analyse des différents types d'erreur nous permet de constater que les dessins ne semblent pas avoir influencé de la même façon les lecteurs débutants et les lecteurs plus avancés. En effet, les élèves ont chaque fois lexicalisé un mot en tenant compte des indices graphiques. Toutefois, il semblerait que les élèves plus avancés en lecture, qui ont eu recours aux dessins, se soient basés sur davantage d'indices graphiques que les élèves de CP. En effet, en ce qui concerne les mots « cordeaux » et « pompe », on remarque que les noms des dessins que les élèves ont prononcés, sont tous deux des voisins orthographiques des mots du texte : ce sont les mêmes mots à une lettre près (*pompe/pomme* et *cordeaux/corbeaux*). Concernant le mot « hirondeau », on constate qu'il n'y a que la fin du mot qui diffère : *hirondelle*. Enfin, quant à « l'ennui » lu principalement *les nuits*, il se pourrait que ce soit l'article élidé « l' » qui soit en partie la cause de ces erreurs. L'élision aurait ainsi entraîné une resyllabation du mot et les élèves auraient cherché à produire un mot qu'ils connaissent. Ainsi, comme au CP, chaque verbalisation s'est appuyée sur des indices graphiques saillants (début et milieu du mot pour « hirondeau » ou début et fin pour les mots « l'ennui », « pompe » et « cordeaux ») et le mot produit semble être un mot qui leur est plus familier. En revanche, alors que les élèves de CP produisent des mots qui ne respectent que la syntaxe, on remarque que les élèves plus âgés ont verbalisé des mots qui sont aussi en rapport avec le contexte sémantique proche, c'est-à-dire le syntagme ou la phrase dans lequel ils s'inscrivent. Il semble alors apparaître ici une nouvelle stratégie. En effet, au-delà d'une stratégie compensatoire lors de l'identification des mots pour les lecteurs débutants, il semble

que pour tous les élèves, l'utilisation des dessins soit aussi une aide à la construction de sens, que ce soit au niveau lexical ou syntagmatique voir phrastique.

En conséquence, le fait que ce type d'erreur soit présent dès les premières années de l'apprentissage de la lecture laisse penser que les jeunes lecteurs ne font pas que décoder, ils essaient aussi de produire du sens.

CONCLUSION

Ce chapitre avait pour objectif de voir dans quelle mesure les dessins présents autour du texte de *L'alouette* influencent la lecture des élèves et, de façon plus générale, quelles sont les informations qu'ils nous apportent en ce qui concerne l'apprentissage de la lecture. Ce test repose sur une conception classique de l'acquisition de la lecture, ce qui suppose que lors de cet apprentissage, les élèves passent d'abord par une étape de décodage avant d'accéder au sens. Nous avons donc supposé que, face à des mots difficiles à lire, les apprentis lecteurs auraient davantage recours aux dessins car ils sont plus accessibles que l'écrit. Ils les utiliseraient donc comme une source d'indice pertinent. De cette façon, les dessins devaient donc influencer de moins en moins les élèves au fur et à mesure qu'ils progressent dans l'apprentissage de la lecture.

Lors d'une première analyse, nous avons constaté que l'utilisation des dessins n'est pas la même selon le niveau scolaire. En effet, on observe bien une discontinuité entre les élèves de CP et les autres. Toutefois, cette discontinuité n'est pas celle à laquelle on s'attendait : les élèves qui ont eu le plus recours aux dessins ne sont pas les lecteurs débutants mais les lecteurs plus avancés. De plus, il apparaît que le nombre d'élèves ayant eu ce type de comportement augmente d'année en année. Autrement dit, plus un élève sait lire et plus il aurait tendance à avoir recours à ce type de réponse. L'utilisation des dessins ne serait donc pas seulement une caractéristique des apprentis lecteurs.

Toutefois, ces résultats sont à prendre avec précautions. En effet, les élèves n'ayant pas tous lu le même nombre de mots, nous nous sommes rendue compte que ne pouvions pas réunir dans une même analyse tous les élèves. A partir de ce constat, et pour plus de rigueur, nous avons repris l'examen des données en analysant d'abord la partie du texte que tous les élèves ont lue, puis le reste du texte lu, partie commune aux lecteurs plus avancés. Nous avons remarqué alors que, pour le début du texte, les élèves de CP se différencient là aussi des autres mais en sens inverse : bien qu'à ce niveau scolaire, les élèves à avoir commis une erreur en

lien avec les dessins soient minoritaires (moins d'un quart), ils restent toutefois les plus nombreux à avoir donné ce type de réponse, en comparaison avec les élèves des autres niveaux scolaires. En revanche, à un niveau plus « expert », on remarque que les élèves à avoir utilisé les dessins sont beaucoup plus nombreux : presque la totalité d'entre eux a commis au moins une fois ce type d'erreur. Les résultats de cette seconde analyse semblent donc bien confirmer le fait qu'il existe une discontinuité dans l'utilisation des dessins qui ne serait pas une caractéristique des apprentis lecteurs mais plutôt une spécificité des lecteurs plus avancés.

Il semble donc que le recours aux dessins lors de la lecture du texte de *L'alouette* ne soit pas uniquement une stratégie compensatoire des lecteurs débutants puisque les lecteurs les plus « experts » sont ceux qui s'avèrent être les plus touchés par ce phénomène. Lire ne serait donc pas seulement une identification des mots.

Partant de ce constat, nous avons donc réalisé une troisième analyse avec une approche plus qualitative. A partir des types d'erreurs en lien avec les dessins, nous avons essayé de déterminer pour chaque niveau scolaire des stratégies quant à l'utilisation des images durant la lecture du texte de *L'alouette*. On remarque que là aussi, les élèves de CP se démarquent des autres. Il semblerait qu'à ce niveau, la lecture soit lexicale, c'est-à-dire que les élèves cherchent à identifier un mot sans se préoccuper du sens qu'il peut avoir dans la phrase. En revanche, à partir du CE1, les élèves produisent un mot mais cherchent aussi à ce que ce dernier soit sémantiquement acceptable dans la phrase.

D'autre part, l'erreur qui n'est dominante qu'en CP est un mot qui ne présente pas une forte ressemblance grapho-phonémique avec le nom induit par le dessin. La verbalisation s'appuie sur quelques indices en début du mot : le *ha-* du mot « haies » lu *chat*. On remarque que le mot produit appartient à la même catégorie grammaticale, celle des noms communs, sans apparaître vraiment sémantiquement pertinent : « dans les haies vives » → *dans les chats vives*. Ce constat montre qu'au début de l'apprentissage de la lecture, les élèves cherchent à produire du sens, ils ne font pas que décoder, mais cette recherche de sens ne concerne que le mot, pas la phrase.

C'est à partir du CE1 que les mots produits sont en cohérence avec le contexte syntagmatique ou phrastique : les mots produits ont un sens dans le syntagme ou la phrase :

Ex *alouette fait un œuf [avec un rien de paille]* → sens au niveau du syntagme

Ex. *l'hirondelle piaille sous la pente des bardeaux [...]* → sens au niveau de la phrase.

A ce niveau, on observe aussi que le nombre d'indices pris en compte augmente : les mots produits présentent davantage de ressemblances grapho-phonémiques avec les noms correspondants aux dessins (*hirondeau/hirondelle*, *écueil/écureuil*). D'ailleurs, pour certains, il s'agit même de voisins orthographiques (*pompe/pomme*, *cordeaux/corbeaux*) : les élèves se sont donc davantage appuyés sur les mots écrits (début et fin du mot).

Par conséquent, il apparaît que les apprentis lecteurs ne feraient pas que du décodage, ils s'intéresseraient aussi au sens mais à un sens local, lexical. Ainsi, plus un élève avance dans l'apprentissage de la lecture et plus sa recherche de sens s'élargit : du sens lexical au sens syntagmatique voire phrastique⁴.

Les verbalisations en cycle 2, comme nous l'avons remarqué, sont effectuées à partir de quelques indices graphiques, souvent un seul graphème, contrairement à celles des autres élèves qui en prennent en compte davantage. On peut alors se demander si, pour les lecteurs débutants, les dessins ne seraient pas une aide, alors que pour les lecteurs plus avancés elles seraient plutôt perturbatrices. Ce constat rejoint celui de Stanovich (1981) qui, suite à une expérimentation concernant les effets du contexte en lecture chez des enfants du primaire, considère que le recours au contexte serait une stratégie compensatoire des mauvais lecteurs. Cependant, rappelons que Samuels (1974) a mené une expérience afin de voir quelle est l'influence de l'image dans un apprentissage de mots chez les enfants non-lisants et dans le décodage de mots chez les lecteurs débutants. Pour cela, il a demandé à des enfants âgés de six à sept ans débutants en lecture de lire un texte en le leur présentant soit tel quel, soit illustré. Les résultats montrent que l'image ne change pas la performance de ceux qui lisent bien et que ceux qui lisent mal décodent plus de mots en l'absence de dessins. Les dessins détourneraient donc l'attention du lecteur et affecteraient ainsi davantage les apprentis-lecteurs.

Par ailleurs, cette étude nous amène à nous poser la question suivante. Les dessins présents autour du texte ont-ils incité les élèves à commettre des erreurs ou ont-ils confirmé leur lecture ?

En considérant que l'élève suppose un lien entre le texte et les dessins, nous pouvons en effet émettre deux hypothèses en ce qui concerne les stratégies de lecture : soit l'élève a d'abord vu le dessin, puis s'en est resservi pour décoder le mot, soit il a tenté un premier décodage qu'il a confirmé par le dessin. Ces deux hypothèses de processus de lecture se déclinent en trois temps. Dans le premier cas, l'élève voit d'abord le dessin (et s'en

⁴ Nous ne parlerons pas ici du sens textuel car dans le texte de *L'alouette*, seuls des syntagmes ou des phrases sont porteurs de sens. Le texte dans sa globalité ne peut pas mesurer la compréhension.

imprègne), puis tente un décodage et, à partir de ce qu'il a déjà décodé, il en déduit un mot par rapport au dessin. Dans le second cas, l'élève effectue d'abord un décodage partiel du mot, puis tente de lui donner du sens à l'aide de son lexique interne et s'appuie sur le dessin pour confirmer sa lecture. Même s'il nous est impossible de déterminer le rôle qu'ont joué les dessins dans la lecture de certains mots, nous constatons toutefois que, hormis pour le mot « haies » lu *chat*, les élèves ont essentiellement produit des mots plus ou moins en cohérence avec les phrases. Comme nous l'avons supposé, les élèves semblent donc essayer de donner du sens à leur lecture. Ils ont tout d'abord effectué un décodage partiel du mot, puis ils ont anticipé la fin (en contrôlant ou non la graphie) du mot à l'aide de leur lexique interne afin de lire un mot qui leur est certainement plus familier.

Le rôle des dessins présents autour du texte de *L'alouette* est par conséquent difficile à cerner. Pour en savoir plus, il faudrait faire lire le texte avec les dessins et sans les dessins à deux groupes équivalents. On ne peut pas expliquer ces erreurs seulement par la présence des dessins, car d'autres facteurs (ressemblance graphophonémique, décodage partiel, lexicalisation, contextualisation...) entrent aussi en jeu.

Toutefois, cette analyse nous a permis de constater que l'utilisation des dessins s'effectue bien de manière discontinue. Discontinue, certes, mais pas de la façon dont on pensait : ce sont les lecteurs plus avancés qui utilisent le plus les dessins. Or, au départ, avec ce test, Lefavrais avait pour intention de dépister toute déficience en lecture. Les dessins, apparaissent ici comme des *artefacts* et celui qui tombe dans les pièges est celui qui présente des difficultés de lecture. Il est donc probable que l'utilisation des dessins soit une stratégie compensatoire alphabétique des « mauvais » lecteurs, mais pas que cela. L'analyse qualitative nous permet de constater que l'utilisation des dessins est aussi une preuve de construction de sens, lexical pour les apprentis lecteurs et syntagmatique voire phrastique pour les plus avancés.

L'analyse des autres types d'erreurs lors de la lecture du texte de *L'alouette* va nous permettre de poursuivre la réflexion quant au déroulement de l'acquisition de la lecture.

Chapitre 12

L'INFLUENCE DES SIMILITUDES LORS DE LA LECTURE DU TEXTE DE *L'ALOUETTE*

Plan du chapitre

INTRODUCTION	260
PREDICTIONS	260
METHODOLOGIE D'ANALYSE	261
I. L'influence des similitudes lors de la lecture du texte de <i>L'alouette</i>	264
<i>Quels sont les élèves les plus influencés par les similitudes?</i>	<i>264</i>
<i>Dans quelle proportion les élèves ont-ils été influencés?</i>	<i>264</i>
1. L'influence des similitudes en cycle 2	2645
<i>Quels sont les élèves les plus influencés en cycle 2?</i>	<i>265</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>266</i>
2. L'influence des similitudes en cycle 3	266
<i>Quels sont les élèves les plus influencés en cycle 3?</i>	<i>267</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>267</i>
II. L'effet des similitudes dans les différentes parties du texte	270
1. L'influence des similitudes dans le début du texte	26670
<i>Quels sont les élèves les plus influencés dans le début du texte?</i>	<i>270</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>270</i>
2. L'influence des similitudes dans le reste du texte	26671
<i>Quels sont les élèves les plus influencés dans le reste du texte?</i>	<i>271</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>272</i>
III. Quels sont les types d'erreurs les plus induits par les similitudes?	274
<i>Les types d'erreurs des lecteurs débutants</i>	<i>274</i>
<i>Les types d'erreur des lecteurs plus avancés</i>	<i>276</i>
CONCLUSION	280

INTRODUCTION

L'effet des similitudes lors de l'apprentissage de la lecture a fait l'objet de plusieurs recherches notamment celles portant sur les analogies. De nos jours, « *presque tout le monde s'accorde pour reconnaître que les analogies sont utilisées dans l'apprentissage de la lecture et de l'écriture* » (Gombert, Bryant et Warrick, 1997, p.331)¹. Cependant, un désaccord apparaît lorsqu'il s'agit de voir à quel moment ce processus a lieu et quelles sont les unités traitées par analogie.

En effet, pour certains chercheurs, le recours aux analogies s'effectuerait au début de l'apprentissage de la lecture (Goswami 1986, 1988, 1990 ; Faure, Ecalle et Magnan, 2000) : les enfants traiteraient les mots nouveaux sur la base de ceux qu'ils connaissent déjà. Il s'agirait alors « *d'un transfert de connaissances basé sur la découverte de ressemblances entre les unités sonores que l'enfant maîtrise et des séquences de lettres connues dans leurs formes phonologique et graphique* ». Ces différentes recherches font apparaître que les enfants auraient essentiellement recours à un traitement analogique pour les unités infra-lexicales, notamment les rimes des mots.

Pour d'autres chercheurs, ce processus serait plutôt une caractéristique des lecteurs plus avancés (Bowey, Vaughan et Hansen, 1998). D'ailleurs, la plupart des modèles développementaux soutiennent que le recours aux analogies interviendrait plus tard car cette procédure nécessiterait une maîtrise des correspondances graphèmes-phonèmes. Ainsi, ce processus ne se mettrait en place qu'à partir du 3^e stade, le stade orthographique.

PREDICTIONS

Nous nous demandons alors **dans quelle mesure les erreurs des élèves dues aux similitudes lors de la lecture du texte de *L'alouette* font apparaître le caractère discontinu ou continu de l'apprentissage de la lecture.**

Nous supposons (contrairement aux modèles développementaux) que les élèves auront recours à un traitement par analogie pour pallier toute défaillance dans le traitement alphabétique. De cette façon, la maîtrise du code devrait faire diminuer ce type d'erreur.

¹ Dans « *L'apprentissage de la lecture, fonctionnement et développements cognitifs* » (Ecalle et Magnan, 2007)

Ainsi, d'un point de vue quantitatif, on s'attend à ce que ce type d'erreur diminue dans le temps. D'un point de vue qualitatif, on s'attend à ce que la nature des erreurs rende compte d'une évolution dans le rapport code/sens : des erreurs de type lexical chez les plus jeunes et des erreurs de type grapho-sémantique chez les plus âgés.

METHODOLOGIE D'ANALYSE

Certains mots ou syntagmes du texte se ressemblent graphiquement et/ou phonémiquement et peuvent de ce fait être lus de la même façon par l'élève. Or ce qui semble identique ne l'est pas nécessairement. Nous supposons qu'un mot *B* qui présente des similitudes (grapho-phonémiques) avec un mot *A* placé avant lui est lu comme le mot *A*. Dans ce cas, nous dirons que la lecture du mot *A* a interféré sur celle du mot *B*. Prenons pour exemple le syntagme « *sous la mousse ou sur le toit* » : si l'élève a lu correctement « *sous la mousse* » et s'il lit « ***sous la* toit / *sous le* toit » alors nous supposons qu'il y a eu une interférence.**

Nous avons relevé plusieurs cas de similitudes sur lesquels au moins un élève s'est trompé (*cf.* tableau 31 + cartographie 6 des similitudes dans le texte de *L'alouette*, p.263):

Tableau 31 : erreurs susceptibles d'être dues à la présence de similitudes

Mots/Syntagmes du texte	Types d'erreurs
« Sous la mousse ou sur le toit »	<i>Sous la mousse ou sous le toit</i> <i>Sous la mousse ou sur la toit</i> <i>Sous la mousse ou sous la toit</i>
« Dans les haies vives ou le chêne »	<i>Dans les haies vives ou les chêne</i> <i>Dans les haies vives ou sous le chêne</i> <i>Dans les haies vives sous le chêne</i> <i>Dans les haies vives ou sur le chêne</i> <i>Dans les haies vives ou dans le chêne</i>
« Le printemps a mis ses nids »	<i>Le printemps a mis ses mis</i>
« Le printemps a nids au bois »	<i>Le printemps a mis au bois</i> <i>Le printemps a mis ses nids au bois</i> <i>Le printemps a mis nids au bois</i> <i>Le printemps mis ses nids</i> <i>Le printemps mis nids au bois</i> <i>Le printemps a mis des nids au bois</i>
« Annie, amie du renouveau »	<i>Annie, annie du renouveau</i>
« C'est le doux temps »	<i>C'est le doux printemps</i>
« Amie Annie au bois joli »	<i>Annie Annie au bois joli</i> <i>Amie Amie au bois joli</i> <i>Annie amie au bois joli</i>
« Amie Annie au bois joli »	<i>Amie Annie du bois joli</i>
« Dans les buis »	<i>Dans les bois</i>
« Annie Annie au doigt joli »	<i>Annie Amie au doigt joli</i> <i>Amie Annie au doigt joli</i> <i>Amie, amie au doigt joli</i>
« au doigt joli »	<i>Du doigt joli</i>
« au doigt joli »	<i>au bois joli</i>
« au bout du temps »	<i>Au bois du temps</i>
« L'alouette fait [...] ; alouette fait un nœud »	<i>L'alouette fait [...] ; l'alouette fait un nœud</i> <i>L'alouette fait [...] ; alouette fait ses nœuds</i>
« avec un rien de paille »	<i>Alouette fait un rien de paille</i>
« L'hirondeau piaille »	<i>L'alouette piaille</i>
« L'hirondeau piaille sous la pente »	<i>L'hirondeau paille sous la pente</i>
« et, vif et gai, le geai »	<i>Et, vif et gai, et geai</i>
« Et vif et gai, le geai sur l'écaïlle... »	<i>Et vif et gai, le gai sur l'écaïlle</i>
« Et vif et gai, le geai sur l'écaïlle... »	<i>Et vif et gai, le geai sous l'écaïlle</i>
« l'écaïlle argentée du bouleau »	<i>L'écaïlle argentée sur bouleau</i> <i>L'écaïlle argentée sur du bouleau</i>
« Dans les pampres d'or »	<i>Dans les pompes d'or</i>
« On surprend des rixes »	<i>On suspend des rixes</i>
« Et lourd et lent l' envol »	<i>Et lourd et lent vol</i>
« L'envol des corbeaux »	<i>L'envol des cordeaux</i>
« Dans la moire de l'eau »	<i>Dans le miroir de l'eau danse l'ombre</i>
« danse l'ombre... »	<i>dans l'ombre</i>
« Au clair de lune »	<i>Au clair de la lune mon amie annie</i>
« Au clair de la lune »	<i>Au clair de lune</i>
« prête-moi la plume »	<i>Prête-moi la lune</i>

Cartographie 6 : mots sur lesquels les élèves ont commis une erreur du fait des similitudes dans le texte de L'alouette

L'alouette.

Sous la mousse ou sur le toit,
dans les haies vives ou le chêne fourchu,
le printemps a mis ses nids.
Le printemps a nids au bois.

Annie amie du renouveau, c'est le doux temps.
Amie Annie au bois joli gamine le pinson.
Dans les buis gîte une biche, au bois chantant.
Annie, Annie! au doigt joli, une églantine laisse du sang :
au bout du temps des féeries viendra l'ennui.

L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau piaille sous la pente des bardeaux et, vif et gai, le geai sur l'écaille argentée du bouleau, promène un brin d'osier.
Au verger, dans le soleil matinal, goutte une pompe dégelée.
On voit un bec luisant qui trille éperdument des notes claires et, dans les pampres d'or que suspend la grille antique, on surprend des rixes de moineaux.
Au potager s'alignent les cordeaux; l'if est triste à l'horizon et lourd et lent l'envol des corbeaux.

Un lac étire ses calmes rives et, quand le soir descend, le miroir de ses eaux reflète les poisons des brignoles perfides. Et, quand descend le soir, quand joue la pourpre du couchant, le ciel rougit ses eaux.
Dans la moire de l'eau danse l'ombre d'un écueil.
Tout est cris! Tout est bruits!

Une amarre est décochée... une barque est arrimée... des matelots jettent leurs cassettes sur le rivage...
Tout est cris! Tout est bruits!
Au clair de la lune mon ami Pierrot...
Au clair de lune mon amie annie...
Au clair de la lune mon ami Pierrot, prête-moi la plume pour écrire un mot.

o u e i a
le la les un dans des do ti pu mi

Plusieurs influences d'erreurs possibles

I. L'influence des similitudes lors de la lecture du texte de *L'alouette*

Quels sont les élèves les plus influencés par les similitudes?

De façon générale, il semble que les élèves de notre échantillon sont une majorité (62%) à avoir commis au moins une erreur d'interférence due aux similitudes (cf. tableau 32)

Tableau 32 : pourcentage et nombre d'élèves ayant commis ou non des erreurs dues à une similitude selon le cycle de scolarisation

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
Cycle 2	52% (169)	48% (154)	100% (323)	43.405	.000
Cycle 3	30% (159)	70% (379)	100% (538)		
Ensemble	38% (328)	62% (533)	100% (861)		

Lecture : 52% des élèves scolarisés en cycle 2 (soit 169 sur 323) n'ont commis aucune erreur due à une similitude et 48% en ont commis au moins une (soit 154 sur 323).

Dans le détail, il semble que ce type d'erreur augmente avec le niveau de scolarisation (cf. tableau 32). En effet, alors qu'un peu plus de la moitié (52%) des élèves scolarisés en cycle 2 n'a pas commis d'erreur de ce type, leur nombre diminue en cycle 3, puisqu'ils sont 30%. Autrement dit, on constate que la majorité des élèves en cycle 2 n'a pas commis d'erreur de ce type alors qu'à l'inverse, la majorité des élèves en cycle 3 en ont commis au moins une.

On observe donc ici une rupture dans la progression des élèves, puisqu'il apparaît une augmentation de 22 points de la proportion d'élèves qui ont commis ce type d'erreur entre le cycle 2 et le cycle 3.

Dans quelle proportion les élèves ont-ils été influencés?

Sur les 861 élèves, 533 ont commis au moins une erreur due aux similitudes. On constate que parmi eux, un peu moins de la moitié (246, soit 46%) a commis plusieurs erreurs de ce type alors qu'un peu plus de la moitié n'en a commis qu'une seule (287, soit 54%) (cf. tableau 33).

Tableau 33 : pourcentage et nombre d'élèves ayant commis une ou plusieurs erreur(s) due(s) à une similitude selon le cycle de scolarisation

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
Cycle 2	64% (99)	36% (55)	100% (154)	8.916	.003
Cycle 3	50% (188)	50% (191)	100% (379)		
Ensemble	54% (287)	46% (246)	100% (533)		

Lecture : 64% des élèves scolarisés en cycle 2 qui ont commis une erreur due aux similitudes (soit 99 sur 154) n'ont commis qu'une seule erreur due à une similitude et 36% en ont commis plusieurs (soit 55 sur 154).

Plus précisément, alors que les élèves scolarisés en cycle 2 ont tendance à n'avoir été influencés qu'une seule fois (64%), ce n'est pas le cas pour ceux du cycle 3 qui sont autant à avoir été influencés une seule fois ou plusieurs fois (50%).

Ainsi, la rupture constatée précédemment semble se confirmer ici, puisque non seulement les élèves en cycle 3 sont les plus nombreux à avoir commis ce type d'erreur mais ils en commettent aussi davantage. Les élèves plus âgés, et par conséquent plus avancés en lecture, seraient donc ceux qui semblent être le plus touchés par cet effet de similitude.

Voyons maintenant dans le détail ce qu'il en est des élèves scolarisés dans les classes en cycle 2 et cycle 3 : ces tendances se retrouvent-elles au sein de chaque classe ?

1. L'influence des similitudes en cycle 2

Parmi les 861 élèves, 323 sont scolarisés en cycle 2 : 143 sont en CP et 180 en CE1.

Quels sont les élèves les plus influencés en cycle 2 ?

Rappelons que, dans l'ensemble, la majorité des élèves en cycle 2 (52%) n'a pas commis d'erreur d'interférence due à une similitude (cf. tableau 34).

Tableau 34 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude selon la classe en cycle 2

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CP	64% (91)	36% (52)	100% (143)	12.367	.000
CE1	43% (78)	57% (102)	100% (180)		
Ensemble	52% (169)	48% (154)	100% (323)		

Lecture : 64% des élèves scolarisés en CP (soit 91 sur 143) n'ont commis aucune erreur due à une similitude et 36% en ont commis au moins une (soit 52 sur 143).

Si on regarde plus précisément les résultats obtenus à chaque niveau scolaire, on remarque que la tendance générale ne se retrouve pas au sein des classes qui composent le

cycle 2. En effet, alors que les élèves de CP sont une majorité à ne pas avoir commis d'erreur de ce type (64%), c'est l'inverse pour les élèves de CE1 qui sont 57% à avoir commis au moins une erreur de ce type.

La rupture observée précédemment entre le cycle 2 et le cycle 3 semble se situer plus précisément entre les élèves de CP et de CE1.

Dans quelle proportion ont-ils été influencés?

On remarque que, parmi les élèves qui ont commis au moins une erreur de ce type, 64% d'entre eux n'ont commis qu'une seule erreur (*cf.* tableau 35).

Tableau 35 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude selon la classe en cycle 2

	Une seule erreur	Plusieurs erreurs	Total	χ^2	<i>p.</i>
CP	69% (36)	31% (16)	100% (52)	.543	.461
CE1	62% (63)	38% (39)	100% (102)		
Ensemble	64% (99)	36% (55)	100% (154)		

Lecture : 69% des élèves scolarisés en CP qui ont commis une erreur due aux similitudes n'ont commis qu'une seule erreur (soit 36 sur 52) et 31% en ont commis plusieurs (soit 16 sur 52).

Cette tendance se retrouve aussi bien en CP qu'en CE1 : quelle que soit la classe fréquentée en cycle 2, les élèves ne commettent essentiellement qu'une seule erreur de ce type. Ce constat nous permet de voir qu'il existe une certaine continuité en ce qui concerne la quantité d'erreurs de ce type qui ont été commises.

Par conséquent, bien qu'en cycle 2 les élèves en CE1 soient plus nombreux que les élèves en CP à avoir commis au moins une erreur d'interférence due aux similitudes présentes dans le texte de *L'alouette*, ils n'en commettent cependant pas davantage.

2. L'influence des similitudes en cycle 3

Sur les 861 élèves de notre échantillon, 538 sont scolarisés en cycle 3. Parmi eux, 174 sont en CE2, 188 en CM1 et 176 en CM2.

Quels sont les élèves les plus influencés en cycle 3 ?

Rappelons que 70% des élèves en cycle 3 ont commis au moins une erreur due aux similitudes.

Tableau 36 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude selon la classe en cycle 3

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CE2	29% (50)	71% (124)	100% (174)	3.720	.156
CM1	25% (48)	75% (140)	100% (188)		
CM2	35% (61)	65% (115)	100% (176)		
Ensemble	30% (159)	70% (379)	100% (538)		

Lecture : 29% des élèves scolarisés en CE2 (soit 50 sur 174) n'ont commis aucune erreur due à une similitude et 71% en ont commis au moins une (soit 124 sur 174).

Cette tendance se retrouve au sein de chaque classe qui compose le cycle 3. En effet, 71% des élèves en CE2 ont commis au moins une erreur de ce type. En CM1, ils sont 75% et 65% en CM2. On remarque donc que, quelle que soit la classe fréquentée en cycle 3, les similitudes ont eu un effet lors de la lecture du texte.

Ainsi, il semble qu'il y ait une certaine continuité dans la progression des élèves de ce cycle. Progression, certes, mais négative puisque ce n'est pas la performance qui s'améliore mais la proportion d'élèves qui ont commis ce type d'erreur qui augmente.

Dans quelle proportion ont-ils été influencés ?

On remarque que, parmi les élèves qui ont commis au moins une erreur de ce type, on remarque qu'il y a autant d'élèves qui ont commis une erreur ou plusieurs.

Tableau 37 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude selon la classe en cycle 3

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CE2	55% (68)	45% (56)	100% (124)	4.391	.111
CM1	51% (72)	49% (68)	100% (140)		
CM2	42% (48)	58% (67)	100% (115)		
Ensemble	50% (188)	50% (191)	100% (379)		

Lecture : 55% des élèves scolarisés en CE2 qui ont commis une erreur due aux similitudes n'ont commis qu'une seule erreur (soit 68 sur 124) et 45% en ont commis plusieurs (soit 56 sur 124).

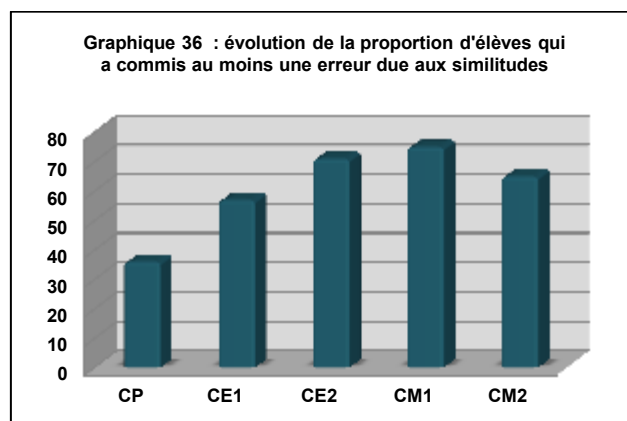
Si on regarde les résultats par classe, on s'aperçoit que la tendance est la même pour les élèves de CE2 et CM1. En effet, on n'observe pas de différence significative entre les élèves

qui ont commis une seule erreur et ceux qui en ont commis plusieurs. En revanche en CM2, il apparaît que les élèves ont tendance à en commettre plusieurs (58%).

Par conséquent, les élèves en cycle 3, contrairement à ceux scolarisés en cycle 2, sont autant à avoir commis plusieurs erreurs qu'une seule erreur. De plus, on remarque que la proportion d'élèves qui ont commis plusieurs erreurs augmente d'année en année.

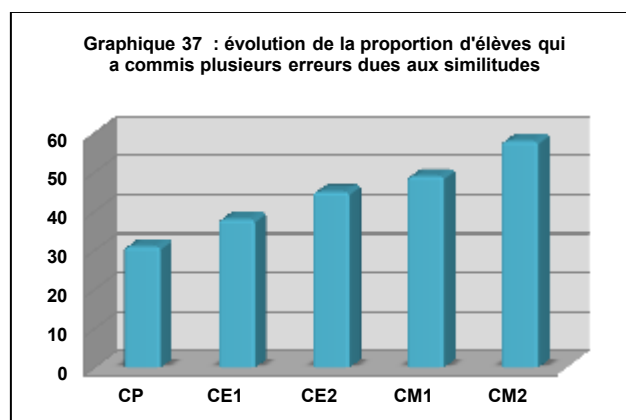
En **conclusion**, cette première analyse nous permet de constater que les similitudes présentes dans le texte de *L'alouette* ont influencé les élèves à tous les niveaux scolaires mais à des degrés différents. En effet, on remarque une discontinuité quantitative en ce qui concerne la proportion d'élèves qui se sont trompés : alors qu'ils sont une minorité en cycle 2, ils deviennent une majorité en cycle 3. Ce seraient donc les élèves scolarisés dans le dernier cycle qui seraient le plus touchés.

Si on regarde les résultats de classe en classe, on remarque que la rupture dans la progression s'observe plus particulièrement entre le CP et le CE1 : ainsi, alors qu'une majorité des élèves de CP ne semble pas avoir commis d'erreurs dues à une similitude, on constate qu'à partir du CE1, et jusqu'en CM2, la tendance s'inverse : les élèves sont une majorité à avoir été influencé au moins une fois par des similitudes. (cf. graphique 36).



Non seulement les élèves à avoir commis ce type d'erreur deviennent majoritaires à partir du CE1, mais ils sont aussi de plus en plus nombreux jusqu'en CM1. En CM2, bien que leur nombre diminue légèrement, cette catégorie d'élèves est encore importante.

Nous avons aussi remarqué que, le nombre d'erreurs commises augmente avec le niveau scolaire. En effet, alors que jusqu'en CE1, la plupart des élèves a tendance à ne faire qu'une seule erreur de ce type, à partir du CE2 ils tendent à en faire plusieurs (cf. graphique 37).



La rupture observée précédemment se retrouve donc ici en ce qui concerne les nombre d'erreurs commises.

Ainsi, cette première analyse nous permet de constater qu'il existe une discontinuité quantitative qui nous laisse penser à une rupture dans l'apprentissage de la lecture. Nous avons supposé que les lecteurs débutants seraient davantage influencés par les similitudes que les lecteurs plus avancés. En effet, nous avons émis l'hypothèse que le recours à un traitement analogique serait le résultat d'une défaillance dans la maîtrise du code entraînant une lecture identique de deux mots similaires. Or, il apparaît ici que ce n'est pas le cas. Il est alors possible que le processus de traitement des mots par analogie intervienne bien lors d'une étape ultérieure dans l'apprentissage de la lecture comme le supposent les modèles développementaux (Frith, 1985).

Toutefois, rappelons que les élèves n'ont pas tous lu la même quantité de mots. Les comparaisons précédentes semblent par conséquent être biaisées. De ce fait, au risque de perdre de l'information, nous nous proposons maintenant de reprendre cette étude mais en comparant maintenant ce qui est comparable, c'est-à-dire la partie du texte que tous les élèves ont lue.

II. L'effet des similitudes dans les différentes parties du texte

1. L'effet des similitudes dans le début du texte

Quels sont les élèves les plus influencés dans le début du texte?

Il apparaît que lorsqu'on ne tient compte que de la partie du texte qui a été lue par tous les élèves, le constat n'est plus le même. En effet, alors que dans l'analyse précédente les élèves en cycle 2 ne commettaient essentiellement pas d'erreurs et ceux en cycle 3 en commettaient au moins une, il s'avère qu'ici le fait de commettre des erreurs dues aux similitudes ne dépend plus du niveau de scolarisation (*cf.* tableau 38)

Tableau 38 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude dans le début du texte selon la classe

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CP	71% (101)	29% (42)	100% (143)	1.942	.746
CE1	72% (130)	28% (50)	100% (180)		
CE2	72% (126)	28% (48)	100% (174)		
CM1	75% (142)	25% (46)	100% (188)		
CM2	69% (122)	31% (54)	100% (176)		
Ensemble	72% (621)	28% (240)	100% (861)		

Lecture : 71% des élèves scolarisés en CP (soit 101 sur 143) n'ont commis aucune erreur due à une similitude dans le début du texte et 29% en ont commis au moins une (soit 42 sur 143).

En effet, dans cette partie du texte, les élèves qui n'ont pas commis d'erreur sont les plus nombreux que ce soit en CP (71%), CE1 (72%), CE2 (72%), CM1 (75%) ou CM2 (69%). On observe donc une régularité dans la progression des élèves, ce qui laisse penser que, à toute chose comparable, l'apprentissage de la lecture se déroulerait de façon continue.

Ainsi, on constate que lorsqu'on ne tient compte que de la partie lue par tous les élèves, les similitudes présentes dans le texte de *L'alouette* ne semblent plus influencer les élèves, quel que soit leur niveau de scolarisation.

Dans quelle proportion ont-ils été influencés ?

Ici, aussi les résultats changent. En effet, dans la première analyse, nous avons constaté que les élèves de CP, CE1, CE2 et CM1 commettaient essentiellement une erreur et les élèves en CM2 en commettaient plusieurs. On remarque que dans le début du texte, les élèves ont tendance à ne commettre qu'une seule erreur.

Tableau 39 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude dans le début du texte selon la classe

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CP	81% (34)	19% (8)	100% (42)	5.625	.229
CE1	82% (41)	18% (9)	100% (50)		
CE2	67% (32)	33% (16)	100% (48)		
CM1	85% (39)	15% (7)	100% (46)		
CM2	76% (41)	24% (13)	100% (54)		
Ensemble	78% (187)	22% (53)	100% (240)		

Lecture : 81% des élèves scolarisés en CP qui ont commis une erreur due aux similitudes n'ont commis qu'une seule erreur (soit 34 sur 42) et 19% en ont commis plusieurs (soit 8 sur 42).

En effet, quelle que soit la classe de scolarisation, les élèves à avoir commis une seule erreur sont à chaque niveau, les plus nombreux. Le phénomène de continuité observé précédemment semble donc se confirmer par ces résultats.

Ainsi, il semble que lorsqu'on analyse seulement la partie lue par tous les élèves, les similitudes n'ont pas majoritairement influencé les élèves. L'apprentissage de la lecture se déroulerait alors de façon continue.

Toutefois, cette analyse reste incomplète puisque nous ne disposons pas de toute l'information (seulement le début du texte). C'est pourquoi, nous avons poursuivi cette étude en ne tenant compte, cette fois-ci, que de l'autre partie du texte, c'est-à-dire les 89% restant.

2. L'effet des similitudes dans le reste du texte

Parmi les 861 élèves, 193 ont lu le texte en entier. On constate qu'aucun CP n'a lu tous les mots du texte. Cette analyse porte donc sur les élèves plus avancés en lecture, c'est-à-dire ceux qui ont lu le texte en entier et qui sont scolarisés du CE1 au CM2.

Quels sont les élèves les plus influencés dans le reste du texte?

On remarque que les lecteurs les plus avancés sont une majorité à avoir commis au moins une erreur due aux similitudes lors de la lecture du reste du texte : ils sont 53% (cf. tableau 40).

Tableau 40 : pourcentage et nombre d'élèves ayant été influencés ou non par une similitude dans le reste du texte selon la classe

	Aucune erreur	Au moins une erreur	Total	χ^2	<i>p.</i>
CE1	17% (1)	83% (5)	100% (6)	5,274	,153
CE2	33% (8)	67% (16)	100% (24)		
CM1	46% (31)	54% (37)	100% (68)		
CM2	53% (50)	47% (45)	100% (95)		
Ensemble	47% (90)	53% (103)	100% (193)		

Lecture : 17% des élèves scolarisés en CE1 (soit 1 sur 6) n'ont commis aucune erreur due à une similitude dans le reste du texte et 83% en ont commis au moins une (soit 5 sur 6).

Si on regarde les résultats dans le détail, par classe, on constate que la tendance est la même jusqu'en CM1 : ils sont une grande majorité à avoir été influencés par cette variable : 83% des élèves de CE1 qui ont lu le texte en entier a commis au moins une fois ce type d'erreur. En CE2, ils sont 67% et 54% en CM1.

Ainsi, on observe donc une continuité dans la progression des élèves lors de la lecture du reste du texte de *L'alouette*. Une continuité certes, non pas dans la performance, mais dans la proportion d'élèves qui a commis ce type d'erreur.

Dans quelle proportion ont-ils été influencés ?

Parmi les élèves qui se sont trompés dans le reste du texte, on remarque que plus de la moitié d'entre eux (52%) a commis plusieurs erreurs (*cf.* tableau 41).

Tableau 41 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par une similitude dans le reste du texte selon la classe

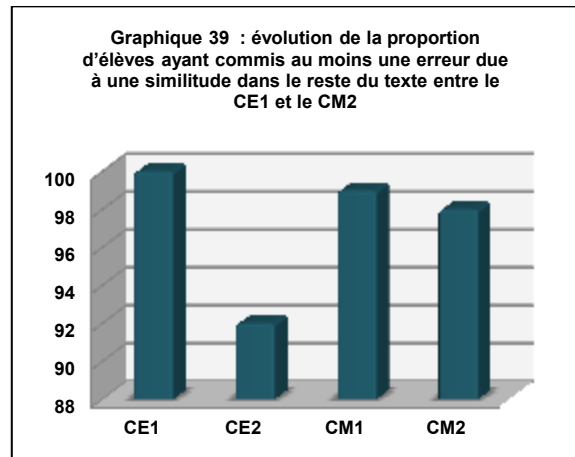
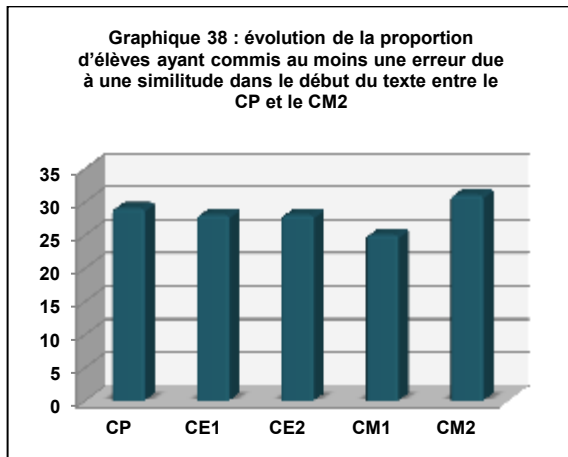
	Une seule erreur	Plusieurs erreurs	Total	χ^2	<i>p.</i>
CE1	80% (4)	20% (1)	100% (5)	2,850	,415
CE2	81% (13)	19% (3)	100% (16)		
CM1	60% (22)	40% (15)	100% (37)		
CM2	64% (29)	36% (16)	100% (45)		
Ensemble	66% (68)	34% (35)	100% (103)		

Lecture : 80% des élèves scolarisés en CE1 qui ont commis une erreur due aux similitudes n'ont commis qu'une seule erreur (soit 4 sur 5) et 20% en ont commis plusieurs (soit 1 sur 5).

Dans le détail, il apparaît que, mis à part en CM2, les élèves ont commis plusieurs erreurs de ce type. Ainsi, lorsqu'on analyse le reste du texte de *L'alouette*, on remarque que

non seulement les élèves scolarisés du CE1 au CM2 ont été majoritairement influencés par les similitudes, et les élèves qui se sont trompés ont essentiellement commis plusieurs erreurs de ce type.

Cette seconde analyse nous permet donc de constater que les résultats sont différents si on considère le début du texte (partie lue par tous les élèves) ou le reste du texte (partie lue par les lecteurs plus avancés). En effet, alors que dans le début du texte, les similitudes ne semblent pas avoir influencé la lecture de la majorité des élèves (*cf.* graphique 38), il apparaît que c'est l'inverse pour les lecteurs plus avancés qui, eux, ont essentiellement commis au moins une erreur de ce type (*cf.* graphique 39).



Ainsi, à un niveau débutant, on observe une régularité dans la progression des élèves puisque la performance est sensiblement la même d'un moment à l'autre du cursus scolaire. A un niveau plus avancé, on observe aussi une régularité dans la progression des élèves mais de façon négative : les élèves scolarisés entre le CE1 et le CM2 ont été essentiellement influencés par les similitudes.

Par conséquent, cette différence dans la progression des élèves (positive/négative) nous laisse penser qu'il existe une discontinuité quantitative entre les élèves plus avancés et les lecteurs débutants : ce sont les lecteurs plus avancés qui semblent être le plus influencés par les similitudes.

Le recours à un processus de traitement par analogie ne serait donc pas une caractéristique des apprentis lecteurs comme nous avons pu le supposer au départ.

III. Quels sont les types d'erreurs les plus induits par les similitudes?

Si les similitudes ont une influence lors de la lecture du texte de *L'alouette*, alors nous nous attendons à ce que les élèves les moins avancés en lecture commettent des erreurs lexicales et les plus âgés des erreurs grapho-sémantiques. Nous avons remarqué que, pour l'ensemble des élèves, la lecture de sept mots semble avoir été particulièrement influencée par les similitudes présentes dans le texte de *L'alouette* (cf. tableau 42).

Tableau 42 : mots sur lesquels le plus grand nombre d'élèves a commis une erreur due à une similitude selon le niveau de scolarisation.

	Début du texte		Reste du texte			
	36 Amie	38 Au	56 doigt	75 alouette	85 piaille	96 geai
CP (N= 143)	11% (10)	11% (9)	25% (13)	3% (1)		19% (4)
CE1 (N= 180)	2% (4)	5% (9)	12% (21)	4% (7)	11% (17)	27% (39)
CE2 (N= 174)	5% (8)	7% (13)	4% (7)	11% (19)	17% (29)	44% (70)
CM1 (N= 188)	5% (9)	6% (12)	3% (5)	15% (28)	21% (38)	46% (84)
CM2 (N= 176)	5% (9)	6% (10)	2% (4)	18% (32)	15% (27)	35% (61)
Ensemble (N= 861)	5% (40)	7% (53)	7% (50)	12% (87)	16% (111)	38% (258)



Mots situés trop loin dans le texte pour être considérés.

Lecture : En CP, parmi les élèves qui ont lu le mot « Amie », dix d'entre eux (soit 11%) l'ont mal lu à cause d'une similitude.

Alors que la lecture de trois mots (« amie », « au », et « doigt ») semble avoir été influencée par une similitude pour les lecteurs débutants, nous en comptons trois pour les lecteurs plus avancés (« alouette », « piaille » et « geai »).

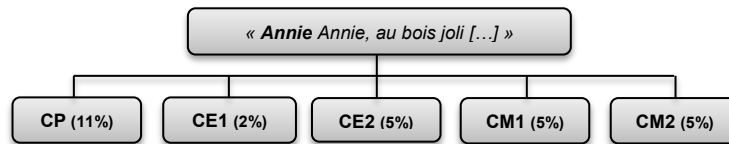
Les types d'erreurs des lecteurs débutants

Parmi les 55 mots lus en moyenne par les élèves de CP, deux ont été particulièrement mal prononcés à cause d'une similitude avec un mot placé juste avant. Dans l'ordre du texte, il s'agit de : « Amie » lu *Annie* et « au » lu *du*.

Le premier mot qui a été mal lu par le plus grand nombre d'élèves de CP est donc « Amie » dans la phrase « **Amie** Annie, au bois joli gamine le pinson » située au début du second paragraphe.

Annie amie, du renouveau, c'est le doux temps.
Amie Annie, au bois joli gamine le pinson.

Il apparaît que 11% des élèves de cette classe a lu *Annie Annie, au bois joli [...]*. La ligne du dessus commençant par « Annie amie », il se pourrait que les élèves, induits par la forme graphique et phonémique du mot « Amie » (la majuscule et les lettres du mot), aient cru que cette phrase commençait de la même façon, c'est-à-dire par le mot « Annie ».

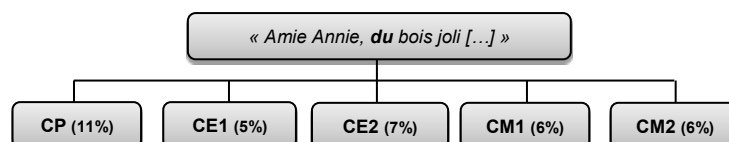


On constate que cette erreur concerne bien essentiellement les CP puisqu'à partir du CE1, elle est moins fréquente. On note néanmoins que les élèves de CM2 sont encore 5% à la commettre.

En second, on trouve le mot « au » qui se situe dans la même phrase: « Amie Annie, **au** bois joli gamine le pinson ».

Annie amie, du renouveau, c'est le doux temps.
Amie Annie, **du** bois joli gamine le pinson.

Comme nous l'avons remarqué pour l'erreur précédente, la phrase du dessus semble commencer de la même façon, il se pourrait donc que, là aussi, les élèves aient cru que l'amorce (« Annie amie ») se répétait et ont lu « du ». Nous avons donc regardé si les élèves qui ont commis cette erreur se sont trompés sur l'amorce de la phrase. Il apparaît que seulement trois élèves sur les 53 qui ont lu « au » : *du*, sont dans ce cas. Le début de la phrase ne serait donc pas ce qui les aurait incités à produire cette erreur. Il se pourrait alors que la typographie du texte les ait aussi influencés et les aient amenés à une confusion visuelle entre les lettres **a** et **d**. Toutefois, nous ne pouvons pas écarter l'hypothèse d'une influence de l'amorce de la phrase du dessus. En effet, il est possible que ces élèves n'aient pas remarqué que les deux premiers mots de la phrase aient été inversés et par conséquent, confondent le mot « au » en lisant *du* comme dans la phrase précédente.



Les élèves de CP sont les plus nombreux à avoir commis une erreur sur ce mot par suite d'une similitude puisqu'ils sont, là aussi, 11% à avoir lu *Amie Annie, du bois joli [...]*. En

CE1 leur nombre diminue et ils ne sont plus que 5%. Ils sont à peu près aussi nombreux en CE2 (7%), CM1 (6%) et CM2 (6%).

Par conséquent, on constate que, outre la proximité graphique et phonémique, la proximité spatiale semble jouer un rôle. En effet, pour le mot « Amie » lu *Annie*, on remarque que ces deux mots comportent les mêmes phonèmes initiaux et finaux. Pour le mot « au » lu *du*, il est possible que la similitude (graphique et phonémique) avec le syntagme situé juste au-dessus associé à la typographie du texte ait aussi joué un rôle. Bien que ces erreurs apparaissent aux autres niveaux scolaires, il semblerait que la ressemblance avec un syntagme précédemment lu, associée à une ressemblance graphique et phonémique du mot cible, aient particulièrement influencé la lecture des CP.

Outre ces erreurs spécifiques aux élèves de CP, d'autres portant sur des mots différents, sont spécifiques aux lecteurs plus avancés.

Les types d'erreur des lecteurs plus avancés

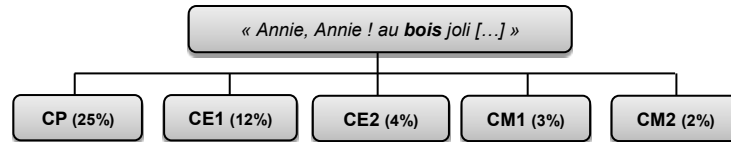
Du côté des lecteurs plus avancés, on remarque que la lecture de trois mots semble avoir été influencée par des similitudes avec un autre mot. On constate qu'un type d'erreur diminue alors qu'un autre augmente. Pour deux autres erreurs, on observe des variations irrégulières dans la proportion d'élèves qui se sont trompés.

➤ Le type d'erreur qui diminue : « *doigt* »

Il s'agit du mot « doigt » dans la phrase « Annie, Annie ! au **doigt** joli [...] » située dans le second paragraphe du texte.

Amie Annie, au bois joli gamine le pinson. Dans les buis, gîte une biche, au bois chantant. Annie, Annie! au doigt joli, une églantine laisse du sang :
--

En effet, le mot « doigt » a été lu *bois*. Ici aussi il se pourrait que la ressemblance avec un syntagme précédemment lu ait joué un rôle. En effet, deux lignes plus haut, on retrouve une amorce de phrase similaire d'un point de vue graphique et phonémique: « *Amie Annie, au bois joli [...]* ». Il est possible que les élèves aient cru que ce début de phrase se répétait et ont donc lu *bois* au lieu de « doigt ». De plus, on remarque que les mots « bois » et « doigt » comportent le même phonème final [wa]. Il est possible aussi que les élèves aient confondu les lettres b et d.



Bien qu'en moyenne les élèves en CP n'aient pas lu ce mot, on constate que 25% de ceux qui l'ont lu ont commis cette erreur. En CE1, ils sont moins nombreux mais sont encore 12%. A partir du CE2, le nombre d'élèves à avoir commis ce type d'erreur représente moins de 10%.

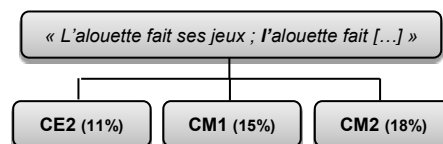
On remarque que cette erreur semble être, là aussi, due à une ressemblance avec un syntagme. Il se pourrait qu'ici aussi, le voisinage phonologique et la structure du mot aient joué un rôle. L'hypothèse d'une confusion visuelle pour les mots « bois » et « doigt » ne peut pas non plus être écartée : il est possible que les élèves les plus jeunes aient confondu les lettres initiales de ces deux mots.

➤ Le type d'erreur qui augmente : « *alouette* »

Il s'agit ici du mot « *alouette* », situé dans la première phrase du troisième paragraphe.

L'alouette fait ses jeux; **alouette** fait un nœud avec un rien de paille.

On remarque que le mot « **alouette** » a été lu *l'alouette*. On remarque que cette erreur porte sur un syntagme beaucoup plus petit, à la limite du mot, en comparaison avec les précédentes qui portent sur des syntagmes plus longs (plusieurs mots). En effet, ce mot apparaît trois fois dans le texte, dont deux fois sous la forme « l'alouette ». Ainsi, il est probable que les élèves aient retenu cette forme et, par conséquent, le prononcent de cette façon à la troisième apparition.



Il apparaît ici que la proportion d'élèves ayant commis cette erreur augmente avec le niveau scolaire. En effet, alors que 11% des élèves de CE2 l'ont prononcé comme tel, ils sont encore un peu plus nombreux en CM1 (15%) et en CM2 (18%). Il est donc possible que cette erreur soit une erreur des lecteurs les plus avancés qui se seraient trompés par inadvertance.

Au-delà des erreurs qui diminuent ou augmentent avec le niveau scolaire, d'autres semblent varier de façon irrégulière selon la classe fréquentée.

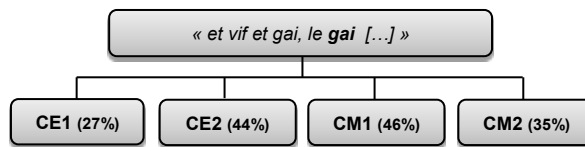
➤ Les types d'erreurs à progression irrégulière : « *piaille* » et « *geai* »

Dans cette catégorie, deux mots semblent avoir été mal lus suite à une similitude : il s'agit des mots « *piaille* » et « *geai* ».

Celui sur lequel le plus grand nombre d'élèves s'est trompé est le mot « *geai* » situé au début du troisième paragraphe.

L'hirondeau *piaille* sous la pente des bardeaux et, vif et gai, le **geai**

Il se pourrait que le mot qui le précède (« *gai* ») ait influencé sa lecture : les élèves ont donc lu *gai* au lieu de « *geai* ». On remarque que, là aussi, cette erreur porte sur un mot et non sur un syntagme. Notons par ailleurs qu'ici ces deux mots ne diffèrent que par une seule lettre (-e). Il se pourrait donc qu'outre la proximité spatiale, la proximité graphique et donc aussi phonémique, aient aussi influencé la lecture des élèves. On peut aussi se demander, pour les élèves les plus jeunes, si cette erreur n'est pas aussi due à la non maîtrise des sons [g] et [ʒ].

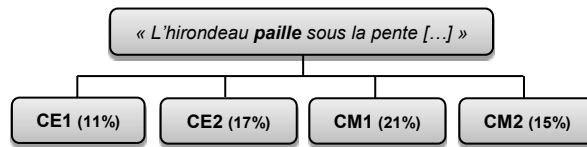


Ainsi, on s'aperçoit que les élèves de CE1 sont 27% à avoir commis cette erreur. En CE2, leur nombre augmente puisqu'ils sont 44%. Légèrement plus nombreux en CM1, ils sont 46%. Enfin, en CM2, les élèves qui ont commis cette erreur sont un peu moins nombreux puisqu'ils sont 35%. Le fait que le nombre d'élèves qui ont commis cette erreur diminue en CM2 laisse penser qu'au milieu de l'apprentissage de la lecture ce ne soit plus le graphème contextuel qui pose difficulté mais la similitude. Il est donc possible que pour les élèves les plus âgés ce soit plutôt une erreur d'inattention relative à leur lecture rapide due à cette fausse répétition.

Le second mot est « *piaille* », situé à la fin de la première phrase du troisième paragraphe.

L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau **piaille** sous la pente des bardeaux et, vif et gai, le geai

Le mot similaire se trouve juste avant : « paille ». Ainsi, les élèves ont été amenés à lire *l'hirondeau paille* [...] au lieu de « l'hirondeau **piaille** [...] ». On remarque là aussi que ces deux mots ne se distinguent que par une lettre (le « -i »). Au-delà de la proximité spatiale, la proximité grapho-phonémique a donc pu aussi les influencer.



On note que la part d'élèves qui ont commis cette erreur varie légèrement entre le CE1 et le CM2 puisqu'elle oscille entre 11% et 21%. En effet, en CE1, ils sont 11% à s'être trompé et à lire *paille*. En CE2, cette proportion augmente légèrement puisqu'ils sont 17%. En CM1, ils sont encore un peu plus nombreux puisque 21% d'entre eux ont commis cette erreur. En fin d'école primaire, bien que leur nombre diminue légèrement, ils sont encore 15% à avoir commis une erreur due à la similitude. Comme pour l'erreur précédente, il se pourrait que les causes soient différentes selon le niveau d'avancement en lecture. En effet, les lecteurs débutants ont sans doute eu des difficultés à lire ce mot à cause de sa structure syllabique alors que les lecteurs plus avancés ont peut être commis cette erreur par manque d'attention.

En bref, on remarque que pour les élèves un peu plus avancés en lecture, ce n'est plus une similitude avec un syntagme mais davantage avec un mot qui semble les avoir influencés. Il se pourrait que, outre la proximité spatiale des mots, ce soit davantage la proximité orthographique qui ait joué un rôle. Pour les mots « piaille » (lu *paille*) et « geai » (lu *gai*), il est possible que leur structure syllabique ait amené les élèves à traiter ces mots comme ceux qui les précèdent et qui sont plus faciles à lire. La part des élèves qui ont commis ces erreurs étant relativement encore importante en fin d'école primaire, nous nous demandons si en cours d'apprentissage de la lecture, les causes ne changent pas. Ainsi, il est probable que pour les lecteurs débutants ces erreurs soient dues à une difficulté de décodage entraînant la répétition d'un mot (plus facile à lire) précédemment lu alors que pour les lecteurs plus avancés, il se pourrait que ce soit des erreurs dues à un manque d'attention.

Cette analyse des types d'erreurs nous permet de constater que les élèves se comportent différemment face aux similitudes. En effet, il semblerait que les élèves débutant dans

l'apprentissage de la lecture, soient davantage influencés par les ressemblances entre syntagmes alors que les élèves les plus avancés en lecture sont davantage influencés par la ressemblance de mots. De plus, il semblerait que, au-delà de la proximité spatiale, la proximité phonologique ou orthographique des mots influence la lecture. Des variables comme la fréquence, la familiarité et la structure du mot peuvent jouer aussi un rôle pour ce type d'erreur (cf. « *L'influence des variables linguistiques* »). En effet, il est possible que les élèves aient davantage tendance à commettre une erreur due à une similitude lorsqu'ils rencontrent un mot plus difficile à lire ou moins fréquent que celui qu'ils viennent de lire. Une autre hypothèse s'ajoute : celle de la rapidité de lecture. En effet, comme nous l'avons vu précédemment (cf. Chapitre 6 « *La performance lexicale* »), les élèves lisent de plus en plus vite au fur et à mesure de l'apprentissage. De ce fait, la rapidité de lecture associée à la ressemblance grapho-phonémique, à la proximité spatiale, à la fréquence et à la complexité de certains mots fait que les élèves ont pu commettre ces erreurs par manque de concentration.

CONCLUSION

Ce chapitre avait pour objectif de voir dans quelle mesure les similitudes présentes dans le texte de *L'alouette* influencent la lecture des élèves et, de façon plus générale, quelles sont les informations qu'elles nous apportent sur l'apprentissage de la lecture, notamment en termes de continuité ou de discontinuité. Ce test repose sur une conception classique de la lecture, c'est-à-dire qu'il mesure essentiellement le niveau de décodage. Partant de ce constat, certaines erreurs de lecture seraient donc le résultat d'une défaillance dans le décodage. Nous avons donc supposé qu'au début de l'apprentissage de la lecture, les élèves auraient davantage recours à un traitement par analogie lors de la lecture de certains mots. La maîtrise du code devait faire diminuer ce type d'erreur dans le temps.

Dans une première analyse, nous avons constaté que ces fausses répétitions semblent avoir influencé les élèves à tous les niveaux scolaires mais à des degrés différents : les élèves de cycle 3 semblent être le plus touchés et commettent davantage d'erreurs dues aux similitudes, que les autres. L'analyse détaillée par classe montre que c'est plus précisément à partir du CE1 que les élèves sont le plus influencés par les similitudes. On observe donc ici, une rupture dans la progression des élèves. De plus, il apparaît que la proportion d'élèves à

avoir commis ce type d'erreur augmente d'année en année. La maîtrise du code ne ferait donc pas diminuer les erreurs dues aux similitudes comme nous avons pu le supposer au départ.

Toutefois, ces résultats sont à prendre avec précaution. En effet, les élèves n'ayant pas tous lu le même nombre de mots, réunir dans une même analyse tous les élèves est discutable. A partir de ce constat, et pour plus de rigueur, nous avons repris l'examen des données en analysant d'abord la partie du texte que tous les élèves ont lue, ce qui correspond au début du texte puis le reste du texte.

On s'aperçoit alors que quand on analyse seulement le début du texte les résultats sont différents : les élèves de cycle 3 se comportent de la même façon que ceux en cycle 2 : ils sont une majorité à ne pas être influencés par ces similitudes. En revanche, lorsqu'on ne tient compte que du reste du texte, on s'aperçoit que les résultats s'inversent : les élèves sont une majorité à avoir commis ce type d'erreur quel que soit le niveau scolaire. La discontinuité observée précédemment entre les lecteurs débutants et les plus avancés semble se confirmer ici : ce sont les lecteurs plus avancés qui seraient le plus influencés par les similitudes. Le recours à un processus de traitement par analogie ne serait donc pas une caractéristique des apprentis lecteurs comme nous avons pu le supposer au départ.

Néanmoins, certaines erreurs de lecture peuvent s'expliquer par un effet d'autres variables (l'effet des dessins, l'effet du co-texte sémantique). C'est pourquoi, pour plus de rigueur, nous avons repris les deux analyses précédentes sans tenir compte des mots dont la lecture erronée peut être due à d'autres facteurs.

Les résultats obtenus lors de cette troisième analyse, confirment l'effet des similitudes chez les lecteurs plus avancés. En effet, même si on ne tient pas compte des mots dont la lecture peut être influencée par d'autres facteurs, les résultats ne changent pas. Ainsi, contrairement à notre supposition de départ, les similitudes n'influencent pas les lecteurs débutants mais plutôt les plus avancés. Il ne s'agirait donc pas seulement d'une stratégie compensatoire à une défaillance du décodage. Il se pourrait que les stratégies mises en place diffèrent d'un moment à un autre du cursus scolaire. Pour tenter de répondre à cette hypothèse, nous avons ensuite mené une étude qualitative des différents types d'erreurs commises. Partant d'une conception classique de la lecture, nous avons supposé que l'étude des types d'erreurs devait faire apparaître une évolution dans le rapport code/sens : les lecteurs débutants devant faire des erreurs lexicales alors que pour les plus avancés les erreurs seraient plutôt d'ordre grapho-sémantique.

Si on regarde de plus près les types d'erreurs commises pour chaque classe, on se rend compte qu'il existe une certaine discontinuité dans les stratégies adoptées par les élèves. En effet, outre la proximité spatiale des mots, il semblerait que la proximité grapho-phonémique les influence également.

Ainsi, on remarque que pour les élèves de cycle 2, les mots sur lesquels les élèves se sont trompés sont essentiellement des voisins phonologiques² des mots qui précèdent, c'est-à-dire qu'ils ne diffèrent que par un ou deux phonèmes (Amie/Annie, doigt/bois). De plus, on remarque que ces mots font partie d'un syntagme similaire précédemment lu. On suppose alors que les lecteurs débutants, à cause de cette similitude phonologique, font une similitude les amenant à effectuer un traitement par analogie avec le mot antérieur. Ce constat se rapproche de celui de Goswami (1986) qui montre que les apprentis lecteurs utilisent précocement l'analogie pour lire des mots nouveaux. Toutefois, elle montre que les élèves utilisent surtout l'analogie des rimes (une procédure de nature non lexicale). Or, nous constatons que nos élèves s'appuient aussi sur d'autres indices : le phonème final (doigt/bois) et parfois le phonème initial et final (Annie/amie). Pour les élèves des niveaux suivants, c'est davantage le voisinage orthographique qui semble jouer un rôle (piaille → paille, geai → gai). De plus, les mots cibles semblent être plus difficiles et moins connus des mots qui précèdent. On suppose alors que pour les lecteurs un peu plus avancés, cette similitude orthographique entraîne une similitude les amenant à traiter, là aussi, de façon analogique les mots les moins connus. Ici, ce constat semble se rapprocher de celui de Marsh et *al.* (1981) et de Morton (1989) qui considèrent que l'analogie interviendrait plus tard, dans la phase orthographique, car elle nécessite une maîtrise des correspondances graphèmes-phonèmes. Contrairement à ce que suppose Goswami, cette procédure serait donc de nature lexicale.

Ainsi, d'après nos résultats, il semblerait que dès les premières années de l'apprentissage de la lecture, les élèves aient recours à une stratégie lexicale, ce qui va à l'encontre des modèles développementaux qui supposent que cette procédure n'arrive qu'en fin d'apprentissage.

Par ailleurs, au-delà de la proximité spatiale et grapho-phonémique, nous avons émis l'hypothèse que la rapidité de lecture pouvait aussi expliquer ces erreurs. En effet, lors d'un chapitre précédent (*cf.* Chapitre 6 « *Performance lexicale* »), nous avons constaté que plus un élève avance dans l'apprentissage de la lecture et plus il lit rapidement. Ainsi, il se pourrait que les erreurs constatées en cours et en fin d'apprentissage de la lecture soient des erreurs par

² Deux mots sont considérés comme voisins phonologiques lorsqu'ils sont constitués des même phonèmes à un près.

inadvertance puisqu'elles ne semblent porter que sur un élément (alouette → l'alouette, paille → paille, geai → gai).

Comme pour le chapitre précédent, le rôle de ces similitudes est donc difficile à cerner. Toutefois, cette analyse semble aller dans le sens du constat effectué dans le précédent chapitre, à savoir que l'apprentissage de la lecture ne se déroule pas nécessairement de façon strictement discontinue, c'est-à-dire par étapes dont chacune serait caractérisée par une stratégie donnée. Comme nous l'avons dit, la réalité semble plus complexe : bien que les élèves semblent être influencés à des degrés différents par ces fausses répétitions, le phénomène de similitude apparaît à tous les niveaux scolaires. La constance dans ce comportement va donc dans le sens d'un apprentissage continu. Néanmoins, lorsqu'on regarde les stratégies adoptées, on note une certaine discontinuité.

L'analyse des autres types d'erreurs va nous permettre de poursuivre la réflexion afin de savoir dans quelle mesure l'apprentissage de la lecture est continu ou discontinu.

Chapitre 13

L'INFLUENCE DU CO-TEXTE SEMANTIQUE LORS DE LA LECTURE DU TEXTE DE *L'ALOUETTE*

Plan du chapitre

INTRODUCTION.....	285
PREDICTIONS	286
METHODOLOGIE D'ANALYSE	286
I. L'effet du co-texte sémantique	288
<i>Quels sont les élèves les plus influencés par le co-texte sémantique ?.....</i>	<i>288</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>288</i>
1. L'influence du co-texte sémantique en cycle 2	289
<i>Quels sont les élèves les plus influencés par le co-texte sémantique en cycle 2 ?.....</i>	<i>289</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>289</i>
2. L'influence du co-texte sémantique en cycle 3.....	290
<i>Quels sont les élèves les plus influencés par le co-texte sémantique en cycle 3 ?.....</i>	<i>290</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>291</i>
II. L'influence du co-texte sémantique dans les différentes parties du texte.293	
1. L'influence du co-texte sémantique dans le début du texte	289
<i>Quels sont les élèves les plus influencés par le co-texte sémantique dans le début du texte?</i>	<i>289</i>
2. L'influence du co-texte sémantique dans le reste du texte	290
<i>Quels sont les élèves les plus influencés par le co-texte sémantique dans le reste du texte?</i>	<i>290</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>291</i>
III. Quels sont les types d'erreurs les plus induits par le co-texte sémantique297	
<i>Le type d'erreur qui varie de façon irrégulière</i>	<i>298</i>
<i>Les types d'erreurs qui augmentent.....</i>	<i>299</i>
CONCLUSION.....	300

INTRODUCTION

Le rôle du contexte en lecture a fait l'objet de plusieurs recherches dont nous proposons de rendre compte ci-dessous (effet d'amorçage sémantique, rôle du contexte dans l'identification du mot). Mais avant, nous tenons à préciser que dans cette étude, nous emploierons de préférence le terme co-texte, qui fait davantage référence au contexte immédiat plutôt que le terme « contexte » qui se réfère plutôt aux éléments extra-textuels.

Ainsi, le rôle que joue la sémantique a été étudié à travers des tâches d'amorçage sémantique au cours desquelles il est présenté successivement à l'élève deux mots écrits : un mot cible et une amorce sensée faciliter la compréhension du premier mot. Les résultats des recherches font apparaître que les effets d'amorçage sémantique semblent diminuer avec l'âge et, par conséquent, le niveau d'expertise en lecture. Autrement dit, les bons lecteurs ne seraient pas ceux qui ont le plus recours aux informations sémantiques.

D'autres recherches, menées sur le rôle du contexte sémantique, font apparaître des résultats similaires.

En 1981, Perfetti et Roth, ont mené une recherche sur des élèves scolarisés en 3^e et 4^e année d'école primaire afin de voir de quelle façon le contexte influence la lecture. Dans leur étude, ces chercheurs font varier le contexte de trois façons différentes afin de savoir si les items dans un contexte fortement prédictible sont mieux lus que ceux qui apparaissent dans un contexte non prédictible ou non congruent. Les résultats font apparaître que les lecteurs plus habiles plus âgés sont les moins dépendants du contexte.

Cette même année, Stanovich (1981), effectue une expérience basée sur l'hypothèse suivante : si le contexte joue un rôle important dans l'identification du mot, alors cet effet devrait être plus fort dans le cas de congruence que de non-congruence puisque le lecteur aura pu anticiper sa réponse. Dans cette optique, il fait lire à des enfants âgés de 9 à 11 ans des phrases non terminées. Le mot à trouver apparaît donc soit dans un contexte congruent ou non congruent. Il constate que ce sont les lecteurs les plus jeunes qui profitent au mieux des effets facilitateurs du contexte.

Les résultats de ces deux expériences sont similaires : ce seraient les lecteurs les plus jeunes et par conséquent les moins habiles qui dépendraient le plus du contexte.

On suppose donc que le contexte et plus particulièrement le co-texte sémantique joue un rôle différent selon le niveau en lecture.

PREDICTIONS

On se demande alors **dans quelle mesure les erreurs des enfants, dues au co-texte sémantique, lors de la lecture du texte de *L'alouette* font apparaître un caractère discontinu ou continu dans l'apprentissage de la lecture.**

Lefavrais a construit le test de *L'alouette* sur une conception classique de la lecture, c'est-à-dire qu'il mesure essentiellement le décodage ; ce qui correspond dans les modèles développementaux à la phase alphabétique du modèle de Frith (1985). De ce point de vue, certaines erreurs apparaissent donc comme une défaillance dans les capacités de décodage. Dans cette optique et compte tenu des résultats des recherches menées sur le rôle du contexte sémantique, nous supposons que cette défaillance serait alors compensée par un recours au co-texte sémantique. Nous supposons ainsi que l'apprentissage du code devrait faire diminuer ce type d'erreur dans le temps. Du point de vue qualitatif, nous nous attendons à ce que la nature des erreurs rende compte d'une évolution dans le rapport code/sens : des erreurs de type lexical chez les plus jeunes et des erreurs de type grapho-sémantique chez les plus âgés.

METHODOLOGIE D'ANALYSE

Nous avons remarqué qu'à certains endroits du texte les mots à reconnaître ne sont pas potentiellement ceux que l'on croit être à cause du co-texte sémantique dans lequel ils s'inscrivent. Ce type d'erreur est donc induit de façon implicite par le texte. Nous avons répertorié neuf sources potentielles d'erreur (*cf. cartographie 7 du texte, p.287*):

Tableau 43 : types d'erreurs commis à cause du co-texte sémantique




Mots/Syntaxe	Type d'erreur
« alouette fait un nœud »	<i>Alouette fait un nid</i> <i>Alouette fait un œuf</i>
« un bec luisant qui trille »	<i>Un bec luisant qui brille</i>
« le ciel rougit ses eaux »	<i>Le soleil rougit ses eaux</i>
« reflète les poisons »	<i>Reflète les poissons</i>
« une amarre est décochée »	<i>Une amarre est décrochée</i>
« des matelots jettent leurs cassettes »	<i>Des matelots jettent leurs casquettes</i>
« au clair de lune »	<i>Au clair de la lune</i>
« mon amie annie »	<i>Mon amie Pierrot</i>
« prête-moi la plume »	<i>Prête-moi ta plume</i>

Mots dont la source d'erreurs peut provenir de plusieurs influences¹



¹ Bien que nous en fassions le détail ici, ces mots ne sont pas retenus pour cette analyse à cause de leur surdétermination. Ils seront étudiés au cours du Chapitre 14 « *L'influence des mots surdéterminés* »

Cartographie 7 : mots sur lesquels les élèves ont commis une erreur à cause du co-texte sémantique dans le texte de L'alouette




L'alouette.



Sous la mousse ou sur le toit,
dans les haies vives ou le chêne fourchu,
le printemps a mis ses nids.
Le printemps a nids au bois.


animaux → Annie amie, du renouveau, c'est le doux temps.
Amie Annie au bois joli gamine le pinson.
Dans les buis, gîte une biche, au bois chantant.
Annie Annie! au doigt joli, une églantine laisse du sang :
au bout du temps des féeries viendra l'ennui.

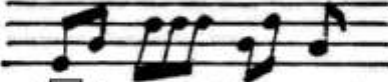
L'alouette fait ses jeux; alouette fait un *nid* [noeud] avec un rien de paille.
L'hirondeau piaille sous la pente des bardeaux et, vif et gai, le geai
sur l'écaille argentée du bouleau, promène un brin d'osier.
Au verger, dans le soleil matinal, goutte une [pompe] dégelée.
On voit un bec luisant qui trille éperdument des notes claires
et, dans les pampres d'or que suspend la grille antique, *pomme*
on surprend des rixes de moineaux.
Au potager s'alignent les cordeaux; l'if est triste à l'horizon
et lourd et lent l'envol des corbeaux.

Un lac étire ses calmes rives et, quand le soir descend,
le miroir de ses eaux reflète les [poissons] des brignoles perfides.
Et, quand descend le soir, quand joue la pourpre du couchant,
le ciel rougit ses eaux.
Dans la moire de l'eau danse l'ombre d'un écueil.
Tout est cris! Tout est bruits!



casquettes → Une amarre est décrochée... une barque est arrimée... des matelots
jettent leurs cassettes sur le rivage...
Tout est cris! Tout est bruits!
Au clair de la lune mon ami Pierrot...
Au clair de lune mon amie annie...
Au clair de la lune mon ami Pierrot, prête-moi la plume pour écrire un mot.



Pierrot → Au clair de la lune
ta → la plume pour écrire un mot

o u e i a
le la les un dans des do ti pu mi

I. L'influence du co-texte sémantique

Quels sont les élèves les plus influencés par le co-texte sémantique ?

Dans l'ensemble, il apparaît que sur les 861 élèves de notre échantillon, 53% d'entre eux (soit 455) semblent avoir été influencés au moins une fois par le co-texte sémantique (cf. tableau 44).

Tableau 44 : pourcentage et nombre d'élèves ayant commis ou non des erreurs dues au co-texte sémantique selon le cycle de scolarisation

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
Cycle 2	84% (271)	16% (52)	100% (323)		
Cycle 3	25% (135)	75% (403)	100% (538)	277.752	.000
Ensemble	47% (406)	53% (455)	100% (861)		

Lecture : 84% des élèves scolarisés en cycle 2 (soit 271 sur 323) n'ont commis aucune erreur due au co-texte sémantique et 16% en ont commis au moins une (soit 52 sur 323).

Dans le détail, il apparaît que le nombre d'élèves qui ont commis des erreurs ou non varie avec le cycle de scolarisation. Ainsi, alors qu'en cycle 2, la majorité des élèves ne semble pas avoir été influencée (84%), on remarque que c'est l'inverse pour les élèves de cycle 3 : ils sont une majorité à avoir été influencés au moins une fois (75%). Il semble donc qu'il existe une discontinuité quantitative entre le cycle 2 et le cycle 3. En effet, on constate qu'au lieu d'augmenter, la performance des élèves diminue puisque la proportion d'élèves qui a commis au moins une erreur de ce type augmente fortement entre les deux cycles (+ 54 points). Les élèves plus âgés sont donc plus nombreux à avoir commis ce type d'erreur que les plus jeunes.

Dans quelle proportion ont-ils été influencés ?

Parmi les 532 élèves dont la lecture a été influencée par le co-texte sémantique, on remarque que 71% d'entre eux ont commis plusieurs erreurs (cf. tableau 45).

Tableau 45 : pourcentage et nombre d'élèves ayant commis une ou plusieurs erreur(s) due(s) au co-texte sémantique selon le cycle de scolarisation

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
Cycle 2	94% (49)	6% (3)	100% (52)		
Cycle 3	68% (276)	32% (127)	100% (403)	13.723	.000
Ensemble	71% (325)	29% (130)	100% (455)		

Lecture : 94% des élèves scolarisés en cycle 2 qui ont commis au moins une erreur due au co-texte sémantique, ont commis qu'une seule erreur (soit 49 sur 52) et 6 % en ont commis plusieurs (soit 3 sur 52).

Dans le détail, ces résultats diffèrent des précédents puisque les élèves de cycle 2 sont une majorité (94%) à n'avoir commis qu'une seule erreur, tout comme les élèves en cycle 3 qui sont 68%.

Toutefois, si on regarde les élèves qui ont commis plusieurs erreurs, on constate que les élèves en cycle 3 sont les plus nombreux : 32% contre 6% des élèves en cycle 2.

Ainsi, la discontinuité constatée précédemment entre le cycle 2 et le cycle 3 semble se confirmer puisque non seulement les élèves en cycle 3 sont plus nombreux à avoir été influencés par le co-texte sémantique, mais ils commettent aussi davantage d'erreurs de ce type que ceux en cycle 2

Voyons maintenant ce qu'il en est plus précisément pour les élèves scolarisés dans les classes en cycle 2 et cycle 3.

1. L'influence du co-texte sémantique en cycle 2

Sur les 861 élèves de notre échantillon, 323 sont scolarisés en cycle 2 dont 143 sont en CP et 180 en CE1.

Quels sont les élèves les plus influencés par le co-texte sémantique en cycle 2 ?

Rappelons que sur les 323 élèves scolarisés en cycle 2, 84% d'entre eux n'ont commis aucune erreur (cf. tableau 46).

Tableau 46 : pourcentage (et nombre d'élèves) ayant été influencés ou non par le co-texte sémantique selon la classe en cycle 2

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CP	99% (142)	1% (1)	100% (143)		
CE1	72% (129)	28% (51)	100% (180)	43.031	.000
Ensemble	84% (271)	16% (52)	100% (323)		

Lecture : 99 % des élèves scolarisés en CP (soit 142 sur 143) n'ont commis aucune erreur due au co-texte sémantique et 1 % en a commis au moins une (soit 1 sur 143).

Dans le détail, bien que la majorité des élèves de CP (99%) et de CE1 (72%) n'aient pas commis d'erreur liée au co-texte sémantique, on constate toutefois que la proportion d'élèves qui a commis ce type d'erreur augmente fortement d'une classe à l'autre : ils passent de 1% à 28%. Cette forte augmentation (+27 points) laisse penser qu'il existe une discontinuité quantitative dans la progression des élèves, plus particulièrement entre le CP et le CE1.

Dans quelle proportion ont-ils été influencés ?

Parmi les 323 élèves scolarisés en cycle 2, 52 d'entre eux ont été influencés au moins une fois par le co-texte sémantique. On remarque que les élèves ont majoritairement (94%) commis une seule erreur de ce type (cf. tableau 47)

Tableau 47 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par le co-texte sémantique selon la classe en cycle 2

	Une seule erreur	Plusieurs erreurs	Total	χ^2	<i>p.</i>
CP	100% (1)	0% (0)	100% (1)		
CE1	94% (48)	6% (3)	100% (51)	.000	1.000
Ensemble	94% (49)	6% (3)	100% (52)		

Lecture : 100% des élèves scolarisés en cycle 2 qui ont commis au moins une erreur due au co-texte sémantique, ont commis qu'une seule erreur.

En effet, qu'ils soient scolarisés en CP ou en CE1, la plupart des élèves n'ont été influencés qu'une seule fois par le co-texte sémantique. On observe donc ici une continuité entre les CP et les CE1, en ce qui concerne la quantité d'erreurs de ce type qui ont été commises.

Ainsi, il apparaît qu'en cycle 2 la tendance à commettre des erreurs en lien avec le co-texte sémantique est faible. Néanmoins, on remarque que, pour ceux qui ont pu commettre ce type d'erreur, il existe un écart important entre les CP et les CE1 ; ce qui laisse croire en l'existence d'une rupture dans l'apprentissage de la lecture entre ces deux classes.

2. L'influence du co-texte sémantique en cycle 3

Sur les 538 élèves scolarisés en cycle 3, 174 sont en CE2, 188 en CM1 et 176 en CM2.

Quels sont les élèves les plus influencés par le co-texte sémantique en cycle 3 ?

Rappelons que sur les 538 élèves scolarisés en cycle 3, 75% d'entre eux ont commis des erreurs en lien avec le co-texte sémantique (cf. tableau 48).

Tableau 48 : pourcentage et nombre d'élèves ayant été influencés ou non par le co-texte sémantique selon la classe en cycle 3

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CE2	44% (76)	56% (98)	100% (174)		
CM1	21% (39)	79% (149)	100% (188)	51.516	.000
CM2	11% (20)	89% (156)	100% (176)		
Ensemble	25% (135)	75% (403)	100% (538)		

Lecture : 44% des élèves scolarisés en CE2 (soit 76 sur 174) n'ont commis aucune erreur due au co-texte sémantique et 56% en ont commis au moins une (soit 98 sur 174).

Dans le détail, 56% des élèves en CE2 ont commis au moins une fois ce type d'erreur. Ils sont 79% en CM1 et 89% en CM2. Il apparaît donc que, du CE2 au CM2, les élèves qui ont commis ce type d'erreur deviennent majoritaires. On remarque une augmentation de 23 points entre le CE2 et le CM1 et de 10 points entre le CM1 et le CM2. Ainsi, au-delà de la discontinuité observée entre le CP et le CE1 en cycle 2, on constate ici un accroissement du nombre d'élèves qui a commis des erreurs de ce type en cycle 2.

Dans quelle proportion ont-ils été influencés ?

Rappelons que parmi les 538 élèves en cycle 3, 403 d'entre eux ont été influencés au moins une fois par le co-texte sémantique (*cf.* tableau 49).

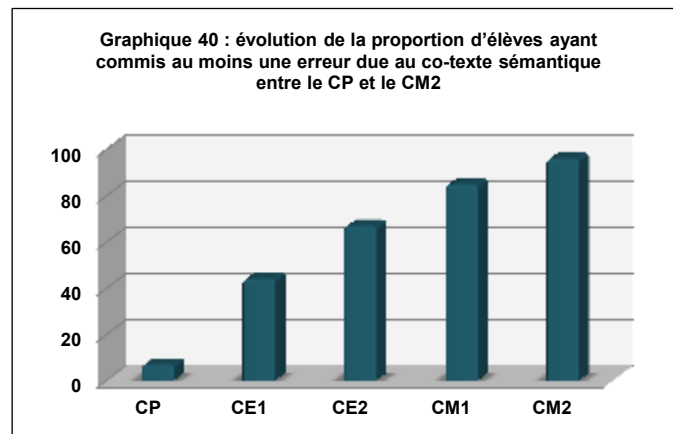
Tableau 49 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par la sémantique selon la classe en cycle 3

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CE2	82% (80)	18% (18)	100% (98)		
CM1	74% (110)	26% (39)	100% (149)	22.713	.000
CM2	55% (86)	45% (70)	100% (156)		
Ensemble	68% (276)	32% (127)	100% (403)		

Lecture : 82% des élèves scolarisés en CE2 qui ont commis au moins une erreur due au co-texte sémantique, ont commis qu'une seule erreur (soit 80 sur 98) et 18% en ont commis plusieurs (soit 18 sur 98).

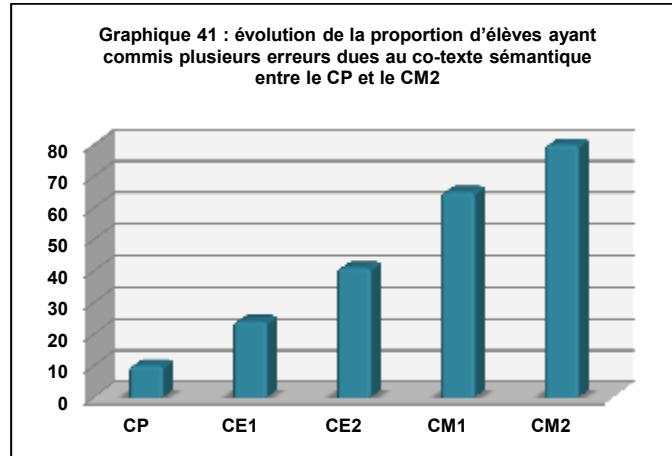
Contrairement à ce que nous avons observé en cycle 2, on constate que le fait de commettre une ou plusieurs erreurs dépend de la classe fréquentée. Ainsi, bien que la majorité des élèves qui se sont trompés ont commis qu'une seule erreur de ce type (ils sont 82% en CE2, 74% en CM1 et 55% en CM2), on constate que plus un élève avance dans la scolarisation et plus il a tendance à commettre plusieurs erreurs : alors que 18% des élèves de CE2 ont commis plusieurs erreurs, ils sont 26% en CM1 et 45% en CM2.

En **conclusion**, cette première partie nous permet de constater qu'**il existe des discontinuités quantitatives dans la progression** des élèves, laissant entrevoir des ruptures dans l'apprentissage de la lecture. En effet, il apparaît un écart important entre les élèves en cycle 2 et en cycle 3 : alors que les premiers sont une majorité à ne pas avoir commis d'erreurs liées au co-texte sémantique, c'est l'inverse pour les seconds. Si on synthétise les résultats vus partiellement auparavant, on constate que la proportion d'élèves ayant commis ce type d'erreur augmente fortement entre le CP et le CE1 (+27 points) (*cf.* graphique 40).



Ainsi, il semble que plus un élève avance dans l'apprentissage de la lecture et plus il a tendance à utiliser le co-texte sémantique (*cf.* graphique 39). Nous avons supposé que les erreurs de ce type seraient le résultat d'une défaillance du décodage et que par conséquent, le recours au co-texte sémantique serait un processus compensatoire. Or la maîtrise du code ne ferait pas diminuer ce type d'erreurs, au contraire, elle les ferait augmenter.

D'autre part, la présence de ce type d'erreur dès le CP laisse penser que les élèves essaient de donner du sens à leur lecture dès les premières années de l'apprentissage. Ce constat va donc à l'encontre des modèles développementaux de l'apprentissage de la lecture qui supposent que les élèves passent d'abord par une étape exclusive de décodage. Ce constat se confirme par les résultats observés en ce qui concerne le nombre d'erreurs commises. En effet, non seulement, les élèves sont de plus en plus influencés par le co-texte sémantique mais ils commettent davantage d'erreurs (*cf.* graphique 41). Les lecteurs débutants ne feraient donc pas que décoder.



Cependant, comme pour le chapitre précédent, nous nous devons de rappeler que les élèves n'ont pas tous lu le même nombre de mots. De plus, on remarque que les mots qui apparaissent dans un co-texte sémantique incohérent se situent davantage dans la seconde moitié du texte (*cf.* cartographie, p.287). Par conséquent, les résultats observés dans cette analyse semblent être biaisés puisque durant le temps imparti (trois minutes), seuls les élèves de CP n'ont pas lu la seconde moitié du texte. C'est pourquoi, afin que les résultats puissent être comparables, nous avons repris cette analyse (comme dans le chapitre précédent) en ne tenant compte que de la partie lue par tous les élèves, c'est-à-dire le début du texte. Nous nous sommes donc demandée dans quelle mesure le co-texte sémantique a influencé les élèves du CP au CM2 lors de la lecture du début du texte de *L'alouette*.

II. L'influence du co-texte sémantique dans les différentes parties du texte

1. L'influence du co-texte sémantique dans le début du texte

Quels sont les élèves les plus influencés par le co-texte sémantique dans le début du texte ?

Il apparaît que lorsqu'on ne tient compte que de la partie lue par tous les élèves, les résultats diffèrent : le co-texte sémantique du début du texte n'a pas influencé leur lecture. Aucun élève n'a commis au moins une erreur de ce type dans cette partie (*cf.* tableau 50).

Tableau 50 : pourcentage et nombre d'élèves ayant été influencés ou non par le co-texte sémantique dans le début du texte selon la classe

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CP	100% (143)	0% (0)	100% (143)	Test statistique impossible	
CE1	100% (180)	0% (0)	100% (180)		
CE2	100% (174)	0% (0)	100% (174)		
CM1	100% (188)	0% (0)	100% (188)		
CM2	100% (176)	0% (0)	100% (176)		
Ensemble	100% (861)	0% (0)	100% (861)		

Lecture : 100 % des élèves scolarisés en CP (soit 143 sur 143) n'ont commis aucune erreur due à la sémantique dans le début du texte.

Il semble donc que, si on ne tient compte que de la partie du texte lue par tous les élèves, on observe une certaine continuité dans la progression des élèves puisque leur performance ne varie pas d'une classe à l'autre.

Toutefois, cette analyse reste incomplète puisque nous ne disposons pas de toute l'information (seulement le début du texte). C'est pourquoi, nous avons poursuivi cette étude en ne tenant compte, cette fois-ci, que de l'autre partie du texte, c'est-à-dire les 89% restant.

2. L'influence du co-texte sémantique dans le reste du texte

Parmi les 861 élèves, 193 ont lu le texte en entier. Sachant qu'aucun CP n'a lu tous les mots du texte, cette analyse porte donc sur les élèves les plus avancés en lecture, c'est-à-dire ceux qui ont lu le texte en entier et qui sont scolarisés du CE1 au CM2.

Quels sont les élèves les plus influencés par le co-texte sémantique dans le reste du texte?

Alors que dans le début du texte, aucun élève n'a commis d'erreur de lecture due au co-texte sémantique, on constate que lorsqu'on analyse l'autre partie du texte, une grande majorité des élèves scolarisés entre le CE1 et le CM2 ont commis au moins une erreur de ce type (*cf.* tableau 51).

Tableau 51 : pourcentage et nombre d'élèves ayant été influencés ou non par le co-texte sémantique dans le reste du texte selon la classe

	Aucune erreur	Au moins une erreur	Total	χ^2	<i>p.</i>
CE1	17% (1)	83% (5)	100% (6)	.738	.864
CE2	12% (3)	88% (21)	100% (24)		
CM1	10% (7)	90% (61)	100% (68)		
CM2	8% (8)	92% (87)	100% (95)		
Ensemble	10% (19)	90% (174)	100% (193)		

Lecture : 17% des élèves scolarisés en CE1 (soit 1 sur 5) n'ont commis aucune erreur due au co-texte sémantique dans le reste du texte et 83% en ont commis au moins une (soit 5 sur 6).

On remarque que cette proportion ne varie pas de façon significative avec la classe fréquentée. Ces élèves sont 83% en CE1, 88% en CE2, 90% en CM1 et 92% en CM2.

Ainsi, à un niveau de lecture plus avancé, on remarque une continuité dans la progression des élèves puisque la part de ceux qui ont commis une erreur en lien avec le co-texte sémantique varie peu.

Dans quelle proportion les élèves ont-ils été le plus influencés par le co-texte sémantique ?

Parmi les 193 élèves qui ont lu le texte en entier et qui ont commis au moins une erreur liée à la sémantique, 55% d'entre eux en ont commis une seule erreur (*cf.* tableau 52).

Tableau 52 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par le co-texte sémantique dans le reste du texte selon la classe

	Une seule erreur	Plusieurs erreurs	Total	χ^2	<i>p.</i>
CE1	100% (5)	0% (0)	100% (5)	7.075	.070
CE2	57% (12)	43% (9)	100% (21)		
CM1	61% (37)	39% (24)	100% (61)		
CM2	47% (41)	53% (46)	100% (87)		
Ensemble	55% (95)	45% (79)	100% (174)		

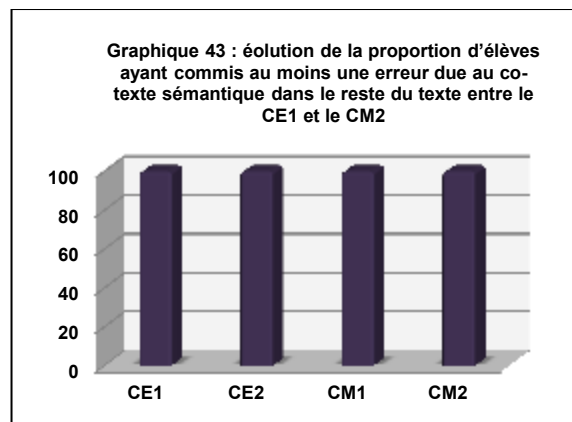
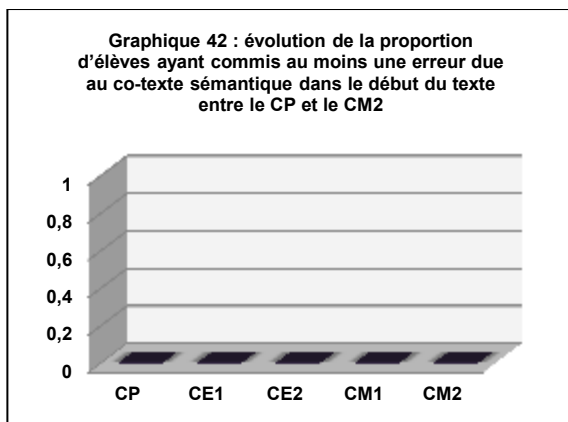
Lecture : 100% des élèves scolarisés CE1 qui ont commis au moins une erreur due au co-texte sémantique, ont commis qu'une seule erreur (soit 5 sur 5).

Bien que l'on observe des proportions différentes d'élèves qui ont commis une ou plusieurs erreurs, il n'existe cependant pas de relation significative entre la classe et le fait de commettre une ou plusieurs erreurs de ce type. Ainsi, malgré le fait qu'en CE1 tous les élèves qui se sont trompés ont commis une seule erreur, qu'en CE2 ils sont 57%, puis 61% en CM1 et 47% en CM2, la relation n'est pas significative. La continuité quantitative observée

précédemment pour le reste du texte semble se confirmer aussi en terme de fréquence d'erreur.

Ainsi, lorsqu'on ne tient compte que du reste du texte, on constate que la grande majorité des élèves entre le CE1 et le CM2 a commis des erreurs en lien avec le co-texte sémantique. On observe donc une certaine continuité puisque les résultats ne varient pas significativement d'une classe à l'autre.

En **résumé**, cette seconde partie nous permet de constater que les résultats diffèrent selon que l'on considère le début ou le reste du texte. En effet, si on ne tient compte que de la partie lue par tous les élèves, il apparaît que personne n'a commis d'erreurs en lien avec le co-texte sémantique dans le début du texte (*cf.* graphique 42). En revanche, lorsqu'on ne tient compte que du reste du texte, on observe un changement : tous les élèves ont commis au moins une erreur de ce type (*cf.* graphique 43).



Il semble donc qu'il existe une **discontinuité quantitative entre les lecteurs débutants et les plus avancés**, plus précisément entre les élèves de CP et les autres. En effet, à un niveau débutant, on constate que la progression est la même puisque la quasi-totalité des élèves scolarisés entre le CP et le CM2 n'a pas commis d'erreurs dues à la sémantique. En revanche, à un niveau plus avancé, on remarque que presque tous les élèves ont commis au moins une erreur de ce type.

Ainsi, contrairement à nos suppositions de départ, il semblerait que les erreurs liées au co-texte sémantique ne soient pas caractéristiques d'une insuffisance dans le traitement alphabétique puisque ce sont les élèves les plus âgés qui commettent ce type d'erreur.

Toutefois, on peut se demander si ces résultats ne pourraient pas s'expliquer aussi par une hétérogénéité du texte, qui ferait que la seconde moitié du texte contiendrait davantage de pièges sémantiques.

Afin de compléter cette analyse, nous avons effectué une analyse qualitative, basée sur les précédentes, pour voir quels sont les types d'erreurs les plus induits.

III. Quels sont les types d'erreurs les plus induits par le co-texte sémantique ?

Rappelons que, pour chaque niveau scolaire, à partir du nombre de mots lus moyen², nous avons regardé quelles sont les erreurs dues au co-texte sémantique les plus fréquemment commises, c'est-à-dire celles commises par plus de 10% des élèves. Toutefois, comme aucun élève n'a commis d'erreur dans le début du texte, cette analyse porte sur l'autre partie du texte et ne concerne pas les élèves de CP puisqu'aucun d'entre eux n'a été aussi loin dans leur lecture.

Cet inventaire, nous a permis de remarquer que la lecture de trois mots semble avoir été particulièrement influencée par le co-texte sémantique (*cf.* tableau 53).

Tableau 53 : mots dont la lecture a été la plus influencée par le co-texte sémantique pour chaque niveau scolaire

	178 poisons	216 décochée	225 cassettes
CE1 (N= 180)	83% (50)	3% (1)	8% (2)
CE2 (N=174)	70% (75)	15% (11)	29% (17)
CM1 (N=188)	86% (141)	19% (26)	17% (21)
CM2 (N=176)	83% (140)	29% (46)	31% (48)
Ensemble (N=861)	81% (408)	21% (84)	24% (88)

Mots lus par moins de la moitié des élèves

Lecture : 70% des élèves de CE2 qui ont lu le mot « poisons » ont commis une erreur due au co-texte sémantique

² Rappel du nombre moyen de mots lus selon la classe : CP : 55 mots, CE1 : 154, CE2 : 197, CM1 : 234 et CM2 : 253.

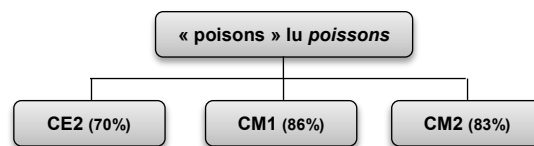
On constate que les erreurs varient d'une classe à l'autre. En effet, alors que pour un mot, la proportion d'élèves qui se sont trompés varie de façon irrégulière (« poisons »), on remarque que pour les deux autres cette proportion semble augmenter d'une classe à l'autre (« décochée » et « cassette »).

Le type d'erreur qui varie de façon irrégulière

Il s'agit du mot « **poisons** » lu [pwasõ].

Un lac étire ses calmes rives et, quand le soir descend,
le miroir de ses eaux reflète les **poisons** des brignoles perfides.
Et, quand descend le soir, quand joue la pourpre du couchant,
le ciel rougit ses eaux.
Dans la moire de l'eau danse l'ombre d'un écueil.
Tout est cris! Tout est bruits!

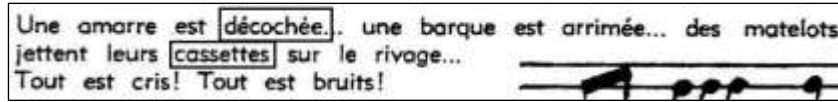
Ainsi, 70% des élèves de CE2 ont commis cette erreur de lecture. En CM1, ils sont encore plus nombreux puisqu'ils sont 86%. En CM2, on voit leur nombre diminuer légèrement et passer à 83%.



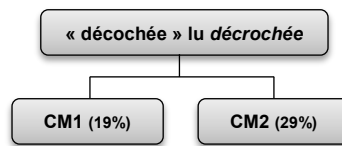
On remarque que ce mot apparaît dans un co-texte sémantique explicite et est, par conséquent, davantage prédictible. En effet, à la lecture de ce paragraphe, on remarque que le champ lexical de l'eau est fortement présent (« lac, rives, eaux, ... »). De plus, le mot produit est un voisin orthographique du mot du texte puisqu'une seule lettre diffère (poisons/poissons). Ainsi, on suppose que cette ressemblance graphique associée au champ lexical de l'eau a amené les élèves à commettre cette erreur. Il est aussi possible que, pour les élèves les plus jeunes, certains aient prononcé [pwasõ] à cause de la non maîtrise de la règle de prononciation du graphème contextuel *s*. En effet, si on regarde du côté des élèves de CE1, on remarque que même si moins de la moitié d'entre eux ont lu ce mot (60 sur 180), ils sont toutefois 83% à avoir commis une erreur et à avoir lu [pwasõ]. Il est donc possible que les causes de cette erreur soient différentes entre les élèves les plus jeunes et les plus âgés.

Les types d'erreurs qui augmentent

On remarque que pour les élèves de CM1 et CM2, la lecture de deux mots semble avoir été particulièrement influencée par le co-texte sémantique. Il s'agit des mots « décochée » et « cassettes » qui se situent dans le paragraphe d'après.

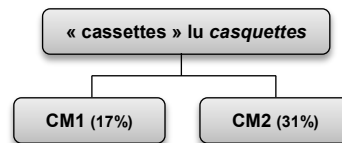


On remarque que le mot « **décochée** » a été lu [dekʁoʃe] par 19% des élèves de CM1 et 29% des CM2.



Ce paragraphe poursuit l'évocation de l'eau du paragraphe précédent et semble se centrer sur l'image d'un bateau (« amarre, barque, matelots »). Ainsi, il se pourrait que les élèves, influencés par le mot « amarre », se soient laissés porter par ces idées et aient donc lu [dekʁoʃe]. Ici aussi, le mot prononcé ne diffère que par une lettre du mot d'origine (décochée/décrochée). Le fait que cette erreur augmente en CM2 nous laisse penser que les élèves plus âgés, qui ont une lecture plus rapide, donc davantage confiants, ne traitent pas systématiquement tous les indices du mot pour l'identifier.

Il en est de même pour le second mot « **cassettes** » lu [kaskɛtə].



Là aussi, portés par le thème du bateau, il se pourrait que les élèves qui ont commis cette erreur n'aient pas fait attention à l'orthographe du mot et l'ont par conséquent associé au mot précédent « matelots ». Comme pour les mots « poisons » et « décochée », là aussi on remarque qu'une seule lettre change entre les deux mots : cassettes / casquettes.

Ainsi, comme pour le mot « poisons », on remarque que les erreurs portant sur les mots « décochée » et « cassettes » correspondent à des mots auxquels on s'attendrait dans ce co-texte. Les élèves semblent donc être tombés dans le piège du co-texte sémantique.

CONCLUSION

Ce chapitre avait pour objectif de voir dans quelle mesure le co-texte sémantique dans le texte de *L'alouette* affecte la lecture des élèves et, de façon plus générale, quelles sont les informations qu'il nous apporte sur l'apprentissage de la lecture. Dans les modèles les plus classiques de l'apprentissage de la lecture, la construction du sens est toujours postérieure à celle du code, ce qui signifie que cet apprentissage se déroulerait de façon discontinue. Ce test de lecture, basé sur une conception classique, mesure essentiellement le décodage et les erreurs qui en découlent sont considérées comme étant une défaillance dans le traitement alphabétique. Partant de là, nous avons supposé, que pour certains mots, le recours au co-texte sémantique serait une façon de compenser cette insuffisance dans le décodage. Par conséquent, nous nous attendions à ce que la maîtrise du code fasse diminuer ce type d'erreur.

Dans une première partie, nous avons constaté une discontinuité entre les élèves scolarisés en cycle 2 et ceux scolarisés en cycle 3. En effet, alors que les plus jeunes ne sont qu'une minorité à avoir commis des erreurs liées au co-texte sémantique, il apparaît que c'est l'inverse en cycle 3 puisque ces derniers sont une majorité à avoir commis ce type d'erreur. Plus précisément, cette proportion augmente fortement entre le CP et le CE1 et le CE2 et le CM1. D'autre part, l'analyse du nombre d'erreurs commises nous permet de constater que non seulement les élèves plus âgés sont plus nombreux à avoir été influencés mais qu'ils ont aussi commis davantage d'erreur de ce type. Ainsi, il apparaît que plus un élève avance dans la maîtrise de la lecture, et plus il semble être influencé par le co-texte sémantique.

Néanmoins, ces résultats sont discutables dans la mesure où les élèves n'ont pas tous lu la même quantité de mots. Ainsi, partant de ce constat, nous avons repris cette analyse en ne tenant compte d'abord que de la partie lue par tous les élèves, puis du reste, partie que seuls les lecteurs plus avancés ont lue.

On s'aperçoit alors que lorsqu'on examine le début du texte les résultats sont différents : le co-texte sémantique ne semble plus affecter les élèves, quel que soit leur niveau de scolarisation. En revanche, à un niveau plus avancé en lecture, on constate que c'est quasiment tous les élèves qui ont été influencés par le co-texte sémantique. Il semble donc qu'il existe une discontinuité entre les élèves de CP et les autres. Le fait que, dans cette analyse, ce soient les lecteurs plus avancés qui commettent le plus ce type d'erreur semble

confirmer le fait que plus un élève avance dans l'apprentissage de la lecture et plus il aurait tendance à être influencé par le co-texte sémantique.

Cependant, la lecture de certains mots du texte peut être influencée par d'autres variables, telles que les dessins présents autour du texte. C'est pourquoi, pour plus de rigueur, nous avons effectué une troisième analyse, sur les mêmes principes que les deux premières, mais sans tenir compte des mots dont la source d'effet peut être multiple.

Les résultats obtenus nous permettent de constater que les tendances ne changent pas : ce sont les élèves les plus âgés qui semblent être le plus influencés. En effet, on observe toujours une discontinuité entre les élèves en cycle 2 et cycle 3 et, dans le détail, on retrouve aussi les ruptures entre le CP et le CE1 et le CE2 et CM1. On remarque par ailleurs, qu'à un niveau plus avancé, les résultats ne changent pas puisque la plupart des élèves scolarisés entre le CE1 et le CM2 ont commis au moins une erreur due à la sémantique.

Par conséquent, le recours au co-texte sémantique ne serait pas seulement le résultat d'une défaillance dans le traitement alphabétique puisque ce sont les lecteurs plus âgés qui semblent être le plus influencés. Il se pourrait donc que cette discontinuité entre les lecteurs débutants et les lecteurs plus experts soit le reflet d'un changement de stratégie dans l'apprentissage de la lecture.

Partant de ce constat, nous avons réalisé une quatrième analyse avec une approche plus qualitative. A partir des types d'erreurs répertoriés, nous avons cherché à identifier des stratégies pour chaque niveau scolaire.

Cette analyse, nous permet de voir que les types d'erreurs sont différents d'un moment à l'autre du cursus scolaire. En effet, alors que les lecteurs débutants commettent plutôt des erreurs de type lexical (« Annie » lu *animaux*), ceux qui sont plus avancés font des erreurs grapho-sémantiques « poisons » lu *poissons*, « cassettes » lu *casquettes*, ...).

Ainsi, ces différentes analyses nous permettent de constater que, dès les premières années de l'apprentissage de la lecture, les élèves essaient de donner du sens à leur lecture, que ce soit d'abord au niveau du mot, puis à celui de la phrase. En réalité, plus un élève avance dans l'apprentissage de la lecture et plus sa recherche de sens s'élargit. Lire ne serait donc pas que décoder ; or, le test de *L'alouette* mesure les capacités de décodage. Par conséquent, l'élève qui tombe dans les pièges sémantiques tendus par Lefavrais serait un

mauvais lecteur. Or, les différentes analyses menées dans ce chapitre nous montrent que ce sont les lecteurs plus avancés qui tombent dans ces pièges. Ainsi, les mauvais lecteurs pour Lefavrais seraient en réalité les bons lecteurs, ceux qui cherchent à donner du sens à leur lecture.

Toutefois, comme nous l'avons dit précédemment, nous ne pouvons pas écarter l'idée d'un effet d'autres variables. En effet, pour certains de ces mots d'autres caractéristiques peuvent aussi expliquer ces erreurs.

Ex. « **pompe** » lu *pomme* → effet du mot « verger » ou effet du dessin du fruit ?

Ex. « **poisons** » lu *poissons* → effet de la sémantique ou confusion de son due au graphème positionnel « s » ?

A partir de ces résultats on peut faire l'hypothèse que les élèves n'ont peut-être pas recours au co-texte sémantique pour les mêmes raisons. En effet, il se pourrait que cette stratégie fasse partie d'une recherche de sens lexical pour les plus jeunes alors que pour les lecteurs plus confirmés ce serait davantage une recherche de sens phrastique. Ce constat semble rejoindre celui de Stanovich qui postule que l'effet du contexte varie selon le niveau de lecture des élèves. En effet, selon ce chercheur, le contexte aurait un effet lors de l'identification du mot pour les plus jeunes (et donc les moins habiles en lecture) alors que pour les plus âgés il jouerait un rôle plutôt dans la compréhension, au sens plus large. Reste à vérifier par la mise en perspective des chapitres suivants. Ce chapitre nous permet seulement de conclure que **lire, ce n'est pas seulement décoder** et que **dès les premières années de l'apprentissage, les élèves cherchent à donner du sens à leur lecture**. L'analyse des autres types d'erreurs lors de la lecture du texte de *L'alouette* va nous permettre de poursuivre la réflexion quant au déroulement de l'apprentissage de la lecture.

Chapitre 14

L'INFLUENCE DES MOTS SURDETERMINES LORS DE LA LECTURE DU TEXTE DE *L'ALOUETTE*

Plan du chapitre

INTRODUCTION	304
PREDICTIONS	304
METHODOLOGIE D'ANALYSE	305
I.L'influence des mots surdéterminés lors de la lecture du test de <i>L'alouette</i>	307
<i>Quels sont les élèves les plus influencés par les mots surdéterminés?</i>	<i>307</i>
<i>Dans quelle proportion ont-ils été influencés?.....</i>	<i>307</i>
1. L'influence des mots surdéterminés en cycle 2	308
<i>Quels sont les élèves les plus influencés par les mots surdéterminés en cycle 2?.....</i>	<i>308</i>
<i>Dans quelle proportion ont-ils été influencés?</i>	<i>309</i>
2. L'influence des mots surdéterminés en cycle 3.....	309
<i>Quels sont les élèves les plus influencés par les mots surdéterminés en cycle 3?.....</i>	<i>309</i>
<i>Dans quelle proportion ont-ils été influencés?.....</i>	<i>310</i>
II. L'influence des mots surdéterminés dans le reste du texte	312
<i>Quels sont les élèves les plus influencés par les mots surdéterminés dans le reste du texte?.....</i>	<i>312</i>
<i>Dans quelle proportion les élèves ont-ils été influencés?.....</i>	<i>312</i>
III.Quels sont les types d'erreurs les plus induits par les mots surdéterminés	314
<i>Le type d'erreur des lecteurs intermédiaires</i>	<i>315</i>
<i>Les types d'erreurs des lecteurs plus avancés</i>	<i>316</i>
CONCLUSION	317

INTRODUCTION

Comme nous venons de le voir dans les chapitres précédents, le test de *L'alouette* renferme un certain nombre de spécificités et la lecture de certains mots peut donc être influencée soit par les dessins, les similitudes ou encore le co-texte sémantique. Toutefois, il apparaît que pour certains mots, la source d'influence peut être multiple c'est-à-dire que la lecture de certains mots peut être influencée non seulement par les dessins mais aussi par le co-texte sémantique et/ou encore les similitudes. On se demande alors si cette surdétermination ne constitue pas une difficulté supplémentaire lors de la lecture de certains mots.

Les chapitres précédents avaient pour objectif de voir dans quelle mesure les variables inhérentes au test de Lefavrais ont une influence lors de la lecture du texte de *L'alouette* et plus précisément, de quelle façon les erreurs commises nous renseignent sur l'apprentissage de la lecture. Les résultats font apparaître que, quel que soit le type d'influence, ce sont les lecteurs plus âgés qui sont le plus influencés.

PREDICTIONS

Rappelons que le test de *L'alouette* a été conçu pour mesurer essentiellement le niveau de décodage et toute erreur commise lors de la lecture de ce dernier traduit alors une défaillance dans les capacités à décoder. Nous supposons alors que les élèves auront recours à différents indices pour pallier cette défaillance.

Compte tenu des résultats obtenus dans les précédentes analyses, nous nous attendons à ce que les mots surdéterminés entraînent aussi des difficultés de lecture. Plus précisément, nous nous attendons à observer, ici aussi, une discontinuité quantitative entre les élèves en cycle 2 et cycle 3. D'un point de vue qualitatif, nous nous attendons à observer là aussi une discontinuité se traduisant par un type d'erreur différent entre les lecteurs débutants et plus avancés : des erreurs lexicales pour les plus jeunes et des erreurs grapho-sémantiques pour les plus âgés.


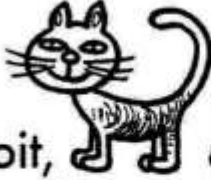

METHODOLOGIE D'ANALYSE


Ainsi, nous avons répertorié cinq mots dont la prononciation peut être le résultat de plusieurs influences :


Tableau 54 : mots dont la lecture peut être influencée par plusieurs variables du test de *L'alouette*





Mots	Types d'erreurs	Types d'influences
« alouette fait un nœud »	<i>Alouette fait un nid</i> (4) <i>Alouette fait un œuf</i> (61)	Dessin ou co-texte sémantique
« au clair de lune »	<i>Au clair de la lune</i> (282)	Dessins, co-texte sémantique ou similitudes
« mon amie annie »	<i>Mon ami Pierrot</i> (4)	Dessins, co-texte sémantique ou similitudes
« prête-moi la plume »	<i>Prête-moi ta plume</i> (146)	Dessins, co-texte sémantique
« prête-moi la plume »	<i>Prête-moi la lune</i> (1)	Dessins, co-texte sémantique ou similitudes



Cartographie 8 : mots surdéterminés dans le test de L'alouette


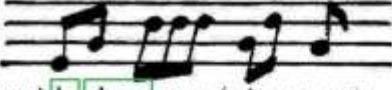
 L'alouette.  

Sous la mousse ou sur le toit,
dans les haies vives ou le chêne fourchu,
le printemps a mis ses nids.
Le printemps a nids au bois. 

Annie amie, du renouveau, c'est le doux temps.
Amie Annie, au bois joli gamine le pinson.
Dans les buis, gîte une biche, au bois chantant.
Annie, Annie! au doigt joli, une églantine laisse du sang :
au bout du temps des féeries viendra l'ennui. 

 L'alouette fait ses jeux; alouette fait un nœud avec un rien de paille.
L'hirondeau piaille sous la pente des bardeaux et, vif et gai, le geai
sur l'écaille argentée du bouleau, promène un brin d'osier.
 Au verger, dans le soleil matinal, goutte une pompe dégelée.
On voit un bec luisant qui trille éperdument des notes claires
et, dans les pampres d'or que suspend la grille antique,
on surprend des rixes de moineaux.
Au potager s'alignent les cordeaux; l'if est triste à l'horizon
et lourd et lent l'envol des corbeaux.  

 Un lac étire ses calmes rives et, quand le soir descend,
le miroir de ses eaux reflète les poisons des brignoles perfides.
Et, quand descend le soir, quand joue la pourpre du couchant,
le ciel rougit ses eaux.
Dans la moire de l'eau danse l'ombre d'un écueil.
Tout est cris! Tout est bruits! 

 Une amarre est décochée... une barque est arrimée... des matelots
jettent leurs cassettes sur le rivage...
Tout est cris! Tout est bruits!
Au clair de la lune mon ami Pierrot... 
Au clair de lune mon amie annie..
Au clair de la lune mon ami Pierrot, prête-moi la plume pour écrire un mot.

o u e i a
le la les un dans des do ti pu mi

I. L'influence des mots surdéterminés lors de la lecture du test de L'alouette

Quels sont les élèves les plus influencés par les mots surdéterminés?

Dans l'ensemble, il apparaît que 59% des élèves n'ont pas commis d'erreur lors de la lecture de ces mots (cf. tableau 55).

Tableau 55 : pourcentage et nombre d'élèves ayant commis ou non des erreurs dues aux mots surdéterminés selon le cycle de scolarisation

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
Cycle 2	83% (269)	17% (54)	100% (323)	127.675	.000
Cycle 3	44% (236)	56% (302)	100% (538)		
Ensemble	59% (505)	41% (356)	100% (861)		

Lecture : 83% des élèves scolarisés en cycle 2 (soit 269 sur 323) n'ont commis aucune erreur due aux mots surdéterminés et 17% en ont commis au moins une (soit 54 sur 323).

Dans le détail, on remarque que, alors que les élèves en cycle 2 sont une majorité à ne pas avoir commis d'erreur (83%), ils sont une majorité en cycle 3 à avoir été influencés au moins une fois (56%). Il apparaît donc, que comme pour les variables précédentes, ce sont les lecteurs plus âgés qui sont le plus influencés. On retrouve donc ici une discontinuité entre la progression des élèves entre le cycle 2 et cycle 3 que nous avons constaté dans les précédents chapitres.

Dans quelle proportion ont-ils été influencés ?

Sur les 861 élèves qui composent notre échantillon, 356 ont été influencés au moins une fois, soit 41% d'entre eux. On remarque que, parmi eux, la plupart des élèves n'ont commis qu'une seule erreur de ce type (61%) (cf. tableau 56).

Tableau 56 : pourcentage et nombre d'élèves ayant commis une ou plusieurs erreur(s) due(s) aux mots surdéterminés selon le cycle de scolarisation

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
Cycle 2	91% (49)	9% (5)	100% (54)	22.277	.000
Cycle 3	56% (168)	44% (134)	100% (302)		
Ensemble	61% (217)	39% (139)	100% (356)		

Lecture : 91% des élèves scolarisés en cycle 2 qui ont commis au moins une erreur due aux mots surdéterminés (soit 49 sur 54) n'ont commis qu'une seule erreur de ce type et 9% en ont commis plusieurs (soit 5 sur 54).

En effet, que ce soit en cycle 2 ou cycle 3, les élèves à avoir été influencés ont essentiellement commis une seule erreur de ce type. Ils sont 91% en cycle 2 et 56% en cycle 3. Toutefois, si on regarde les proportions d'élèves qui ont commis plusieurs erreurs, alors on constate que ce sont les élèves en cycle 3 qui sont les plus nombreux dans cette catégorie.

Ainsi, bien que les élèves en cycle 3 soient plus nombreux à avoir été influencés, ils commettent essentiellement une seule erreur. Néanmoins, quand ils se trompent, ce sont eux qui en commettent le plus. On constate donc une certaine continuité en ce qui concerne le nombre d'erreurs commises.

1. L'influence des mots surdéterminés en cycle 2

Rappelons que parmi les 861 élèves de notre échantillon, 323 sont scolarisés en cycle 2 dont 143 en CP et 180 en CE1.

Quels sont les élèves les plus influencés par les mots surdéterminés en cycle 2?

Lorsqu'on regarde la tendance dans les classes qui composent le cycle 2, on remarque qu'elle est la même que celle qui concerne l'ensemble des élèves de ce cycle : la majorité (83%) des élèves en cycle 2 n'ont pas été influencés (cf. tableau 57).

Tableau 57 : pourcentage et nombre d'élèves ayant été influencés ou non par les mots surdéterminés selon la classe en cycle 2

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CP	94% (135)	6% (8)	100% (143)		
CE1	74% (134)	26% (46)	100% (180)	21.394	.000
Ensemble	83% (269)	17% (54)	100% (323)		

Lecture : 94% des élèves scolarisés en CP (soit 135 sur 143) n'ont commis aucune erreur due aux mots surdéterminés et 6% en ont commis au moins une (soit 8 sur 143).

En effet, on compte 94% des élèves de CP et 74% des élèves de CE1 qui n'ont pas été influencés. Toutefois, on remarque que les élèves qui ont commis au moins une fois ce type d'erreur sont plus nombreux en CE1 (26%) qu'en CP (6%).

Ainsi, il apparaît que les élèves en cycle 2 ne sont pas majoritairement influencés mais quand ils le sont, ce sont les élèves plus âgés qui sont les plus nombreux.

Dans quelle proportion ont-ils été influencés?

Parmi les 323 élèves scolarisés en cycle 2, rappelons que 54 d'entre eux a commis au moins une fois ce type d'erreur. On note que, les élèves qui se sont trompés ont tendance à ne commettre qu'une seule erreur de ce type (91% des élèves) (cf. tableau 58).

Tableau 58 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par les mots surdéterminés selon la classe en cycle 2

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CP	87% (7)	13% (1)	100% (8)		
CE1	91% (42)	9% (4)	100% (46)	.000	1.000
Ensemble	91% (49)	9% (5)	100% (54)		

Lecture : 87% des élèves scolarisés en CP qui ont commis au moins une erreur due aux mots surdéterminés (soit 7 sur 8) n'ont commis qu'une seule erreur de ce type et 13% en ont commis plusieurs (soit 1 sur 8).

Dans le détail, 87% des élèves de CP et 91% des élèves de CE1 n'ont commis qu'une seule fois ce type d'erreur. On observe donc une continuité en ce qui concerne le nombre d'erreurs commises : que les élèves soient scolarisés en CP ou en CE1, ils ne commettent qu'une seule erreur de ce type.

Voyons maintenant ce qu'il en est pour les élèves des classes qui composent le cycle 3.

2. L'influence des mots surdéterminés en cycle 3

Les élèves en cycle 3 sont 538 dont 174 en CE2, 188 en CM1 et 176 en CM2.

Quels sont les élèves les plus influencés par les mots surdéterminés en cycle 3?

Rappelons que les élèves en cycle 3, contrairement à ceux en cycle 2, sont une majorité (56%) à avoir commis au moins une erreur (cf. tableau 59).

Tableau 59 : pourcentage et nombre d'élèves ayant été influencés ou non par les mots surdéterminés selon la classe en cycle 3

	Aucune erreur	Au moins une erreur	Total	χ^2	p.
CE2	65% (114)	35% (60)	100% (174)		
CM1	46% (86)	54% (102)	100% (188)	72.570	.000
CM2	20% (36)	80% (140)	100% (176)		
Ensemble	44% (236)	56% (302)	100% (538)		

Lecture : 65% des élèves scolarisés en CE2 (soit 114 sur 174) n'ont commis aucune erreur due aux mots surdéterminés et 35% en ont commis au moins une (soit 60 sur 174).

Dans le détail, on constate que la tendance générale ne se retrouve pas selon la classe considérée. Ainsi, on observe que les élèves de CE2, comme les CP et CE1, sont une majorité à n'avoir commis aucune erreur (65%). A partir du CM1, les élèves qui ont commis au moins une fois ce type d'erreur deviennent majoritaires : ils sont 54% en CM1 et 80% en CM2. Par conséquent, il semblerait que la discontinuité observée entre le cycle 2 et cycle 3 se situe plus précisément entre les élèves de CE2 et de CM1.

Dans quelle proportion ont-ils été influencés?

Rappelons que parmi les 538 élèves scolarisés en cycle 3, 302 ont commis au moins une fois ce type d'erreur (cf. tableau 60).

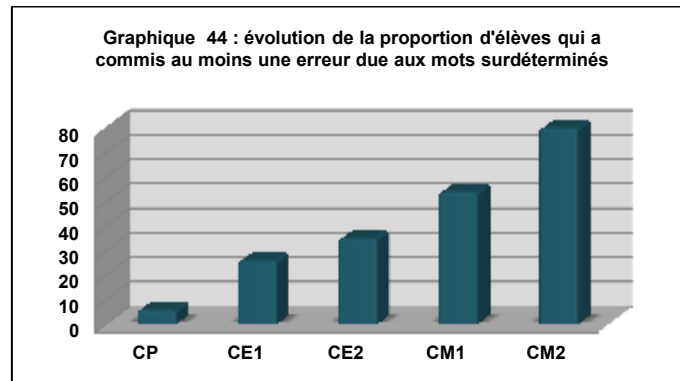
Tableau 60 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par les mots surdéterminés selon la classe en cycle 3

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CE2	67% (40)	33% (20)	100% (60)	7.012	.030
CM1	46% (47)	54% (55)	100% (102)		
CM2	58% (81)	42% (59)	100% (140)		
Ensemble	56% (168)	44% (134)	100% (302)		

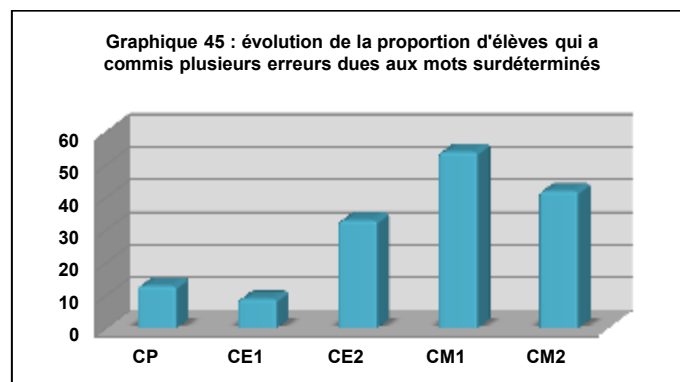
Lecture : 67% des élèves scolarisés en CE2 qui ont commis au moins une erreur due aux mots surdéterminés (soit 40 sur 60) n'ont commis qu'une seule erreur de ce type et 33% en ont commis plusieurs (soit 20 sur 60).

Si on regarde les résultats par classe, on constate que les élèves à n'avoir commis qu'une seule erreur de ce type sont une majorité en CE2 (67%) et CM2 (58%), alors qu'en CM1, les élèves ont tendance à commettre plusieurs erreurs de ce type (54%).

En **résumé**, cette première analyse nous permet de constater que les élèves n'ont pas tous été influencés de la même façon selon leur niveau de scolarisation. En effet, alors qu'en cycle 2, les élèves sont une majorité à ne pas avoir commis d'erreur due à la surdétermination de certains mots, c'est l'inverse pour les élèves en cycle 3 qui sont une majorité à avoir commis au moins une erreur de ce type. Ainsi, on constate une rupture dans la progression des élèves qui serait alors le signe d'une discontinuité dans l'apprentissage de la lecture. Il y aurait donc une discontinuité entre le cycle 2 et le cycle 3. Si on regarde les résultats par classe, on constate plus précisément que la discontinuité apparaît entre les élèves de CP, CE1 et CE2 et les élèves de CM1 et CM2 : à partir du CM1, les élèves qui ont commis ce type d'erreur sont les plus nombreux.



D'autre part, lorsqu'on regarde le nombre d'élèves qui a commis plusieurs fois ce type d'erreur, on constate que ce n'est qu'en CM1 que les élèves ont tendance à être influencés plusieurs fois.



Il apparaît donc que, d'une façon générale, plus un élève avance dans l'apprentissage de la lecture et plus il a tendance à être influencé. Ces premiers résultats semblent donc rejoindre ceux trouvés précédemment lors des analyses de chaque variable inhérente au test de Lefavrais.

Toutefois, il convient de rappeler que tous les élèves n'ont pas lu le même nombre de mots, ce qui biaise notre analyse.

Pendant, il apparaît que les mots dont la lecture peut être influencée par différents indices ne se situent qu'à un niveau avancé du texte. Par conséquent, si on ne tient compte que de la partie lue par tous les élèves, c'est-à-dire le début du texte, alors aucun élève n'a commis d'erreur. C'est pourquoi, nous nous proposons maintenant de regarder les résultats pour le reste du texte seulement.

II. L'influence des mots surdéterminés dans le reste du texte

Rappelons qu'aucun élève de CP n'a lu le texte en entier. Par conséquent, cette analyse porte sur les élèves scolarisés du CE1 au CM2 qui sont au total 193 à avoir lu tout le texte.

Quels sont les élèves les plus influencés par les mots surdéterminés dans le reste du texte?

Parmi les 193 élèves qui ont lu le texte en entier, 96% d'entre eux (soit 185) a commis au moins une erreur de ce type dans le reste du texte (cf. tableau 61).

Tableau 61 : pourcentage et nombre d'élèves ayant été influencés ou non par les mots surdéterminés dans le reste du texte selon la classe

	Aucune erreur	Au moins une erreur	Total	χ^2	<i>p.</i>
CE1	0% (0)	100% (6)	100% (6)	5,483	,140
CE2	12% (3)	88% (21)	100% (24)		
CM1	4% (3)	96% (65)	100% (68)		
CM2	2% (2)	98% (93)	100% (95)		
Ensemble	4% (8)	96% (185)	100% (193)		

Lecture : Aucun élève de CE1 n'a commis d'erreur due aux mots surdéterminés dans le reste du texte

Dans le détail, on remarque que la tendance est la même quelle que soit la classe fréquentée. Ainsi, tous les élèves de CE1 se sont trompés au moins une fois. En CE2, ils sont 88%. Au niveau suivant, en CM1, cette catégorie d'élèves représente 96% puis 98% en CM2.

Il semble donc qu'à un niveau plus avancé de lecture, les élèves ont tendance à être davantage influencés.

Dans quelle proportion ont-ils été influencés ?

Lorsqu'on regarde le nombre d'élèves qui se sont trompés plusieurs fois, on constate que dans l'ensemble, les élèves ont tendance à ne commettre qu'une seule erreur de ce type (55%) (cf. tableau 62).

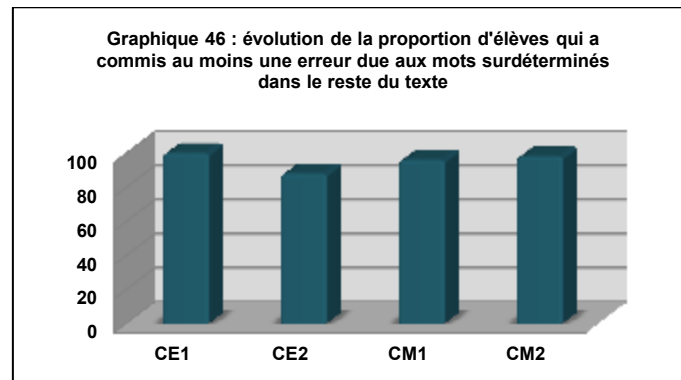
Tableau 62 : pourcentage et nombre d'élèves ayant été influencés une ou plusieurs fois par les mots surdéterminés dans le reste du texte selon la classe

	Une seule erreur	Plusieurs erreurs	Total	χ^2	p.
CE1	100% (6)	0% (0)	100% (6)	8,418	,038
CE2	52% (11)	48% (10)	100% (21)		
CM1	45% (29)	55% (36)	100% (65)		
CM2	59% (55)	41% (38)	100% (93)		
Ensemble	55% (101)	45% (84)	100% (185)		

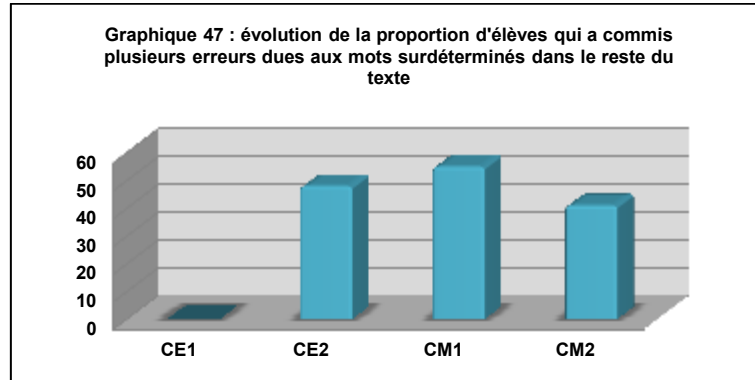
Lecture : Tous les élèves scolarisés en CE1 qui ont commis au moins une erreur due aux mots surdéterminés (soit 6) n'ont commis qu'une seule erreur de ce type.

Plus précisément, on retrouve les résultats de l'analyse précédente, à savoir que mis à part les élèves de CM1, ceux scolarisés en CE1, CE2 et CM2 ont majoritairement commis qu'une seule erreur.

En **résumé**, cette analyse nous permet de constater, qu'à un niveau plus avancé, il existe une continuité quantitative entre le CE1 et le CM2 puisque la majorité des élèves qui sont scolarisés dans ces classes a commis au moins une fois ce type d'erreur.



En revanche, lorsqu'on regarde le nombre d'erreurs commises, on constate que, alors que les élèves de CE1, CE2 et CM2 ont tendance à ne commettre qu'une seule erreur, les élèves de CM1 ont tendance à en commettre plusieurs. Ainsi, bien que les élèves commettent essentiellement des erreurs, certains en font plus que d'autres.




III. Quels sont les types d'erreurs les plus induits par les mots surdéterminés ?

Comme pour les précédents chapitres, nous avons regardé pour chaque niveau scolaire, à partir du nombre de mots lus moyen¹, quelles sont les erreurs dues à plusieurs influences du test de Lefavrais (celles commises par au moins 10% des élèves). Nous avons vu qu'aucun élève n'a commis d'erreur dans le début du texte, de ce fait cette analyse porte seulement sur la deuxième partie du texte et ne concerne que les élèves à partir du CE1. En effet, les élèves de CP sont peu nombreux à avoir été jusque-là dans leur lecture, par conséquent nous n'en tiendront pas compte.

Ainsi, la lecture de trois mots ou syntagme semble avoir posé des difficultés de lecture à la plupart des élèves du CE1 au CM2 (cf. tableau 63).

Tableau 63 : mots dont la lecture a été la plus influencée par les mots surdéterminés pour chaque niveau scolaire

N° mots	78	246	260
Mots du texte	« Alouette fait un nœud »	« Au clair de lune »	« Prête-moi la plume »
CP (N= 143)	20% (7)	100% (1)	100% (1)
CE1 (N= 180)	19% (30)	88% (15)	38% (5)
CE2 (N= 174)	10% (18)	88% (42)	49% (20)
CM1 (N= 188)	4% (8)	91% (95)	54% (54)
CM2 (N= 176)	1% (2)	91% (129)	48% (46)
Ensemble (N= 861)	9% (65)	90% (282)	50% (146)

 Mots lus par moins de la moitié des élèves de la classe

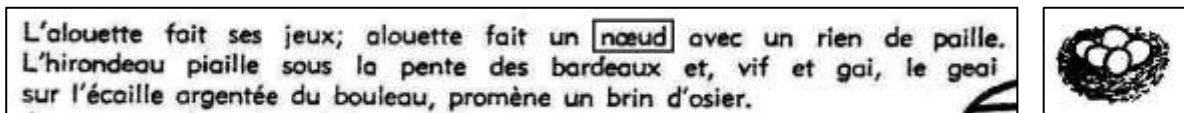
Lecture : En CE1, parmi les élèves qui ont lu le mot « nœud », 19% (soit 30 d'entre eux) ont commis une erreur liée au test de L'alouette.

¹ Rappel du nombre moyen de mots lus selon la classe : CP : 59 mots, CE1 : 154, CE2 : 197, CM1 : 234 et CM2 : 253.

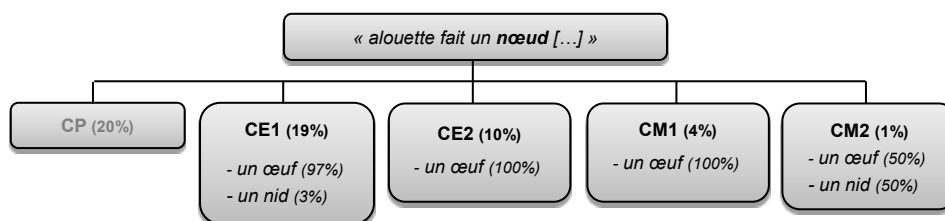
On remarque que les types d'erreurs ne sont pas les mêmes selon la classe que les élèves fréquentent. En effet, en premier lieu, il apparaît que les élèves de CP sont très peu nombreux à avoir été influencés, certainement à cause du fait que ces mots se situent trop loin dans le texte. Ensuite, il apparaît que les élèves de CE1 et de CE2 semblent avoir été influencés sur les mêmes mots et les élèves de CM1 et CM2 sur d'autres. Ainsi, nous allons d'abord voir quelle est l'erreur commune aux lecteurs intermédiaires puis quelles sont les erreurs communes aux lecteurs plus avancés.

Le type d'erreur des lecteurs intermédiaires

On remarque que pour les lecteurs intermédiaires, la lecture d'un mot semble avoir été particulièrement influencée par plusieurs variables liées au test de *L'alouette*. Il s'agit du mot « nœud » dans le syntagme « alouette fait un **nœud** avec un rien de paille »



En effet, 19% des élèves de CE1 ont commis une erreur de lecture. En CE2, leur nombre diminue et ces élèves représentent 10%. Au niveau suivant, ils sont encore moins nombreux puisqu'ils ne sont plus que 4% et seulement 1% en CM2 à avoir commis une erreur lors de la lecture de ce mot.



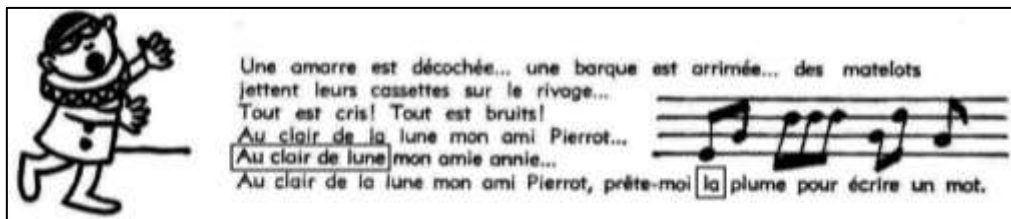
Si on regarde de plus près les types d'erreurs commis à chaque niveau scolaire, on remarque que ce mot a été essentiellement lu [œf] plus précisément [œ nœf]. Ainsi, en CE1, 97% des élèves qui se sont trompés l'ont prononcé comme tel. En CE2, ce sont tous les élèves qui l'ont lu de cette façon, tout comme en CM1. En CM2, sur les deux élèves qui se sont trompés, un a prononcé [œ nœf]. Pour cette erreur, il se pourrait que la ligature située au milieu du mot « -œu » et que le dessin d'un nid rempli d'œufs situé au-dessus les ait amenés à effectuer cette lecture.

Par ailleurs, certains élèves ont prononcé [ni]. Ici, il se pourrait que ce soit le début et la fin du mot (« n » et « d ») qui aient amené les élèves à cette prononciation. De plus, ce mot

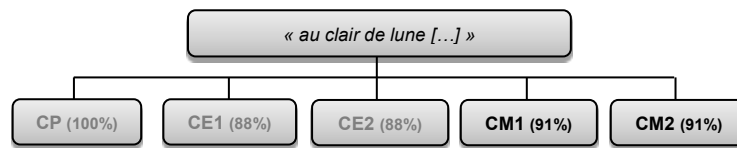
apparaît dans un co-texte sémantique tel, qu'on s'attend à trouver davantage le mot « nid » que le mot « nœud ».

Les types d'erreurs des lecteurs plus avancés

Du côté des lecteurs plus avancés, on constate que la lecture d'un syntagme et d'un mot semble avoir été particulièrement influencée par le test de *L'alouette*. Il s'agit du syntagme « au clair de lune » et du mot « la » dans le groupe de mots « prête-moi **la** plume ». On remarque que ces deux syntagmes ressemblent fortement à la chanson « Au clair de la lune ».

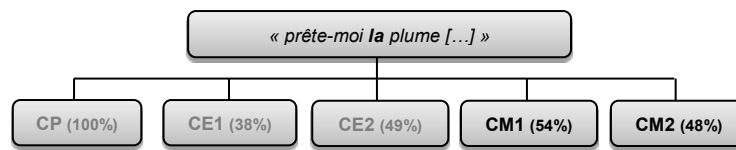


En effet, on remarque 91% des élèves de CM1 et de CM2 ont commis une erreur lors de la lecture du syntagme « au clair de lune » : tous ont rajouté le mot « la » et ont donc lu « au clair de la lune ».



Lorsqu'on regarde le texte de plus près, on s'aperçoit que la phrase qui précède ce syntagme, tout comme celle qui suit, est tirée du refrain de la chanson « au clair de la lune ». Par ailleurs, ces trois lignes sont entourées de dessins : à gauche celui d'un Pierrot et à droite celui d'une partition de notes de musiques. La présence de ces dessins invite et conforte le lecteur à penser à la chanson « au clair de la lune ». On remarque que ce syntagme apparaît dans un contexte fortement influencé par la présence de dessins, de similitudes des débuts de phrases et du co-texte sémantique. C'est le seul dont les sources d'influence sont aussi nombreuses et le fait que les élèves à s'être trompés soient aussi nombreux expliquerait donc un fort effet du test.

Le second type d'erreur des lecteurs plus avancés concerne le mot « la » du syntagme « prête-moi **la** plume ». On constate que 54% des élèves de CM1 et 48% des élèves de CM2 l'ont lu de cette façon.



Ici aussi, le contexte a certainement influencé les élèves mais de manière plus modeste.

En **résumé**, on constate que les mots apparaissant dans un contexte fortement « trompeur » sont ceux sur lesquels le plus grand nombre d'élèves a commis une erreur. De plus, il s'avère que ces élèves ne sont pas les plus jeunes mais plutôt les plus âgés, et par conséquent les plus avancés en lecture. Ce constat semble donc confirmer le fait que le test de *L'alouette* ne mesure simplement pas le décodage et que le fait de tomber dans les pièges de Lefavrais ne traduit pas seulement une défaillance à ce niveau.

CONCLUSION

Ce chapitre avait pour objectif de voir de quelle façon les mots dont la lecture peut être influencée de plusieurs manières nous renseignent sur l'apprentissage de la lecture.

Le test de *L'alouette* contient un certain nombre de pièges dont nous avons fait état dans les chapitres précédents. Les différents résultats obtenus montrent que ce sont les lecteurs plus avancés qui sont le plus influencés par ces différents pièges (iconiques, similitudes et co-texte sémantiques). Ainsi, en toute logique, nous avons supposé que les mots surdéterminés entraîneraient eux aussi des difficultés de lecture aux élèves plus âgés.

En effet, notre première analyse confirme notre supposition : les élèves en cycle 3 sont plus nombreux à avoir commis des erreurs de ce type en comparaison avec ceux du cycle 2. On retrouve donc bien la rupture constatée dans les précédents chapitres entre ces deux moments du cursus scolaire. Cette rupture confirme alors une discontinuité entre les élèves. Cette discontinuité ne se retrouve pas en ce qui concerne le nombre d'erreurs commises puisque, que les élèves soient en cycle 2 ou cycle 3, ils ont majoritairement tendance à ne commettre qu'une seule erreur de ce type.

Par ailleurs, comme pour les précédents chapitres, nous devons rappeler que les élèves n'ont pas tous lu le même nombre de mots. Par conséquent, pour plus de rigueur, nous avons

voulu reprendre cette analyse mais en ne tenant compte que de la partie lue par tous les élèves, c'est à dire le début du texte. Toutefois, nous nous sommes rendu compte que les mots dont la lecture peut être influencée par plusieurs variables du test ne se situent que dans l'autre partie du texte. Par conséquent, nous avons effectué cette analyse seulement pour le reste du texte. Rappelons qu'aucun élève de CP n'a lu tout le texte. Ainsi, cette étude porte sur les élèves scolarisés du CE1 au CM2.

Les résultats font apparaître que quasiment la totalité des élèves qui ont lu tout le texte ont commis au moins une fois ce type d'erreur. Ce constat semble confirmer le fait que ce sont les élèves les plus avancés en lecture qui sont le plus influencés par les variables inhérentes au test de Lefavrais.

Par ailleurs, l'analyse qualitative menée sur les types d'erreurs fait ressortir que ces derniers diffèrent selon le niveau d'avancement des élèves en lecture. Ainsi, les lecteurs intermédiaires (CE1 et CE2) se sont trompés lors de la lecture d'un syntagme (« alouette fait un nœud ») alors que les lecteurs plus avancés ont commis des erreurs en lien avec la chanson « au clair de la lune ». On remarque que ce passage du texte est l'endroit le plus fortement influencé : présence d'un dessin d'un Pierrot et de notes de musiques, présence de similitudes et co-texte sémantique très trompeur. Paradoxalement, c'est le morceau du texte le plus « piégé » et ce sont les plus âgés qui se trompent. Ce constat renforce et confirme ceux des précédents chapitres à savoir que le recours à ces indices ne serait pas seulement une stratégie compensatoire des apprentis lecteurs.

L'influence des variables liées au test de l'alouette

CONCLUSION

Dans la plupart des modèles reconnus de l'acquisition de la lecture, on considère que l'enfant traverse une série de stades dont l'ordre de succession est strict et identique pour tous. Ces stades représentent une progression reposant sur la maîtrise des différents aspects de la langue. Cette progression est alors envisagée de manière discontinue puisque l'enfant passerait par trois ou quatre stades. Par exemple, dans le modèle de Frith, l'enfant passerait d'abord par une étape logographique, puis alphabétique et orthographique. Selon cette conception, l'enfant se centrerait d'abord sur le code avant de construire du sens.

Le test de *L'alouette* est spécifique dans la mesure où Lefavrais a introduit de faux indices parmi les indices pertinents. Ainsi, certains dessins présents autour du texte ne sont pas en rapport avec ce dernier, on trouve aussi des mots, des syntagmes, voire des phrases qui se ressemblent fortement et certains mots apparaissent dans un co-texte sémantique piège, c'est-à-dire qu'à certains endroits du texte, le mot à reconnaître n'est pas celui auquel on s'attend. Compte tenu de toutes ces caractéristiques nous nous sommes donc demandée de quelle façon ce test pouvait nous renseigner sur l'apprentissage de la lecture. Le texte de *L'alouette* est particulièrement représentatif puisqu'il a pour objectif de mesurer essentiellement les capacités de décodage d'un enfant. Par conséquent, les erreurs commises lors de la lecture de ce texte sont considérées comme une défaillance dans cette habileté. Nous avons supposé que les élèves auraient recours à ces différents indices pour pallier cette défaillance. De cette façon, nous nous attendions à ce que les erreurs liées aux différentes variables inhérentes au test de *L'alouette* diminuent dans le temps.

Les résultats des trois études (sur l'influence des dessins, des similitudes et du co-texte sémantique) font apparaître que la proportion d'élèves à avoir commis au moins une erreur en lien avec le test est moins importante en cycle 2 qu'en cycle 3. Autrement dit, ce ne sont pas les lecteurs débutants qui sont le plus influencés par ces variables mais les lecteurs plus avancés. Le recours aux indices suggérés par Lefavrais n'est donc pas seulement une caractéristique des apprentis lecteurs comme nous l'avions supposé. Par conséquent, le test de *L'alouette* ne semble pas mesurer le décodage comme il le prétend puisque ce sont les bons lecteurs qui tombent dans les pièges.

D'autre part, dans une perspective développementale, nous nous attendions à observer une discontinuité en ce qui concerne les types d'erreurs. Autrement dit, nous nous attendions à ce que la nature des erreurs rende compte d'une évolution dans le rapport code/sens : les élèves les plus jeunes devant faire davantage d'erreurs lexicales et les plus âgés des erreurs de type grapho-sémantiques. Les différentes analyses qualitatives menées à ce sujet révèlent effectivement une progression dans les types d'erreurs. Ainsi, nous avons constaté que, alors que les apprentis lecteurs s'attachent à une construction de sens au niveau du mot, les lecteurs plus avancés essaient quant à eux de donner du sens au niveau du syntagme voire de la phrase. Il apparaît donc que, quel que soit le niveau d'avancement en lecture, les élèves cherchent à donner du sens et que cette recherche de sens s'élargit au fil du temps. Ce constat va donc à l'encontre des suppositions des modèles développementaux selon lesquelles la recherche de sens n'interviendrait que plus tard dans l'apprentissage de la lecture.

Chapitre 15

PROCESSUS DE LECTURE ET EVOLUTION DU CP AU CM2

Plan du chapitre

INTRODUCTION	322
PREDICTIONS.....	323
METHODOLOGIE D'ANALYSE	323
I. Quels sont les processus dominants à chaque niveau scolaire ?	324
A) Les élèves de CP.....	324
B) Les élèves de CE1.....	326
C) Les élèves de CE2.....	327
D) Les élèves de CM1.....	328
E) Les élèves de CM2.....	329
II. De quelle façon évoluent ces processus d'une classe à l'autre ?	330
A) Le décodage	331
B) L'intention lexicale.....	331
C) Le processus non identifié.....	331
D) L'abstention	332
CONCLUSION	333

INTRODUCTION

Dans les modèles les plus reconnus, l'apprentissage de la lecture, considéré à travers les processus d'identification du mot, reposerait sur une progression prenant en compte les différents aspects de la langue. Cette progression est alors envisagée de manière discontinue puisque l'enfant traverserait une série d'étapes (ou de stades) durant cet apprentissage. Ces modèles distinguent trois ou quatre étapes. La première est une étape dite logographique. Elle se caractérise par un traitement visuel des mots, la phonologie et l'ordre des lettres n'interviennent pas. Autrement dit, l'enfant traite le mot comme un objet visuel et non linguistique. Certains chercheurs distinguent deux phases dans cette étape logographique : une première, caractérisée par une reconnaissance globale des mots appris par cœur (Frith, 1985, 1986 ; Marsh et *al.*, 1981) et une seconde au cours de laquelle l'enfant reconnaît les mots à partir d'indices graphiques saillants (« discrimination par coup de filet ») (Marsh et *al.*, 1981 et Harris & Coltheart, 1986). Suite à cette étape logographique, l'enfant passerait par une étape alphabétique dans laquelle l'ordre des lettres et la phonologie deviennent cruciaux. La terminologie utilisée pour désigner cette étape et l'unité de traitement de l'enfant varient selon les chercheurs. Ainsi, Frith (1985, 1986) et Morton (1989) parlent de phase alphabétique, dont l'unité de traitement serait le graphème. Harris & Coltheart (1986) parlent quant à eux de « procédure phonologique ». En revanche, Marsh et *al.* (1981) parlent de « décodage séquentiel » dont l'unité de traitement serait la lettre. Enfin, la troisième étape est l'étape orthographique. Durant cette étape, l'enfant, pour reconnaître un mot, n'a plus recours à la phonologie mais à des unités orthographiques. Pour résumer, l'enfant, lors de l'apprentissage de la lecture traverserait trois grandes étapes : une étape logographique, caractérisée par un traitement visuel des mots, une étape alphabétique, caractérisée par la mise en application du code et une étape orthographique caractérisée par la recherche de sens. Selon ces modèles, la succession de ces étapes est stricte et identique à tous les élèves.

Dans cette conception, le test de *L'alouette* semble particulièrement représentatif puisqu'il repose sur un apprentissage traditionnel de la lecture et mesure essentiellement le niveau de décodage.

L'objectif de ce chapitre est donc de voir, à partir du test de *L'alouette*, ***quels sont les processus dominants à chaque niveau scolaire et de quelle façon ils évoluent d'une classe à l'autre.***

PREDICTIONS

Partant, d'une conception traditionnelle de la lecture nous nous attendons à ce qu'au début de l'apprentissage institutionnel de la lecture, les élèves identifient un mot par une procédure de décodage. La recherche de sens n'interviendrait qu'ultérieurement.

Nous nous attendons à ce que la maîtrise du code fasse diminuer le recours au décodage et fasse augmenter les procédures lexicales d'année en année.

METHODOLOGIE D'ANALYSE

Nous avons d'abord répertorié les différents processus utilisés par les élèves (cf. chapitre 7 « *Types de réponses erronées et processus de lecture* »). Nous en distinguons trois :

- **Décodage** : la verbalisation est un non-mot. L'élève essaie de déchiffrer sans vouloir donner du sens.
- **Intention lexicale** : la verbalisation est un mot. L'élève veut donner du sens à sa lecture en ajoutant des éléments extérieurs.
- **Abstention** : le mot n'a pas été lu. L'élève n'a rien prononcé.
- **Processus non identifié** : la verbalisation est un mot mais il est impossible de savoir si l'élève a eu l'intention de produire un mot ou pas.

Puis, nous avons regardé pour chaque élève, parmi le nombre d'erreurs commises qu'elles sont celles qui découlent d'un processus de décodage, d'une intention lexicale, d'une abstention et d'un processus non identifié.

Ensuite, nous avons regardé pour chaque niveau scolaire, qu'elle est en moyenne la proportion d'élèves qui ont eu recours à chaque procédure. Notre distribution ne suivant pas une loi normale, nous avons donc utilisé des tests non paramétriques.

Ainsi, afin de voir si, pour chaque niveau scolaire, il existe une différence significative dans l'utilisation des différentes procédures, nous avons eu recours au test de Friedman, pour échantillon apparié, puisque pour chaque élève il y a quatre procédures. Dans une seconde partie, afin de voir de quelle façon ont évolué ces processus, nous avons eu recours au test de

Kruskal-Wallis, pour échantillons indépendants, puisqu'il y a cinq niveaux scolaires (cf. *CD annexes*).

I. Quels sont les processus dominants à chaque niveau scolaire ?

Lorsqu'on regarde quels sont les processus dominants à chaque niveau scolaire, il apparaît que ces derniers semblent varier selon le niveau de scolarisation (cf. tableau 64).

Tableau 64 : proportion d'élèves qui ont eu recours aux différents processus selon la classe

Classe	Décodage	Intention Lexicale	Non identifié	Abstention
CP (N=143)	54%	12%	12%	22%
CE1 (N=143)	45%	18%	19%	18%
CE2 (N=143)	36%	22%	26%	16%
CM1 (N=143)	29%	30%	29%	12%
CM2 (N=143)	23%	39%	28%	11%

Lecture : 54% des élèves de CP qui ont commis des erreurs ont eu recours au décodage

En effet, il semblerait que jusqu'en CM1, les élèves aient essentiellement recours au décodage. A un niveau plus avancé, en CM2, les élèves semblent être davantage dans la recherche de sens. Voyons, ce qu'il en est plus précisément, pour chaque niveau scolaire.

a) Les élèves de CP

On remarque qu'en CP les élèves qui essaient avant tout de déchiffrer les mots sans chercher à y attribuer du sens sont significativement¹ les plus nombreux (54%).

Parmi les erreurs les plus fréquentes on observe une lecture littérale des graphèmes composés (lettre par lettre) comme par exemple le mot « *printemps* » lu [pɾitɛ], [pɾintɛ] ou encore [pɾi]. On observe aussi des lectures séquentielles entraînant la plupart du temps un découpage erroné du mot. Ainsi, « *renouveau* » est souvent lu [Rãnuvo], [Rãuvo] ou [Rɛnuvɛo]. On note aussi que, les lectures sont souvent incomplètes, comme [pɾɛ], [pɾɛtɛ] ou [Rɛnuvɛ]. De la même façon, on remarque que l'essentiel des élèves ont lu [sɛ] pour le mot « *ses* ». Parmi les erreurs de décodage on trouve aussi quelques erreurs de conversion

¹ Décodage-Intention Lexicale ($\chi^2=10,787$; ddl =4 ; $p<.01$), Décodage-Processus non identifié ($\chi^2=10,993$; ddl =4 ; $p<.01$) et décodage-abstentions ($\chi^2=8,634$; ddl =4 ; $p<.01$)

grapho-phonémique. Par exemple, « *gamine* » lu [ʒaminə] ou [kaminə]. Pour la première erreur, il semble que ce soit la règle de prononciation du « g » qui, lorsqu'il est suivi d'une voyelle, se lit [ʒ], ne soit pas maîtrisée. Pour la seconde, les élèves semblent avoir confondu deux sons proches : [g] et [k].

Concernant les autres processus, on constate qu'il n'y a pas de différence significative entre ceux qui ont verbalisé un mot et ceux qui n'ont rien lu. Ainsi, ils sont 12% à avoir produit un mot -12% semblent avoir eu une intention lexicale- et ils sont autant pour lesquels le processus est difficilement identifiable. Les intentions lexicales portent essentiellement sur le mot « *haies* » lu [ʃa]. Ici, les élèves ont effectué un décodage partiel du mot et se sont peut être aidés du dessin du chat pour produire cette réponse. Les verbalisations de mots dont les processus ne peuvent pas être clairement identifiés apparaissent lors de la lecture des mots « *bois* » et « *amie* » lu respectivement [dwa] et [ani]. Pour le premier mot, il est difficile de savoir si celui-ci a été prononcé suite à une confusion de graphèmes (b/d) ou par volonté de donner du sens. Pour le second mot, on remarque que le mot « *Annie* » apparaît plus tôt dans le texte, ainsi, pour cette erreur il est difficile de savoir si la lecture de l'élève a été interférée par ce mot ou si elle est le résultat d'une méprise entre le « m » et le double « n ».

Les élèves qui n'ont rien prononcé sont 22% en CP. Toutefois, il n'y a pas de différence significative avec ceux qui ont verbalisé un mot, qu'il y ait eu une intention lexicale ou non.

En **résumé**, on remarque qu'en CP, les élèves ont essentiellement recours au décodage pour lire des mots. Ce constat semble donc aller dans le sens des modèles de l'apprentissage de la lecture qui supposent qu'au début de cet apprentissage les enfants passent par une étape alphabétique caractérisée par l'apprentissage du code. Néanmoins, on observe tout de même que 12% des élèves ont eu une intention lexicale et que 12% des élèves ont verbalisé un mot, que ce soit de manière intentionnelle ou non. Autrement dit, 24% des élèves de CP ont verbalisé un mot. Bien qu'il n'y ait pas de différence significative, cela représente tout de même un quart des élèves. Ces résultats semblent donc aller à l'encontre des modèles développementaux qui supposent que la recherche de sens n'intervient que plus tard dans l'apprentissage de la lecture. Par conséquent, il semblerait que le phénomène de discontinuité ne soit pas absolu.

b) Les élèves de CE1

La tendance est la même en CE1, puisque les élèves qui ont commis une erreur de décodage sont à ce niveau aussi les plus nombreux (45%)². Contrairement aux élèves de CP, on remarque que sont surtout les graphèmes dont la prononciation est irrégulière qui ont entraîné des erreurs. Ainsi, par exemple, la majorité des élèves de CE1 qui ont mal lu le mot « *féeries* » ont prononcé [fɛri]. En effet, si on n'a jamais rencontré ce mot, il est tentant de lire de cette façon. Or ce n'est pas la prononciation la plus courante qui est [fœri]. Ainsi, ce mot ne peut être correctement lu si on se base simplement sur le décodage. Les mots « *d'osier* » et « *luisant* » ont été la plupart du temps lus [dosjɛ] et [lɥisɑ̃]. Ici, le « s » est un graphème positionnel, comme le « g » de « *gamine* », c'est-à-dire que selon sa position dans le mot il va se prononcer différemment. Ainsi, il est possible que les élèves ne sachent pas encore que le « s » entouré de voyelles se prononce [z]. Une autre erreur de décodage fréquent en CE1 est la lecture des mots « *les haies* » lu [leze]. Dans ce cas, c'est la liaison avec le « h » aspiré du mot « *haies* » qui pose des difficultés. Il est probable que les élèves aient abordé le phénomène de liaison, mais n'en connaissent pas encore les exceptions, comme c'est le cas avec un « h » aspiré. Les erreurs dues à un mauvais découpage semblent être moins nombreuses qu'au niveau précédent et touchent essentiellement le mot « *paille* » lu [pɛjə] ou [pɛlɐ].

Comme en CP, les intentions lexicales sont aussi nombreuses que les processus non identifiés. En CE1, on compte 18% d'intentions lexicales. Elles portent essentiellement sur les mots « *l'ennui* » et « *nœud* » lus respectivement [lɛnɥi]/[lanɥi] et [nœf]. Comme pour le mot « *haies* » lu [fa], il est possible que ces lexicalisations soient un effet des dessins présents autour du texte.

Les processus non identifiés sont de 19% et portent essentiellement sur les mots « *pompe* » lu [pɔm], « *gai* » lu [ge] et « *rixes* » lu [risk]. Pour la première erreur, il paraît difficile de savoir si cette lecture est due à une méprise des sons [p] et [m] ou si le mot produit est le résultat d'un effet du dessin de la pomme, situé juste à côté. Pour la seconde erreur, le mot « *gai* » est présent juste avant dans le texte ; de cette façon, il est là aussi difficile de savoir si cette prononciation est le résultat d'une interférence avec le mot précédent ou d'une méconnaissance de règle de prononciation. La troisième erreur quant à

² Décodage-Intention Lexicale ($\chi^2=8,798$; ddl =4 ; $p<.01$), Décodage-Processus non identifié ($\chi^2=8,308$; ddl =4 ; $p<.01$) et décodage-abstentions ($\chi^2=10,410$; ddl =4 ; $p<.01$)

elle, peut être envisagée sous deux processus différents : soit l'élève a effectué un décodage partiel suivi d'une lexicalisation, soit c'est une confusion des sons [ks] et [sk].

Les élèves qui n'ont rien prononcé sont 18%. Ce sont principalement les mots « *haies* » et « *l'écaille* » qui n'ont pas été lus.

En **résumé**, les élèves de CE1, comme ceux de CP, ont essentiellement recours au décodage pour lire des mots. Toutefois, si on regroupe les catégories intentions lexicales et processus non identifiés sous l'étiquette « verbalisation de mots », on remarque qu'il n'y a plus de différence significative avec les erreurs de décodage. Autrement dit, en CE1, les élèves qui verbalisent des mots sont aussi nombreux que ceux qui verbalisent des non-mots, ce qui signifie que la recherche de sens devient plus importante. Ce constat semble donc confirmer que l'apprentissage de la lecture ne se déroule pas de façon strictement discontinue puisque, à ce niveau scolaire, on répertorie autant d'élèves qui ont recours au décodage qu'à une procédure lexicale.

c) *Les élèves de CE2*

En CE2, les élèves qui ont eu recours au décodage sont 36% et restent les plus nombreux. On retrouve les erreurs commises sur le graphème irrégulier de « *féeries* » entraînant une lecture erronée : [fɛri]. L'erreur due à une liaison abusive sur les mots « *les haies* » est encore présente. On remarque que le mot « *paille* » est moins lu par décodage, en revanche, il est remplacé par le mot « *piaille* ». Ce dernier est essentiellement lu [pjɛjə] ou [pjɛlə].

On ne constate pas de différence significative entre les élèves qui ont eu une intention lexicale ou dont le processus de lecture est difficilement identifiable. En effet, ils sont 22% dans la première catégorie et 26% dans la seconde. L'erreur lexicale la plus répandue en CE2 est « *l'hirondeau* » lu [lirɔ̃dɛl]. On remarque que ces élèves ont effectué un décodage partiel et ont ensuite produit un mot qui leur est certainement plus familier. Il est possible que la lecture du début du mot ait activé dans leur lexique mental le mot « *hirondelle* ». De plus, là aussi, il est possible que le dessin de l'hirondelle ait joué un rôle dans cette lexicalisation. Parmi les verbalisations de mots dont nous ignorons le processus, on trouve les mots « *poisons* » lu [pwasɔ̃], « *geai* » lu [ge] et « *cordeaux* » lu [kɔ̃ʁbɔ]. En effet, on se demande

si pour le mot « poisons », il a été lu [pwasɔ̃] à cause du co-texte sémantique dans lequel il apparaît ou à cause d'une méprise sur la prononciation du « s ». Il en est de même pour le mot « geai » lu [gɛ], est-ce un effet de similitude ou est-ce une confusion de son ? Le mot « cordeaux » lu [kɔ̃ʁbɔ] nous amène lui aussi à nous demander si cette prononciation est le résultat d'une confusion de lettres (b/d) ou d'un effet du dessin du corbeau situé juste à côté.

Les élèves qui non rien prononcé sont 16%. Contrairement aux niveaux précédents, on remarque ici, que c'est un syntagme qui a été le plus sauté : « *un brin d'osier* ».

En **résumé**, il semblerait qu'en CE2, comme pour les niveaux précédents, les élèves ont essentiellement recours au décodage pour lire un mot. Néanmoins, si comme en CE1, on regroupe les verbalisations de mots, on constate que cette catégorie devient la plus importante. Ainsi, à partir du CE2, les élèves produisent plus de mots que de non-mots (48% contre 36%), ce qui signifie que la recherche de sens prime sur la mise en correspondance des lettres et des sons. Il semble donc qu'à partir du CE2, les élèves essaient davantage de donner du sens à leur lecture plutôt que faire correspondre une chaîne écrite à une chaîne orale. Toutefois, nous précisons que même dans cette comparaison, les élèves qui ont recours au décodage restent nombreux, ce qui semble confirmer un apprentissage de la lecture qui n'est pas strictement discontinu.

d) Les élèves de CM1

A ce niveau, on constate que les erreurs de décodage sont aussi nombreuses que les erreurs lexicales et les verbalisations de mots qui découlent d'un processus non identifié.

Ainsi, on observe 29% d'erreurs dues à un mauvais décodage. On retrouve essentiellement le mot « féeries » lu [fɛri] et la liaison abusive sur « *les haies* ».

Les intentions lexicales représentent 30% des erreurs. La principale erreur ne concerne pas un ajout de lettres mais un ajout de mot. En effet, elle se situe lors de la lecture du syntagme « *au clair de lune* » lu « au clair de la lune » : les élèves sont une majorité à avoir rajouté le mot « la ». La seconde erreur la plus fréquente est le « *la* » du syntagme « *prête-moi la plume [...]* » lu [ta]. On remarque ici, que ces erreurs portent sur un passage du texte qui ressemble fortement à la chanson « Au clair de la lune ». Il est donc probable que les élèves aient cru que c'étaient les paroles de la chanson et ont donc commis ces erreurs.

Les productions de mots dont les processus ne sont pas clairement identifiables sont de 29% et concernent principalement le mot « *poisons* » lu [pwas⁵]. Comme pour les niveaux précédents, on se demande si cette erreur est due à une confusion de son ou s'il y a réellement eu une intention lexicale. Néanmoins, nous pensons qu'à ce niveau de la scolarisation, les élèves maîtrisent les sons [s] et [z] et qu'il est plus probable que ces élèves aient commis cette erreur à cause du co-texte sémantique dans lequel il apparaît.

Les abstentions sont quant à elles 12%. Alors qu'en CP et CE1, on observait des sauts de mots, en CE2 des sauts de syntagme. On remarque en CM1, que ce sont la plupart du temps des sauts de phrase.

En **résumé**, il apparaît qu'en CM1, les élèves ont autant recours au décodage qu'aux productions de mots qu'il y ait eu une intention ou pas. Cependant, si, là aussi on compare les verbalisations de mots et de non mots, on constate que, comme les élèves de CE2, ceux de CM1, produisent plus de mots que de non mots (59% contre 29%), ce qui signifie qu'à ce niveau scolaire c'est la recherche qui est plus importante. Néanmoins, le fait qu'il y ait encore 29% d'élèves qui ont recours au décodage, semble aller dans le sens d'un apprentissage discontinu, certes, mais pas absolu.

e) Les élèves de CM2

En CM2, même si on regroupe les catégories en production de mots et non mots, on remarque que les verbalisations de mots sont les plus nombreuses : 67% contre 23%. Dans le détail on note que ce sont les intentions lexicales les plus nombreuses (39%). Comme précédemment, les élèves ont commis les mêmes erreurs lors de la lecture des syntagmes « *au clair de lune* » et « *prête-moi la plume* ». On retrouve aussi le mot « *poisons* » lu [pwas⁵].

En **résumé**, il apparaît qu'en CM2, les élèves essaient essentiellement de donner du sens à leur lecture. Ce constat semble donc aller dans le sens de la plupart des modèles d'apprentissage de la lecture qui postulent que suite au stade alphabétique caractérisé par la maîtrise du code, se trouve une phase orthographique caractérisée par la recherche de sens. Toutefois, précisons qu'ici aussi, encore presque un quart des élèves a recours au décodage,

ce qui signifie, que l'apprentissage de la lecture ne se déroule pas de façon strictement discontinue.

En **résumé**, cette analyse nous permet de constater que les élèves n'ont pas recours aux mêmes procédures pour lire des mots selon le niveau de scolarisation. En effet, alors qu'au début de l'apprentissage institutionnel de la lecture, les élèves ont essentiellement recours au décodage, on remarque qu'en CM1, il n'y plus de différence significative entre les élèves qui ont recours à ce processus et ceux qui ont verbalisé un mot, que ce soit de façon intentionnelle ou pas. En fin d'école primaire les élèves qui produisent des mots de façon intentionnelle sont les plus nombreux, ce qui signifie que pour eux, la recherche de sens est primordiale. Ce premier constat semble donc aller dans le sens des modèles développementaux qui supposent que les élèves passeraient d'abord par une étape de décodage avant d'accéder au sens. Néanmoins, si on compare les verbalisations de mots avec les verbalisations de non mots, le constat change. En effet, on remarque qu'en CP les élèves ont essentiellement recours au décodage puisqu'ils produisent majoritairement des non-mots. En revanche en CE1 on constate qu'il y a autant de verbalisation de mots que de non mots. Puis à partir du CE2, on remarque que les élèves qui produisent des mots deviennent les plus nombreux. Ainsi, les élèves semblent vouloir donner du sens à leur lecture dès le CE1 et non dès le CM1, ce qui signifierait que la recherche de sens interviendrait tôt dans l'apprentissage de la lecture. En effet, bien qu'en CP, les élèves ont essentiellement recours au décodage, on compte tout de même 24% d'élèves qui ont verbalisé un mot. Autrement dit, l'apprentissage de la lecture ne se déroulerait pas de façon strictement discontinue puisque dès le début, un quart des élèves essaie de donner du sens à leur lecture.

II. De quelle façon évoluent ces processus d'une classe à l'autre ?

Partant d'une conception traditionnelle de l'apprentissage de la lecture, nous nous attendons à ce que la maîtrise du code grapho-phonologique fasse diminuer le nombre d'élèves qui ont recours au décodage et fasse augmenter le nombre d'élèves qui ont recours à une procédure orthographique.

a) Le décodage

Lors de l'analyse précédente, nous avons constaté que 54% des élèves ont commis une erreur suite à un mauvais décodage. On observe qu'au niveau suivant en CE1, leur nombre diminue de façon significative puisqu'ils sont moins de la moitié dans cette catégorie : 45% ($\chi^2(4) = 2.923 ; p < .05$). En CE2, cette proportion continue de baisser et est de 36% ($\chi^2(4) = 3.852 ; p < .01$). Enfin, aux niveaux suivants, le nombre d'élèves ayant recours au décodage semble se stabiliser puisqu'on ne constate pas de différence significative entre ces deux classes : ils sont 29% en CM1 et 24% en CM2 ($\chi^2(4) = 2.765 ; NS$).

Ce constat semble donc aller dans le sens de nos prédictions, puisque le nombre d'élèves ayant recours au décodage diminue d'année en année.

b) L'intention lexicale

Au début de l'école primaire, en CP, 12% des élèves ont eu recours de façon intentionnelle à la lexicalisation lors de la lecture de mots. En CE1, leur nombre augmente significativement : ils sont 18% ($\chi^2(4) = -3.164 ; p < .05$). On remarque qu'il n'y a pas de différence significative avec le CE2, bien qu'ils soient légèrement plus nombreux : 22% ($\chi^2(4) = -2.355 ; NS$). En revanche, l'augmentation redevient significative en CM1, puisque cette proportion est à ce niveau scolaire de 30% ($\chi^2(4) = -4.002 ; p < .01$). Elle augmente davantage en fin d'école primaire, puisqu'elle compte 39% des élèves ($\chi^2(4) = -3.502 ; p < .01$).

Ce constat semble donc confirmer nos prédictions, puisque la proportion d'élèves à avoir recours à une procédure lexicale augmente d'année en année.

c) Le processus non identifié

La proportion d'élèves qui ont verbalisé un mot mais dont nous ignorons s'ils ont eu l'intention ou pas d'effectuer une lexicalisation, suit à peu près la même tendance que celle des intentions lexicales. En effet, ces élèves qui sont 12% en CP représentent 19% en CE1 ($\chi^2(4) = -4.437 ; p < .01$). En CE2, leur nombre continue d'augmenter puisqu'ils représentent à ce niveau 26% des élèves ($\chi^2(4) = -4.328 ; p < .01$). En revanche, à partir du CM1, on observe un effet plafond, car il n'existe plus de différence significative : ils sont 29% en CM1 ($\chi^2(4) = -2.539 ; NS$) et 28% en CM2 ($\chi^2(4) = 1.162 ; NS$).

Si on ajoute à cette catégorie les intentions lexicales, on retrouve à peu près les mêmes résultats : les verbalisations de mots augmentent de façon significative jusqu'en CM1. En effet, alors que 24% des élèves de CP ont verbalisé un mot, ils sont 37% en CE1. Au niveau suivant, cette proportion continue d'augmenter puisqu'ils sont 48%. En CM1, c'est plus de la moitié des élèves qui ont verbalisé un mot (59%). Enfin, en CM2, la proportion d'élèves dans cette catégorie s'élève à 67%.

Ce constat semble donc confirmer, la mise en place d'une procédure lexicale au fur et à mesure de l'apprentissage de la lecture.

d) L'abstention

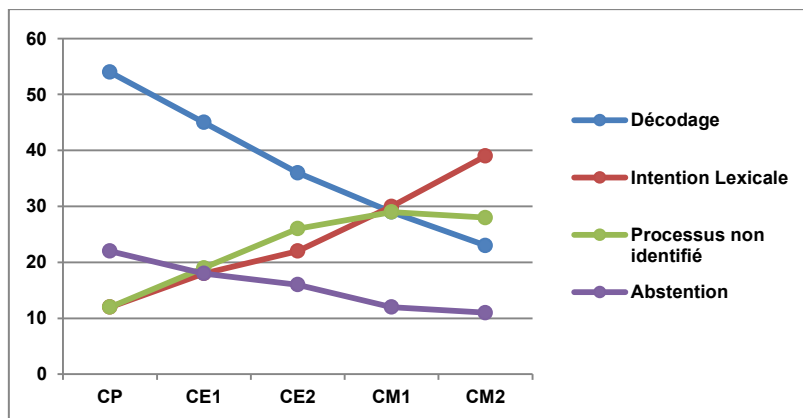
En ce qui concerne le processus d'abstention, on remarque qu'il suit la même tendance que ceux qui ont recours au décodage. Ainsi, 22% des élèves en CP, ont préféré ne pas lire les mots. En CE1, leur nombre diminue mais pas de façon significative, ils sont 18% ($\chi^2(4) = 2.343$; *NS.*). Cette proportion est la même pour les élèves en CE2 (16%) ($\chi^2(4) = .599$; *NS.*). En CM1, bien que ces élèves soient un peu moins nombreux (12%), on n'observe pas de différence significative ($\chi^2(4) = .069$; *NS.*) tout comme en CM2, où ils sont 11% ($\chi^2(4) = .772$; *NS.*). En réalité, le phénomène d'abstention est essentiellement significatif pour les élèves scolarisés en CP³. Cela signifierait qu'au début de l'apprentissage de la lecture, certains élèves préfèrent ne rien prononcer plutôt que de se tromper.

³ CP-CE2 ($\chi^2(4) = 2.889$; $p < .05$), CP-CM1 ($\chi^2(4) = 4.663$; $p < .01$), CP-CM2 ($\chi^2(4) = 5.316$; $p < .01$)

CONCLUSION

Alors que les proportions d'élèves qui ont recours au décodage et à l'abstention, diminuent entre le CP et le CM2, on observe la tendance inverse pour ceux qui ont verbalisé un mot, que ce soit de façon intentionnelle ou pas.

Graphique 47 : évolution des processus de lecture entre le CP et le CM2



Ces résultats semblent donc aller dans le sens des suppositions des modèles traditionnels de l'apprentissage de la lecture puisqu'on observe essentiellement un recours au décodage au début de l'apprentissage et un recours à la procédure lexicale en fin d'apprentissage. Ceci laisse donc penser que cet apprentissage s'effectuerait bien de manière discontinue. Cependant, nous employons le terme « essentiellement » car la procédure de décodage au début de l'apprentissage de la lecture est certes, la plus importante mais elle n'est pas la seule. En effet, on remarque que dès le début, la procédure lexicale est présente : près d'un quart des élèves l'ont utilisée pour identifier des mots, tout comme l'abstention qui concerne 22% des élèves. Ces résultats semblent donc indiquer que, la discontinuité de l'apprentissage de la lecture n'est pas absolue puisqu'à un même niveau scolaire, on retrouve des procédures différentes. Les étapes ne seraient donc pas exclusives ni identiques à tous les élèves.

CONCLUSION- DISCUSSION

Plan de la conclusion

CONCLUSION	335
<i>Une baisse relative du niveau en lecture</i>	<i>335</i>
<i>Idées reçues</i>	<i>336</i>
<i>Comment expliquer cette légère baisse ?.....</i>	<i>337</i>
<i>Les écarts se creusent.....</i>	<i>337</i>
<i>Les garçons lisent aussi bien que les filles.....</i>	<i>339</i>
<i>Les élèves en éducation prioritaire en difficulté.....</i>	<i>339</i>
<i>L'apprentissage de la lecture.....</i>	<i>341</i>
<i>Des erreurs aux processus hypothétiques.....</i>	<i>342</i>
<i>De l'effet des variables linguistiques aux variables contextuelles</i>	<i>343</i>
<i>Les apprentis lecteurs ne font pas que décoder</i>	<i>344</i>
<i>Eviter de lire pour ne pas se tromper.....</i>	<i>345</i>
<i>Un apprentissage discontinu, oui mais pas absolu.....</i>	<i>345</i>
<i>L'influence du test de L'alouette.....</i>	<i>346</i>
DISCUSSION	349
PERSPECTIVES DE RECHERCHE	350

CONCLUSION

Dans nos sociétés où l'écrit est omniprésent, savoir lire est devenu plus que nécessaire. Il est devenu une condition d'intégration sociale et culturelle. Le savoir lire, à la base de tous les autres apprentissages scolaires, est devenu un enjeu majeur et nombreuses sont les enquêtes et recherches menées sur la lecture et son apprentissage. Régulièrement, les polémiques médiatiques évoquent la question du niveau de compétence dans les savoirs de base parmi lesquels le savoir lire-écrire occupe une place centrale.

Afin de nous faire notre propre idée quant à l'évolution du niveau, plus particulièrement en lecture, nous avons fait le choix d'utiliser le test de *L'alouette* (Lefavrais, 1965). Celui-ci nous permet, en effet, d'effectuer une comparaison à quarante ans d'écart puisque l'étalonnage de 1965 est encore disponible. Ainsi, reproduisant scrupuleusement les modalités originelles de passation, nous avons fait passer à nouveau ce test à 1403 élèves scolarisés dans trois types d'établissements (école publique hors éducation prioritaire, école publique en éducation prioritaire et école privée), du CP au CM2, à trois moments de l'année (décembre, février et avril), entre 2006 et 2009. Pour plus de rigueur, nous avons extrait de ce panel, un échantillon qui soit représentatif de la population scolaire en France. En 1965, Lefavrais a constitué son échantillon à partir d'élèves scolarisés en école publique, par conséquent, nous avons comparé le niveau en lecture de cette population aux élèves scolarisés en école publique hors EP¹ des années 2000.

Adoptant la position dominante qui consiste à dire que le niveau a baissé, nous avons donc émis l'hypothèse que le niveau en lecture des élèves d'aujourd'hui est inférieur à celui des élèves de 1965.

Une baisse relative du niveau en lecture

Les résultats de cette première étude valident notre hypothèse mais de façon très modeste. En effet, d'une manière générale, on observe une baisse de seulement six mois du niveau en lecture. La comparaison par niveau scolaire, fait apparaître qu'en dépit de cette baisse générale, le niveau en lecture semble être resté stable en CP, CE1 et CM2, alors qu'il a diminué de six mois pour les CE2 et de neuf mois pour les CM1.

¹ Education Prioritaire. A la rentrée 2009, l'éducation prioritaire comprend les établissements en RAR (Réseaux Ambition Réussite) et RRS (Réseaux de Réussite Scolaire).

Toutefois, cette comparaison nous a semblé biaisée puisque nous n'avons pas tenu compte des élèves scolarisés en éducation prioritaire. Bien que l'échantillon de Lefavrais était constitué d'élèves en école publique, les élèves qui fréquentent aujourd'hui les écoles en EP étaient déjà présents il y a quarante ans, même si les structures spécifiques actuelles n'existaient pas encore. Par conséquent, ces premiers résultats sont discutables. Ainsi, pour plus de rigueur, nous avons effectué une nouvelle comparaison en y incluant les élèves de l'éducation prioritaire.

Les résultats font apparaître que, lorsqu'on inclut les élèves scolarisés en éducation prioritaire, la tendance générale ne change pas : le niveau en lecture a diminué de six mois à quarante ans de distance. Le constat est aussi le même pour les élèves de CP et CE1 dont le niveau en lecture est resté stable et pour les élèves de CE2, dont le retard est toujours de six mois. En revanche, on observe une baisse un peu plus prononcée pour les élèves de CM1 et CM2. En effet, le retard en lecture passe de neuf mois à onze mois pour les premiers. Quant aux élèves de CM2, dont le niveau est stable sans les élèves de l'éducation prioritaire, on remarque qu'il diminue de trois mois lorsqu'on les inclut.

La baisse constatée est donc relativement faible. De plus, on remarque qu'elle n'a pas le caractère cumulatif ni progressif auquel on aurait pu s'attendre suivant l'opinion d'une baisse générale du niveau commençant en CP et croissant d'année en année.

Idées reçues

Si l'on en croit ceux qui remettent en cause les méthodes d'enseignement de la lecture en ses tout débuts, on aurait dû s'attendre à observer une baisse en CP, or, ce n'est pas le cas, le niveau, tel que mesuré par *L'alouette*, reste stable avec ou sans les élèves en EP. Par conséquent, ce résultat invite à dédouaner les méthodes d'enseignement puisqu'à quarante ans d'écart le niveau en lecture au début de l'école primaire reste inchangé.

De plus, la focalisation sur les méthodes d'enseignement au CP laisse penser que tout se passe là. Or, notre analyse montre que finalement, c'est en CE2 et CM1 que la baisse du niveau est la plus forte. Ce constat semble donc indiquer que tout ne se joue pas en CP.

Il en est de même avec l'idée qu'en fin d'école primaire, un quart des élèves ne savent pas lire. En effet, si on considère les élèves scolarisés en école publique hors EP, on observe une stabilité du niveau en lecture en CM2. Si on inclut cette population, alors on observe un léger retard de trois mois en lecture, ce qui reste relativement faible.

Comment expliquer cette légère baisse ?

La baisse du niveau pourrait s'expliquer de plusieurs façons : baisse des exigences de la part des enseignants, diminution du nombre d'heures consacrées à l'enseignement de la langue, primauté de la quantité au détriment de la qualité, etc. Nous pourrions aussi émettre l'hypothèse d'une baisse des pratiques de lecture. En effet, selon une récente étude menée par la Direction des Etudes Statistiques sur les pratiques culturelles entre 1973 et 2008, il apparaît qu'en 1997, 74% des Français de plus de 15 ans ont lu au moins un livre au cours de l'année alors qu'ils sont 70% en 2008. Le nombre d'inscrits en bibliothèque a diminué lui aussi durant cette période, il est passé de 20% à 18%. La lecture de quotidiens est passée de 73% en 1997 à 69% en 2008².

Cependant, cette baisse peut aussi s'expliquer par des facteurs méthodologiques. En effet, rappelons que, pour 1965, nous ne disposons que de très peu d'informations sur la façon dont Lefavrais a constitué son échantillon. Plus précisément, nous ne savons pas à quel moment de l'année le recueil des données a été effectué. Ainsi, le retard de six mois en lecture constaté en 2009 pourrait ne pas être considéré comme tel si les mois où Lefavrais a effectué ses passations sont différents des nôtres. On ne peut donc écarter l'idée que les variations observées soient au moins en partie l'effet de différences sur ce plan.

De plus, ne disposant pas de la proportion d'élèves scolarisés en école élémentaire en éducation prioritaire, nous avons utilisé le nombre d'établissements. Les élèves de l'éducation prioritaire sont, par conséquent, surreprésentés dans notre échantillon.

Par ailleurs, la comparaison avec les écoles en éducation prioritaire met en évidence un fait intéressant en ce qui concerne les élèves les plus faibles. Alors que leur niveau reste stable pour ceux qui sont scolarisés en école publique hors EP, on constate une différence significative lorsqu'on inclut celles situées en EP. Cette différence non visible avec les médianes laisse supposer des variations internes à l'échantillon. Autrement dit, la prise en compte des élèves en EP semble entraîner un accroissement de la dispersion.

Ainsi, nous avons mené une troisième analyse afin de savoir si, à quarante ans de distance, la dispersion des échantillons est la même.

Les écarts se creusent

Cette étude fait ressortir des faits tout aussi intéressants. En effet, on constate que notre échantillon est plus hétérogène que celui de 1965 (accroissement des écarts types). Plus

² Donnat O. (2009). Pratiques culturelles, 1973-2008. Dynamiques générationnelles et pesanteurs sociales. Département des études statistiques (DEPS).

précisément, si à un niveau intermédiaire, les scores des élèves sont homogènes entre les deux années, ce n'est pas le cas en ce qui concerne les extrémités des distributions. Ainsi, il apparaît que les élèves les plus faibles d'hier sont plus faibles aujourd'hui et que les élèves les plus forts d'hier, sont encore plus forts aujourd'hui. On remarque donc un décrochage des élèves les plus en difficulté et un décollage des meilleurs. Il semble donc que nous puissions ici faire notre la conclusion des sociologues de l'éducation selon lesquels : « *L'écart se creuse entre un peloton de tête [...] plus performant que jamais ; et les autres plus démunis que jamais [...]* » (Baudelot et Establet, 1989, p.157).

Plusieurs explications peuvent être données quant à cet accroissement des écarts.

D'abord, nous pourrions émettre l'hypothèse d'effets négatifs du redoublement. Ainsi, selon plusieurs études, le redoublement ne permettrait pas d'améliorer les performances scolaires des élèves concernés (cf. Jackson, 1975, Seibel, 1984, Grisay, 1983). Les enquêtes internationales PISA (2009) et PIRLS (2006) confirment ce constat en indiquant que les pays dont le taux de redoublement est élevé ont de mauvais résultats.

De la même façon, ces enquêtes montrent que le transfert d'élèves en difficulté dans d'autres établissements, ne contribue pas non plus à améliorer les performances. L'enquête menée par PISA conclut que les établissements qui cherchent à regrouper les élèves qui ont des difficultés soit par redoublement, soit par transfert ne contribuent pas à l'amélioration des performances.

D'autres facteurs, plus sociologiques, sont susceptibles de rendre compte de cet accroissement des écarts. Par exemple, la modification de la structure sociale pourrait être une explication. L'accroissement des écarts entre les classes populaires et les classes favorisées serait en lien avec cet accroissement du niveau en lecture. En effet, les nombreuses recherches menées au sujet de l'influence du milieu social montrent clairement que les enfants de milieux défavorisés éprouvent plus de difficultés de lecture que ceux de milieux favorisés. Le capital culturel étant différent d'une classe à l'autre, l'habitus lectural ne se développe donc pas de la même façon et se trouve être plus réduit pour les enfants de milieux peu favorisés. Les sociologues Baudelot et Establet (2009) indiquent que « *on retrouve donc au niveau des sociétés ce qu'on observait dans les écoles. Les résultats sont d'autant meilleurs que les inégalités sociales sont faibles* ».

Si on étudie plus en détail les caractéristiques des groupes en tête et en queue de distribution, on remarque que, chez les plus faibles ou chez les plus forts, on compte autant de filles que de garçons. Le genre ne semble donc pas être ici un facteur discriminant. En revanche, on constate que chez les plus faibles, bien que les élèves scolarisés en EP soient les moins nombreux, ce sont eux qui accusent le plus gros retard en lecture (sept mois contre trois mois pour ceux scolarisés en école hors EP). Quant au groupe des plus forts, on remarque que les élèves en EP, sont minoritaires et ne présentent qu'une légère avance (deux mois par rapport aux neuf mois des élèves hors EP).

Partant de ce constat nous nous sommes demandée si, dans l'ensemble, des variables individuelles telles que le genre ou le milieu social d'appartenance ne peuvent pas influencer le niveau lecture.

Les garçons lisent aussi bien que les filles

Lorsqu'on compare les scores en lecture des filles et des garçons, on constate contre toute attente qu'il n'y a pas d'effet de genre : les garçons lisent aussi bien que les filles quel que soit le niveau de scolarisation et le type d'établissement fréquenté. Notre hypothèse est par conséquent rejetée.

Depuis la généralisation de l'entrée en 6e dans les années 1970, les filles devançaient les garçons à l'école et on avait l'habitude d'entendre que les filles réussissaient mieux que les garçons. D'ailleurs les évaluations en lecture des dernières années indiquent des différences de genre dans les performances en faveur des filles. Néanmoins, ce constat semble indiquer que les garçons ont rattrapé leur retard en lecture. Une enquête menée par la DEPP en 2011 et les résultats de l'enquête PIRLS (2011) semblent aller dans le sens de nos résultats, puisqu'elles observent un resserrement des écarts entre les filles et les garçons depuis quelques années. Pourrions-nous dire alors que les filles et les garçons seraient sur le chemin de l'égalité ?...

Les élèves en éducation prioritaire en difficulté

La comparaison par type d'établissement nous permet de constater que ce sont les élèves scolarisés en EP qui ont le plus faible niveau en lecture, tandis que ceux qui sont scolarisés en école publique hors EP et en école privée ont un niveau équivalent. Ainsi, ces

résultats laissent penser qu'au-delà du type d'établissement, ce serait plutôt le milieu social d'appartenance qui aurait un effet sur le niveau en lecture et, de façon plus générale, le niveau scolaire.

En effet, à la rentrée 2011³, près des trois quarts (72,9%) des collégiens scolarisés dans des établissements ECLAIR⁴ ont des parents ouvriers ou inactifs contre un tiers (34,7%) des élèves scolarisés dans des établissements hors EP. A l'inverse, 9,1% des élèves en ECLAIR ont des parents cadres ou enseignants contre 38,4% de ceux scolarisés hors EP. Il semble donc que les élèves scolarisés dans des établissements Eclair sont massivement issus d'un milieu social peu favorisé contrairement à ceux scolarisés hors EP. Par conséquent, il est probable que ce soient donc les élèves issus de milieu social peu favorisé qui rencontrent le plus de difficultés de lecture. Ce constat doit se confirmer d'une façon plus générale, puisqu'à la rentrée 2011, 22% des élèves en Eclair entrent en retard en 6^{ème} contre 11% des élèves issus d'établissements hors EP. Il semble donc que la politique de mise en œuvre d'une éducation prioritaire dans les années 80 (Savary, 1981), destinée à renforcer l'action éducative pour améliorer les résultats des élèves notamment les plus défavorisés, n'ait pas réussi à supprimer les inégalités sociales, bien au contraire.

Ainsi, il semble difficile de tirer une conclusion tranchée quant à l'évolution du niveau en lecture depuis 1965. En effet, on ne constate ni un net progrès, ni une forte baisse. On remarque seulement une baisse relative pouvant être liée aux conditions d'échantillonnage. Toutefois, ce qui apparaît clairement, c'est qu'à quarante ans de distance, on assiste à un réel accroissement des écarts entre les élèves les plus faibles et les élèves les plus forts.

Jusqu'à présent, nous avons employé le mot « lecture » à propos des scores obtenus avec le test de *L'alouette*. Or, il convient de préciser que ce test ne mesure qu'un aspect de la lecture. En effet, ce test basé sur une conception traditionnelle de la lecture mesure essentiellement le décodage et fait peu appel à la compréhension. Il est donc possible que les résultats auraient été différents si on avait employé un autre test. D'autres tests, plus complets pour mesurer la lecture existent mais les résultats obtenus par le passé ne sont plus disponibles (Par exemple, le test d'Inizan, 1963).

³ « Repères et références statistiques sur les enseignements, la formation et la recherche », édition 2012, p. 67

⁴ A la rentrée 2012, l'éducation prioritaire comprend les établissements ECLAIR (Ecoles, Collèges, Lycées pour l'Ambition, l'Innovation et le Réussite) et les établissements en RRS (Réseaux de Réussite Scolaire).

Ainsi, pour juger du niveau en lecture, il faut d'abord définir ce qu'on entend par savoir lire et apprendre à lire. En effet, il existe plusieurs façons de concevoir l'apprentissage et donc de le mesurer. De la même façon, l'acte de lire et l'évaluation de sa maîtrise font l'objet de résultats différents.

L'apprentissage de la lecture

Dans les modèles les plus reconnus, les modèles développementaux (Marsh & al., 1981, Frith, 1985 ; Harris & Coltheart, 1986 ; Morton, 1989 ; Ehri & Wilce, 1985), l'apprentissage de la lecture, considéré à travers les processus d'identification du mot, reposerait sur une progression prenant en compte différents aspects de la langue. Cette progression est alors envisagée de manière discontinue puisque l'enfant traverserait une série de stades (logographique, alphabétique et orthographique) dont l'ordre est strict et identique à tous. Selon ces modèles, le passage d'un stade à un autre ne pourrait s'effectuer que lorsque les compétences du stade antérieur sont maîtrisées. Ainsi, par exemple, l'enfant aurait recours à la procédure orthographique si et seulement si la procédure alphabétique est acquise. Bien qu'il existe des débats sur la nature et l'existence de certaines procédures, notamment l'étape logographique, tous les modèles se rejoignent sur l'importance de la phase alphabétique (ou médiation phonologique). D'ailleurs, dans cette conception, la lecture se résume au décodage et les facteurs syntaxico-sémantiques et d'une façon plus générale, contextuels, sont totalement écartés.

Le test de *L'alouette* nous a paru particulièrement représentatif ici puisqu'il mesure essentiellement le niveau de décodage. Les erreurs commises lors de la lecture de ce texte apparaîtraient donc comme une insuffisance dans le traitement alphabétique. Par ailleurs, l'utilisation de ce test présente un avantage majeur par rapport aux différentes études menées sur l'identification du mot dans le sens où il présente les mots en contexte alors que les recherches menées sur l'apprentissage de la lecture se font la plupart du temps sur des mots isolés. Cette particularité rendrait donc disponible d'autres types d'indices pour identifier les mots. L'utilisation de ce test semble donc nous rapprocher des conditions « réelles » de lecture.

Par conséquent, mettant à l'épreuve les modèles développementaux, nous avons voulu savoir dans quelle mesure les erreurs commises lors de la lecture du test de *L'alouette* nous

informent sur l'apprentissage de la lecture, et plus précisément dans quelle mesure elles rendent compte d'un apprentissage continu ou discontinu de la lecture.

Avant de nous intéresser aux erreurs, nous avons souhaité regarder du côté des acquisitions. Autrement dit, nous avons cherché à savoir qu'est-ce qui est acquis avant de mesurer les difficultés. Cette étude nous permet de constater que certains mots semblent être « définitivement » acquis alors que d'autres, pourtant simples dans leur structure semblent être acquis mais de façon partielle, c'est-à-dire qu'à certains moments ces mots sont bien lus alors qu'à d'autres non. Ces résultats nous amènent à constater qu'il existe une différence entre ce que les élèves savent lire et ce qu'ils peuvent lire. Cette différence peut être liée certes aux caractéristiques des mots mais elle semble être due aussi au test de *L'alouette* en particulier.

Des erreurs aux processus hypothétiques

Nous avons ensuite procédé à une étude dans laquelle nous avons effectué une typologie des erreurs et des procédures de lecture. Ainsi, nous distinguons des erreurs relatives à la langue et des erreurs relatives au comportement de l'élève. Autrement dit, on constate qu'il y a des erreurs portant sur les unités infra lexicales, lexicales ou supra lexicales. Ces erreurs, relatives à la langue sont le résultat d'un certain nombre de processus liés au comportement de l'élève. C'est ce que nous avons appelé les processus hypothétiques de lecture. Nous en avons répertorié trois : le décodage, l'intention lexicale et l'abstention. Nous avons considéré que l'élève qui a effectué du décodage a verbalisé un non-mot. Nous supposons que, pour lui, lire c'est déchiffrer. La lecture est alors envisagée comme la mise en correspondance d'un signal écrit avec un signal sonore. Par contre, l'élève qui a eu une intention lexicale, a verbalisé un mot à l'aide d'éléments différents de ceux qui composent le mot d'origine. Nous supposons que, pour lui, lire c'est donner du sens à ce qui est écrit. Le processus qui est l'abstention correspond au cas où l'élève n'a rien prononcé (que ce soit un mot ou une phrase). Nous supposons que pour lui, lire c'est ne pas commettre d'erreur. Ainsi, il est possible que pour eux, il soit préférable de ne rien prononcer au risque de se tromper. Parmi les élèves qui ont verbalisé des mots, il est arrivé que, pour certains, le processus qui est à l'origine soit difficilement identifiable. Par exemple, le mot « bois » lu *doigt* : la verbalisation est un mot mais est-ce le résultat d'une erreur de décodage due à la ressemblance des graphèmes « b » et

« d » ou est-ce le résultat d'une intention lexicale ? C'est ce que nous avons appelé les processus non identifiés.

Partant de cette typologie des erreurs et des différents processus hypothétiques de lecture, nous avons ensuite cherché à savoir quelles sont les caractéristiques des mots du texte qui ont été les plus difficiles à lire et de quelle façon les élèves ont procédé pour lire ces mots.

De l'effet des variables linguistiques aux variables contextuelles

D'une manière générale, on constate que les mots qui posent problème, et par conséquent les caractéristiques des mots, ne sont pas les mêmes à tous les niveaux scolaires. En effet, on remarque que jusqu'au début du cycle 2, la plupart des mots qui ont entraîné des difficultés de lecture sont des mots longs contenant un phonème retranscrit par plusieurs lettres, généralement une suite de voyelles (comme le phonème [e] du mot « haies ») ou de consonnes (comme le phonème [ã] dans le mot « printemps »). Le graphème irrégulier du mot « fêtes » a aussi amené la plupart des élèves à commettre une erreur, tout comme certains graphèmes contextuels tels que le « g » du mot « geai » ou le « s » du mot « d'osier ». Face à ces difficultés, les élèves semblent avoir essayé de décoder les mots avant tout. A partir du CM1, les mots sur lesquels le plus d'élèves ont commis une erreur de lecture changent et, il semblerait que ce ne soit plus la structure graphémique du mot ou sa longueur qui posent des difficultés, mais plutôt le contexte dans lequel ces mots s'inscrivent.

Il y a donc des erreurs qui sont le résultat de caractéristiques linguistiques comme la longueur, la fréquence ou la structure graphémique d'un mot. Et il y en a d'autres qui semblent être le résultat d'une influence du test de Lefavrais.

Ainsi, dans un premier temps, nous avons voulu savoir si les effets des variables linguistiques « classiques », observés dans les recherches portant sur l'identification du mot isolé, se retrouvent en situation de lecture de texte.

Si on se place dans le cadre des modèles développementaux qui envisagent un apprentissage discontinu de la lecture, l'enfant traverserait d'abord une étape alphabétique, caractérisée par la maîtrise du code, correspondant au début de l'enseignement institutionnel de la lecture. La recherche de sens n'apparaîtrait que plus tard, dans une étape ultérieure, l'étape orthographique. Par conséquent, nous nous attendons à ce qu'au début de l'apprentissage de la lecture, les variables linguistiques telles que la longueur et la fréquence

d'un mot, ou encore sa structure graphémique, aient une influence lors de la lecture de certains mots.

Ainsi, afin de pouvoir juger de la discontinuité ou de la continuité de l'apprentissage de la lecture, nous avons pris comme point d'analyse, la progression des élèves. Pour vérifier nos suppositions, nous avons effectué, pour chacune des variables, une analyse quantitative et une analyse qualitative. Nous avons considéré que, lorsque dans l'analyse quantitative la progression est régulière (la performance augmente progressivement ou les erreurs diminuent progressivement), il y a continuité. Toute rupture dans cette progression devient alors un critère de discontinuité. En ce qui concerne l'analyse qualitative, nous avons considéré qu'il y a continuité lorsqu'un type d'erreur est le même d'un moment à l'autre du cursus scolaire. Toute différence dans les types d'erreurs est considérée comme un critère de discontinuité.

Les apprentis lecteurs ne font pas que décoder

Les résultats font apparaître que le nombre d'erreurs commises sur des mots longs ou contenant des graphèmes complexes est plus important que sur les mots courts ou contenant des graphèmes simples. Ce premier constat semble donc rejoindre les études menées sur l'identification du mot en confirmant un effet de la longueur écrite ou de la structure graphémique d'un mot. Ce constat semble donc confirmer aussi la mise en place d'une procédure alphabétique au cours de laquelle les mots seraient déchiffrés. Il se pourrait donc que durant cette étape, les mots longs, comme les mots complexes, contenant davantage d'informations grapho-phonémiques, soient plus longs à traiter, ce qui entraînerait davantage de risques de commettre des erreurs de lecture, notamment lors de la mise en relation des graphèmes et des phonèmes.

De plus, conformément à nos suppositions de départ, on remarque que cet effet tend à diminuer entre le CP et le CM2. Ces résultats correspondraient alors avec la maîtrise de la phase alphabétique. Toutefois, bien que cette baisse soit constante d'une année sur l'autre, elle n'est pas pour autant régulière. On observe ainsi un écart plus prononcé entre les élèves de CP et de CE1 qui correspondrait à une rupture dans l'apprentissage de la lecture. Cette rupture, synonyme de discontinuité dans l'apprentissage comme l'envisagent les modèles développementaux, serait alors peut être le signe d'un changement de stratégie des élèves.

En effet, on remarque que face à des mots longs ou dont la structure graphémique est complexe, les lecteurs débutants adoptent essentiellement une procédure de décodage alors que les lecteurs plus avancés adoptent une procédure plus lexicale. Il convient de se demander

ici, si cette présence d'erreurs de décodage ne serait pas relative à une pédagogie dominante. En effet, l'enfant apprend comme on lui enseigne, il est donc possible que les résultats auraient été différents si ces élèves avaient été scolarisés dans des classes où la pédagogie est le langage entier. Si on revient à nos résultats, on remarque que le nombre d'erreurs liées à un mauvais décodage diminuent d'une année à l'autre, et en parallèle, on assiste à une augmentation des verbalisations de mots qui semble donc confirmer l'existence d'une procédure lexicale.

Nous avons remarqué que cette dernière est présente dès les premières années de l'apprentissage institutionnel de la lecture. Cela signifierait donc que, lors de l'automatisation de la procédure par médiation phonologique se mettrait en place une procédure lexicale. Autrement dit, les apprentis lecteurs ne feraient pas que décoder mais chercheraient aussi à donner du sens à leur lecture. Ce constat s'inscrit en faux si l'on en croit les modèles développementaux qui postulent que l'élève passe d'un stade à l'autre si, et seulement si les compétences du stade antérieur sont maîtrisées. Selon nos résultats, les élèves de notre échantillon se situeraient alors à la fois dans le stade alphabétique et orthographique.

Eviter de lire pour ne pas se tromper

Notre analyse des différents types de processus et de leur évolution confirme d'une façon plus générale ce constat. En effet, au début de l'apprentissage de la lecture, les élèves font essentiellement du décodage. Mais pas seulement puisqu'on observe dès le début des erreurs de type lexical et des abstentions. En effet, alors que certains élèves essaient de déchiffrer, d'autres essaient de donner du sens et d'autres encore ne prononcent rien. Ce constat rejoint celui de l'enquête de PISA (2009) qui constate une augmentation des taux de non réponses. Dans un pays, comme le nôtre où la culture de l'évaluation est forte, il se pourrait que pour ces élèves, il soit préférable de ne rien lire plutôt que de commettre une erreur.

Un apprentissage discontinu, oui mais pas absolu

Ces différents résultats semblent donc être en cohérence avec les modèles traditionnels de la lecture puisque les erreurs de décodage diminuent entre le CP et le CM2. Ceci laisserait donc penser à un apprentissage discontinu. Toutefois, l'apparition d'autres stratégies au début de l'apprentissage de la lecture invite à penser que cette discontinuité n'est pas absolue.

En effet, la mise en place en parallèle d'une procédure lexicale invite à penser que les enfants ont recours à plusieurs procédures au même moment. Ce constat va donc à l'encontre de la notion de successivité des modèles développementaux et nous amène à nous tourner vers l'idée que les processus joueraient alors un rôle plus interactif en lecture comme le prétend Seymour (1990). Toutefois, il semblerait qu'ici, dès le début de l'apprentissage de la lecture, les procédures alphabétiques et orthographiques soient disponibles contrairement au modèle de Seymour qui suppose que le lexique orthographique se développe par l'interactivité des procédures alphabétiques et phonologiques.

L'influence du test de L'alouette

Au cours de ces analyses, nous avons remarqué que, parmi les erreurs lexicales, plusieurs sont en lien avec le test de *L'alouette*. En effet, le test de Lefavrais n'est pas constitué de façon anodine. Le texte contient un certain nombre de pièges : les dessins, les similitudes grapho-phonémiques ou encore le co-texte sémantique. De ce fait, l'introduction de faux indices parmi les indices pertinents amènerait les élèves à commettre des erreurs. Rappelons que ce test, basé sur une conception traditionnelle de la lecture, mesure essentiellement le décodage. Ainsi, le fait de tomber dans ces pièges traduirait une défaillance dans le décodage. Autrement dit, on suppose que les élèves qui ne maîtriseraient pas totalement la procédure alphabétique se serviraient de ces indices pour pallier cette difficulté.

Par conséquent, nous avons voulu savoir dans quelle mesure les pièges tendus par Lefavrais ont un effet lors de la lecture du texte de *L'alouette* et plus précisément, dans quelle mesure ces erreurs permettent de rendre compte d'un apprentissage continu ou discontinu de la lecture.

Ainsi, dans une perspective développementale, nous nous attendions à observer une discontinuité dans l'apprentissage : les apprentis lecteurs devaient être plus nombreux que les lecteurs avancés à commettre des erreurs en lien avec le test de *L'alouette*. Normalement, la maîtrise du code devait faire diminuer ces types d'erreurs dans le temps.

Nous avons donc étudié successivement les effets liés aux dessins, aux similitudes et co-texte sémantique lors de la lecture du texte de *L'alouette*.

Les résultats semblent confirmer nos attentes puisqu'on observe une rupture dans la progression des élèves. Or, il apparaît une rupture certes, mais dans le sens inverse : alors que l'effet est moindre chez les apprentis lecteurs, il est plus important chez les lecteurs avancés.

La proportion d'élèves qui a commis ces types d'erreurs est, à chaque fois, plus élevée chez les lecteurs experts. Autrement dit, il semblerait que plus un élève avance dans l'apprentissage de la lecture et plus il aurait tendance à être influencé par les dessins, les similitudes ou le co-texte sémantique. Ces premiers résultats nous amènent à penser que le recours à ces variables ne serait pas seulement une stratégie compensatoire du traitement alphabétique (Stanovich, 1981). Ainsi, il se pourrait que cette discontinuité quantitative observée entre les lecteurs débutants et experts soit le reflet d'un changement de stratégie dans l'apprentissage de la lecture.

L'analyse qualitative des différents types d'erreurs, nous permet de confirmer ce caractère discontinu puisqu'on remarque que les stratégies adoptées semblent varier d'un moment à l'autre du cursus scolaire. En effet, alors que les apprentis lecteurs semblent s'attacher à une construction de sens au niveau du mot, les plus avancés semblent, quant à eux, s'attacher à une construction de sens portant sur des unités plus larges telles que le syntagme, voire la phrase. La présence de ce type d'erreur dès les premières années de l'apprentissage laisse penser qu'au début de l'apprentissage de la lecture, les apprentis lecteurs ne feraient pas que décoder mais chercheraient aussi à construire du sens. Ainsi, il semblerait que plus un élève avance dans l'apprentissage de la lecture et plus sa recherche de sens s'élargit : du sens lexical au sens syntagmatique voire phrastique. Ces résultats confirment donc la mise en place au même moment d'une procédure alphabétique et lexicale et remet donc en cause l'idée des modèles développementaux selon laquelle les élèves ne changeraient de stade que lorsque la procédure du stade antérieur est maîtrisée.

Ces analyses sur l'influence du test de Lefevrais, nous permettent donc de constater que le contexte et le co-texte jouent un rôle lors de l'identification des mots, notamment un rôle compensatoire pour les apprentis lecteurs. Ce constat remet donc en cause la conception des modèles développementaux qui ne tiennent pas compte du rôle du contexte dans l'apprentissage de la lecture. Identifier un mot ne serait donc pas seulement une mise en correspondance des graphèmes et des phonèmes. Ainsi, nos résultats nous renvoient vers Goodman (1967), pour qui le mécanisme de base de la lecture est la recherche de sens à l'aide du contexte.

Rappelons que le test de Lefavrais, conçu pour mesurer les déficiences en lecture, repose sur une conception classique de la lecture, c'est-à-dire qu'il mesure essentiellement le

niveau de décodage. Les pièges tendus par Lefavrais apparaissent donc ici comme des *artefacts* et l'élève qui tomberait dans ces pièges serait alors considéré comme un « mauvais lecteur ». Or nous venons de le voir, ceux qui tombent dans les pièges de ce test semblent être ceux qui ont un bon niveau de lecture. Ainsi, les mauvais lecteurs pour Lefavrais seraient en réalité les bons lecteurs, ceux qui cherchent à donner du sens à leur lecture.

Ainsi, comme il est difficile de juger du niveau en lecture pour les raisons méthodologiques énoncées précédemment, il apparaît aussi difficile de tirer une conclusion tranchée quant à l'apprentissage de lecture.

DISCUSSION

L'objectif de notre recherche était de voir ce qu'il en est de l'apprentissage de la lecture chez les élèves du CP au CM2. Compte tenu du renouveau des préconisations officielles concernant la place à faire au code dans l'enseignement de la lecture, le test de Lefavrais nous a semblé particulièrement intéressant non seulement pour effectuer un état des lieux du niveau en lecture mais aussi pour juger de l'apprentissage de la lecture.

Toutefois, lorsque que nous avons mené notre étude sur le niveau, nous nous sommes heurtée aux difficultés qu'il y a à réaliser une comparaison dans le temps. En effet, bien que nous disposions du nombre d'élèves qui ont servi à établir l'étalonnage en 1965, nous avons manqué d'informations quant au recueil des données. Ainsi, par exemple, le mois de passation n'est pas indiqué. Cette information est importante puisque, comme nous l'avons vu précédemment, elle peut entraîner des différences. De même, des indicateurs sociologiques tels que le milieu social d'appartenance n'est pas précisé. Nous supposons, qu'à cette époque l'échantillon a été constitué selon les commodités du chercheur sans tenir compte de variables sociologiques. Ce manque d'information, participe là aussi à relativiser nos résultats puisqu'il apparaît que le milieu social joue un rôle dans la réussite scolaire.

A l'origine, le test de *L'alouette* est censé mesurer les difficultés de lecture, à travers la capacité qu'à un élève à décoder. En effet, l'introduction de pièges devait permettre de voir quels sont les élèves qui rencontrent des difficultés lors de la mise en application de cette procédure. C'est en fait une méthode expérimentale qui crée des *artefacts* pour mesurer le niveau en lecture d'un élève.

Mais contrairement à ce qu'il prétend, ce test est un outil qui montre que la lecture est bien plus que du simple décodage et qu'elle est bien plus complexe que cela puisqu'au cours de nos analyses nous avons pu voir apparaître d'autres stratégies.

En effet, c'est un test calibré pour tester les stratégies compensatoires lors de la mise en place d'une procédure alphabétique. Or il apparaît que ce sont les lecteurs les plus avancés qui sont le plus influencés par ce test. Cela signifie que ce test ne démontre pas seulement que les apprentis lecteurs sont dépendants d'autres variables, il démontre aussi que les lecteurs experts sont sensibles à ces variables mais pas de la même façon. En effet, nous avons vu que la procédure lexicale a tendance à s'élargir avec le niveau de maîtrise de lecture : du sens

local pour les lecteurs débutants (au niveau du mot) à un sens plus large pour les lecteurs experts (au niveau du syntagme voir de la phrase).

Ainsi, le postulat de Lefavrais selon lequel les bons lecteurs sont ceux qui ne se laissent pas induire en erreur et à l'inverse, les mauvais lecteurs sont ceux qui tombent dans les pièges, semble être en incohérence avec ce que démontre le test de *L'alouette*. En effet, d'après les résultats de notre recherche, les bons lecteurs sont ceux qui tombent dans les pièges. Nous pourrions ainsi dire que ceux qui réussissent au test sont ceux qui sont « malins », qui cherchent à donner du sens à leur lecture.

Ce constat rejoint donc celui de Foucambert (1983) selon lequel dans le test de *L'alouette* « *Ce n'est pas le déchiffrage qui est ici mesuré. [...] Ce qui est mesuré c'est la non-lecture* ».

PERSPECTIVES DE RECHERCHE

Comme nous l'avons remarqué plus haut, l'avantage de ce test est qu'il permet d'étudier l'identification de mots dans des conditions textuelles de lecture et non pas expérimentales.

Toutefois, cet avantage peut aussi être un inconvénient puisque les variables testées ne sont pas isolées. Dans le test de Lefavrais, les intentions lexicales sont provoquées soit par les dessins, les similitudes ou le co-texte sémantique. Les résultats mettant en avant l'utilisation de la voie lexicale au début de l'apprentissage de la lecture auraient peut-être été différents si nous avions utilisés un autre test. Ainsi, pour vérifier l'existence de la procédure lexicale au début de l'apprentissage de la lecture, il faudrait prévoir un plan expérimental permettant d'isoler ces variables. Par exemple, pour vérifier l'influence des dessins, on aurait deux groupes d'élèves de niveau équivalent. On ferait lire à un groupe un texte avec des illustrations et on donnerait à l'autre groupe le même texte mais sans les illustrations. De la même façon, pour l'influence des similitudes, on ferait lire à un groupe le texte avec des mots qui se ressemblent graphiquement et phonémiquement, et on donnerait au second groupe le même texte en ayant pris soin de changer les mots qui présentent des similitudes.

Par ailleurs, il serait aussi intéressant de poursuivre l'analyse des autres caractéristiques du test de *L'alouette* comme par exemple la forme du texte. En effet, l'aspect du texte n'est pas commun : les paragraphes ne sont pas justifiés ce qui pourrait entraîner une difficulté

supplémentaire de lecture. Lors de nos passations, nous avons remarqué que plusieurs élèves (et notamment les élèves les plus âgés) n'avaient pas lu certaines lignes qui se trouvent être souvent les mêmes. On peut alors se demander si ces sauts de lignes sont en lien avec la forme visuelle du texte.

Enfin, le test de *L'alouette* prévoit un questionnaire destiné à l'enseignant afin de connaître ses pratiques pédagogiques. Lors de nos passations, nous en avons récolté un certain nombre qui pourrait faire l'objet d'une recherche complémentaire à celle que nous avons menée.

BIBLIOGRAPHIE

Baudelot, C. et Establet, R. (1989). *Le niveau monte : réfutation d'une vieille idée concernant la prétendue décadence de nos écoles*. Paris, France : Editions du Seuil.

Broccolichi, S., Ben-Ayed, C., Mathey-Pierre, C. & Trancart, D. (2007). Fragmentations territoriales et inégalités scolaires : des relations complexes entre la distribution spatiale, les conditions de scolarisation et des réussites des élèves. *Education et Formations*, n°74, p.31-48.

Capel, F. et Dubet, F (2009). L'avenir de nos enfants en question(s), *Le niveau baisse-t-il vraiment ?* Paris, France : Magnard.

CÉARD M-T, RÉMOND M. et VARIER M., (2003). « L'appréciation des compétences des élèves et des jeunes en lecture et en écriture et l'évolution de ces compétences dans le temps ». HCE. Paris, France.

Cédelle, L. (2007). Orthographe : les collégiens de cinquième sont tombés au niveau des élèves de CM2 de 1987. *Le Monde*. Récupéré du site de la revue : <http://www.lemonde.fr/>

Cédelle L., (2008). Une nouvelle enquête atteste de la baisse du niveau des élèves en fin de CM2. *Le Monde*. Récupéré du site de la revue : <http://www.lemonde.fr/>

Chartier, A. M. (2007). *L'école et la lecture obligatoire : Histoires et paradoxes des pratiques d'enseignement de la lecture*. Mercues, France : Retz.

Chartier, A. M. et Hébrard J. (2000). *Discours sur la lecture 1880-2000*. Paris, France : Fayard.

Chauveau, G. (2004). *Comprendre l'enfant apprenti lecteur : recherches actuelles en psychologie de l'écrit* (dir.). Paris, France : Retz. Collectif.

Chauveau, G (2007). *Le savoir-lire aujourd'hui*. Paris, France : Retz.

Chauveau, G. (2011). *Comment l'enfant devient lecteur. Pour une psychologie culturelle de la lecture*. Paris, France : Retz.

Chervel, A. et Manesse, D. (1989). *La dictée : les français et l'orthographe (1873-1978)*, INRP/Calmann-Lévy, Paris, France.

Chervel, A. (1995). *L'enseignement du français à l'école primaire. Textes officiels Tome 3 : 1940-1995*. Paris, France : INRP et Economica.

Colmant, M. et Dos Santos, S. (2008). Évolution des performances en lecture des élèves de CM1 : Résultats de l'étude internationale PIRLS. *Note d'Information*, pp. 1-6. Paris : DEPP.

De F. Barbier et C. Bertho Lavenir « *Histoire des médias, de Diderot à Internet* » cité par A-M Chartier et J. Hébrard dans « *Discours sur la lecture (1880-2000)* »

Content A. Leybaert L. (1992). L'acquisition de la lecture: influence des méthodes d'apprentissage. In Lecocq P. (Ed.), *La lecture, processus, apprentissage, troubles* (pp181-211). Lille: Presses Universitaires

Cour des Comptes. (2010). « *L'Education nationale face à l'objectif de la réussite de tous les élèves* ». Récupéré du site <http://www.ladocumentationfrancaise.fr>

Don, C. (1982). *The process of reading. A cognitive analysis of fluent reading and learning to read*.

Duru-Bellat, M. (1994). Filles et garçons à l'école, approches sociologiques et psychosociales. In: *Revue française de pédagogie*.

Volume 109, 1994. pp. 111-141.

doi : 10.3406/rfp.1994.1250

Ecalte, J. et Magnan, A. (2002). *L'apprentissage de la lecture : fonctionnement et développement cognitifs*. Paris, France : Armand Colin.

Elley W.B (1992). *“How in the world do students read?”*

Fayol, M., Gombert, J.E., Lecocq, P., Sprenger-Charolles, L., & Zagar, D. (1992). *Psychologie cognitive de la lecture*. Paris, France : PUF.

Fayol, M. et Gombert, J-E. (1999). L'apprentissage de la lecture et de l'écriture. Dans J. A. Rondal & E. Esperet (Eds), *Manuel de Psychologie de l'enfant* (pp. 565-594). Hayen : Mardaga.

Fijalkow, J. (1986). *Mauvais lecteurs pourquoi ?* Paris, France : PUF.

Fijalkow, E. (2003). *L'enseignement de la lecture-écriture au cours préparatoire : entre tradition et innovation*. Paris, France : L'Harmattan.

Fijalkow, J. (1993). *Entrer dans l'écrit*, Edition Magnard, Tournai, France.

Fijalkow, J. et Fijalkow, E. (2003). *La lecture : idées reçues*. Paris, France : Le cavalier bleu.

Fijalkow, J. & Downing, J. (1984). *Lire et raisonner*. Toulouse, France : Privat.

Fédération Nationale des Orthophonistes (2004). *Le Langage écrit. Troubles spécifiques, Dossier de Presse*, pp. 1-20. Paris : FNO.

Fijalkow, J., Pasa, L. & Creuzet V. (2006). Écriture inventée : pluralité des traitements et variabilité selon la structure syllabique. *Education et Francophonie*, vol. XXXIV : 2, pp. 85-100. Montréal, Canada : ACELF.

Fijalkow, J., Pasa, L. & Ragano, S. (1998). Les difficultés en lecture orale : perspective socioconstructiviste. *Repères*, n°18, pp.173-192. Paris : INRP.

Foucambert D., (1998), Les actes de lecture, *La revue de L'association Française pour la Lecture*, n°62.

Frith, U. (1986). *Annals of dyslexia. A developmental framework for developmental dyslexia.* Vol. 36, pp.69-81.

Gak, V.G. (1976). *L'orthographe du français : essai de description théorique et pratique.* Paris, France : SELAF.

Huteau, M. (2001). *Les figures de l'intelligence. Editions et applications psychologiques,* Paris.

Laronche M., (2008), *Pourquoi les filles battent les garçons à l'école ?*, Le Monde, Paris.

Laronche M. (2008), *Les filles têtes de classe en Europe*, Le Monde, Paris

Lecocq, P. (1992). *La lecture : processus apprentissage, troubles.* Lille, France : PUF.

Manesse D., Cogis D., Dorgans M. & Taller C., (2007), *Orthographe à qui la faute ?*, Les Cahiers Pédagogiques, ESF, Paris

Ministère de l'Education Nationale. (2011). *L'état de l'école 2011.* Récupéré du site <http://www.education.gouv.fr>

Ministère de l'Education Nationale (2006). *Qu'apprend-on à l'école élémentaire ? 2006-2007, les programmes.* CNDP/XO Editions.

Observatoire National de la Lecture, Inspection générale de l'éducation nationale-Groupe de l'enseignement primaire (2005), *L'apprentissage de la lecture à l'école primaire, Rapport - n° 2005-123.*

Observatoire National de la Lecture (1996). *Regards sur la lecture et ses apprentissages.* Paris : MEN-ESR.

Repères et références statistiques, (1999-2000), *Enquête dans les écoles publiques et privées de l'enseignement préélémentaire et élémentaire*, édition 2007, Tableaux statistiques n°6709.

Perfetti, C. A., and Roth, S. (1981) Some of the interactive processes in reading and their role in reading skill. In A. Lesgold and C. Perfetti (eds), *Interactive Processes in Reading*, Hillsdale, NJ: Erlbaum Associates, pp. 269-97.

Rieben L. & Perfetti Ch. (1989). *L'apprenti lecteur. Recherches empiriques et implications pédagogiques*. Suisse, Lausanne

Romdhane, M. N., Gombert J.E. et Belajouza M. (2003) *L'apprentissage de la lecture. Perspectives comparatives*.

Rosenwald F., *Les filles et les garçons dans le système éducatif*, (2006), DEPP, Note d'information 06.06 février

Seibel C. (1984). Genèses et conséquences de l'échec scolaire. *Revue française de pédagogie*. Volume 67. pp. 7-28. doi : 10.3406/rfp.1984.1571

Sprenger-Charolles, L. et Casalis, S. (1996). *Lecture et écriture : acquisition et troubles du développement*. Paris, France : PUF.

Van Hout, A. & Estienne, F. (2001). *Les dyslexies : décrire, évaluer, expliquer, traiter*. Paris : Masson.

Sites internet

<http://www.education.gouv.fr>

<http://www.onl.inrp.fr>

<http://www.oecd.org/pisa/pisaenfranglais.htm>

<http://www.legifrance.gouv.fr>

Rapports

Haut Conseil de l'École. (2007). L'école primaire. Bilan des résultats de l'école 2007.

DEPP. (2001). « *Les élèves de CM1, premiers résultats d'une évaluation internationale en lecture (PIRLS)* », Note d'Information 03-22.

DEPP. (2003). « *Compréhension à l'écrit et à l'oral des élèves en fin d'école primaire* »,