



HAL
open science

Description linguistique et évaluation des capacités orthographiques d'adultes en difficulté avec l'écrit.

Carole Blondel

► **To cite this version:**

Carole Blondel. Description linguistique et évaluation des capacités orthographiques d'adultes en difficulté avec l'écrit.. Linguistique. Université Grenoble Alpes [2020-..], 2020. Français. NNT : 2020GRALL001 . tel-02632800

HAL Id: tel-02632800

<https://theses.hal.science/tel-02632800v1>

Submitted on 27 May 2020

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Sciences du langage Spécialité Didactique et Linguistique

Arrêté ministériel : 25 mai 2016

Présentée par

CAROLE BLONDEL

Thèse dirigée par **Catherine BRISSAUD**, Université Grenoble Alpes

préparée au sein du **Laboratoire Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles** dans l'**École Doctorale Langues, Littératures et Sciences Humaines**

Description linguistique et évaluation des capacités orthographiques d'adultes en difficulté avec l'écrit.

Linguistic description and assessment of writing abilities of adults in trouble with the writing.

Thèse soutenue publiquement le **17 janvier 2020**, devant le jury composé de :

Madame Catherine BRISSAUD

Professeur, Université Grenoble Alpes, Directeur de thèse

Monsieur JEAN-PIERRE JEANTHEAU

Directeur d'études, AGENCE NATIONALE DE LUTTE CONTRE L'ILLETTRISME, Examineur

Madame CLARA MORTAMET

Professeur, Université Jean Monnet, Président

Monsieur JEAN-CHRISTOPHE PELLAT

Professeur Emérite, Université de Strasbourg, Rapporteur

Madame CLAUDINE GARCIA-DEBANC

Professeur, Université de Toulouse - Jean Jaurès, Rapporteur

Madame FANNY RINCK

Maître de Conférences, UNIVERSITE GRENOBLE ALPES, Examineur





REMERCIEMENTS

Mes premiers remerciements vont à Catherine Brissaud et à Fanny Rinck, qui m'ont fait confiance quant à la réalisation de ce travail.

Je remercie Catherine Brissaud, qui a été une directrice de thèse exemplaire. Disponible, patiente, je la remercie en particulier pour avoir su trouver les mots justes et pour avoir su s'effacer quand il le fallait.

Merci également à Fanny Rinck pour la qualité de son encadrement, pour ses critiques toujours constructives et pour ses encouragements.

Ma gratitude va également aux membres du jury qui me font l'honneur d'évaluer ce travail : Claudine Garcia-Debanc, Jean-Christophe Pellat, Jean-Pierre Jeantheau et Clara Mortamet.

Je remercie la région Rhône-Alpes, qui a offert un financement permettant de mener à bien ce doctorat, ainsi que la Délégation Générale à la Langue française et aux langues de France pour avoir financé le projet « *Compétences orthographiques des adultes et contextes sociaux d'usage* ». Ce projet m'a permis de réaliser des enquêtes sur le terrain et a enrichi ce travail, tant humainement que pour les données recueillies. Merci en particulier à Jean-Pierre Jeantheau, qui a été à l'initiative de ce projet.

Je souhaite également remercier Clara Mortamet pour nous avoir permis de travailler conjointement avec le laboratoire Dylis (Université de Rouen) sur le corpus de l'INSEE et pour avoir facilité le rapprochement avec sa doctorante, Jeanne Conseil, avec laquelle nous avons eu d'intéressantes discussions sur les graphies.

Un immense merci aux copines de thèse : Armelle pour son soutien indéfectible et pour nos longues conversations ; Lucie et Claire pour avoir été présentes lors de ces dernières semaines bien éprouvantes et pour avoir assuré, presque nuit et jour, un soutien moral et technique dont je leur serai éternellement reconnaissante.

Plus personnellement, à mes amies, pour leur soutien et leur compréhension en cette période de « grande absence » : Audrey de l'autre bout de l'Atlantique, ma petite sœur de cœur ; Audrey mon amie d'enfance qui croit que je rédige un mémoire depuis toutes ces années.

Un dernier mot à mes parents, pour avoir été des baby-sitters hors pairs et pour leur dire à quel point leur soutien sur ces dernières semaines a été précieux. Même s'il leur faudra attendre la soutenance pour qu'ils comprennent « enfin » ce que je fais, ils ont largement contribué à ce que ce travail soit arrivé à son terme.

Enfin, à Lilou et à Éric, tout simplement.

SOMMAIRE

REMERCIEMENTS	3
SOMMAIRE.....	5
INTRODUCTION	13
CADRE THÉORIQUE.....	19
CHAPITRE 1. LA LITTÉRACIE, UN CADRE POUR REPENSER L'ILLETTRISME ET LES COMPÉTENCES SCRIPTURALES DES ADULTES.....	21
1. LA LITTÉRACIE, UN CADRE POUR REPENSER L'ILLETTRISME.....	24
1.1. L'ILLETTRISME AUJOURD'HUI : UNE DÉFINITION IMPOSSIBLE ?.....	24
1.1.1. BREF RAPPEL HISTORIQUE	24
1.1.2. QUELQUES CHIFFRES ISSUS DE DIFFÉRENTES ENQUÊTES EN FRANCE	25
1.1.3. UNE DÉFINITION QUI TÉMOIGNE D'UNE RÉALITÉ COMPLEXE	27
1.1.4. « LES ILLETTRISMES » : UNE MEILLEURE CONNAISSANCE DES PUBLICS AUJOURD'HUI	28
1.2. LES ATTENTES DE L'ANLCI EN MATIÈRE DE POLITIQUE PUBLIQUE	31
1.2.1. LE CADRE NATIONAL DE RÉFÉRENCE DE L'ANLCI	31
1.2.2. LES RÉFÉRENTIELS DES COMPÉTENCES DE L'ANLCI	33
1.3. LA LITTÉRACIE : UNE APPROCHE RÉCENTE DES COMPÉTENCES À L'ÉCRIT	37
1.3.1. L'ORTHOGRAPHE DES ADULTES : UN CONFLIT ENTRE COMPÉTENCES ET ATTENTES SOCIALES	38
1.3.2. LES ENSEIGNANTS ET LA NORME ORTHOGRAPHIQUE.....	43
1.4. CONCLUSION SUR L'ÉCRITURE DES ADULTES DANS LE CONTEXTE DE LA LITTÉRACIE	45
2. L'ÉVALUATION DES ADULTES À L'ÉCRIT ET PLUS PARTICULIÈREMENT EN ÉCRITURE	46
2.1. TOUR D'HORIZON DES ENQUÊTES SUR LA LITTÉRACIE RÉALISÉES AUPRÈS D'ADULTES ET DE JEUNES ADULTES.....	48
2.1.1. DES ENQUÊTES À GRANDE ÉCHELLE	49
2.1.2. LES ÉVALUATIONS AUPRÈS DE PUBLICS PARTICULIERS	51
2.1.3. LA PLACE DE L'ÉCRITURE ET DE L'ORTHOGRAPHE DANS LES ENQUÊTES	59
2.2. L'ENQUÊTE INFORMATION ET VIE QUOTIDIENNE (INSEE, 2002, 2004, 2011)	64
2.2.1. LE PROTOCOLE D'IVQ : ÉVALUER TOUS LES NIVEAUX DE COMPÉTENCES.....	65
2.2.2. FOCUS SUR LES COMPÉTENCES TESTÉES DANS LE MODULE « BAS »	72
2.2.3. CONCLUSIONS DE L'ENQUÊTE ET PRINCIPAUX RÉSULTATS.....	76

CHAPITRE 2. L'ORTHOGRAPHE DU FRANÇAIS ET SES DIFFICULTÉS	84
1. L'ORTHOGRAPHE : UN PROBLÈME DE SOCIÉTÉ	86
1.1. BREF APERÇU HISTORIQUE	87
1.2. L'APPRENTISSAGE DE L'ORTHOGRAPHE À L'ÉCOLE : UNE APPROCHE NORMATIVE.....	89
1.3. LA QUESTION D'UNE RÉFORME	94
1.3.1. LES ENJEUX ET DIFFICULTÉS D'UNE RÉFORME	95
1.3.2. UNE ORTHOGRAPHE PLUS « ÉCONOMIQUE » ?.....	96
1.3.3. UN ATTACHEMENT PARFOIS IRRATIONNEL À L'ORTHOGRAPHE	97
1.4. BILAN SUR L'ORTHOGRAPHE ET LA SOCIÉTÉ.....	98
2. PHONOGRAPHIE ET SÉMIOGRAPHIE	99
2.1. DE L'ORAL À L'ÉCRIT, DE L'ÉCRIT À L'ORAL.....	101
2.2. L'ÉCRITURE DU FRANÇAIS, UN SYSTÈME MIXTE	102
2.2.1. LE PRINCIPE PHONOGRAPHIQUE	102
2.2.2. ENTRE PHONOGRAPHIE ET SÉMIOGRAPHIE : LA NOTION DE MIXITÉ	104
3. LES GRANDS PRINCIPES DU SYSTÈME ORTHOGRAPHIQUE DU FRANÇAIS : ANALYSE DES GRAPHÈMES 105	
3.1. LE « PLURISYSTÈME » DU FRANÇAIS : GRAPHÈMES, SOUS-GRAPHÈMES ET ARCHIGRAPHÈMES (CATACH, 1980).....	108
3.1.1. CATÉGORIES DE GRAPHÈMES.....	109
3.1.2. L'ARCHIGRAPHÈME	111
3.1.3. HIÉRARCHISATION DES GRAPHÈMES SELON CATACH (1995) : L'IMPORTANCE DE LA FRÉQUENCE.....	111
3.2. L'ORGANISATION DES GRAPHÈMES SELON BLANCHE-BENVENISTE ET CHERVEL (1969).....	114
3.2.1. LES VALEURS DU GRAPHÈME SELON BLANCHE-BENVENISTE ET CHERVEL (1969).....	115
3.2.2. LES DIGRAMMES	117
3.2.3. LA SYLLABE GRAPHIQUE ET LA VALEUR ZÉRO	118
3.3. LE CHOIX DES SCRIPTEURS : ENTRE RÉGULARITÉS ET IRRÉGULARITÉS.....	119
4. ZOOM SUR UNE DIFFICULTÉ EMBLÉMATIQUE DU FRANÇAIS : LES MORPHOGRAMMES	122
4.1. INTRODUCTION SUR LES MORPHOGRAMMES	123
4.2. LES LETTRES MUETTES ET LE PRINCIPE DE DÉRIVATION	126
4.2.1. DÉRIVATION SUFFIXALE	127
4.2.2. DÉRIVATION SUR LE PRÉFIXE	128
4.2.3. BILAN SUR LES LETTRES MUETTES ET LA DÉRIVATION.....	129

4.3. LA PRONONCIATION DES FINALES ET LA QUESTION DU « -E » MUET FINAL	130
5. BILAN SUR LES DIFFICULTÉS ORTHOGRAPHIQUES DU FRANÇAIS.....	131
CHAPITRE 3. APPORTS DE LA PSYCHOLINGUISTIQUE SUR L'APPRENTISSAGE DE L'ORTHOGRAPHE DU FRANÇAIS.....	134
1. RENDRE COMPTE DE LA DIFFICULTÉ ORTHOGRAPHIQUE DES MOTS : LES NOTIONS DE CONSISTANCE ET DE FRÉQUENCE.....	136
1.1 L'INDICE DE CONSISTANCE DANS LA BASE MANULEX-INFRA.....	136
1.1.1 LA CONSISTANCE ORTHOGRAPHIQUE : UN INDICE DE LA COMPLEXITÉ ORTHOGRAPHIQUE.....	136
1.1.2 LA BASE DE DONNÉES MANULEX : PRÉSENTATION ET FONCTIONNALITÉS	137
1.2 LA FRÉQUENCE ORTHOGRAPHIQUE : UNE VARIABLE PRIMORDIALE DE LA COMPÉTENCE ORTHOGRAPHIQUE.....	140
1.2.1 LE CALCUL DE L'INDICE DE FRÉQUENCE DANS MANULEX-INFRA.....	141
1.2.2 LA BASE DE DONNÉES LEXIQUE 3 : BRÈVE PRÉSENTATION	142
1.3 L'ACQUISITION DU LEXIQUE.....	144
1.3.1 LES PROCÉDURES MISES EN ŒUVRE DANS L'APPRENTISSAGE DU LEXIQUE	144
1.3.2 L'IMPACT DE LA CONSISTANCE SUR LES COMPÉTENCES EN ORTHOGRAPHE LEXICALE	146
1.3.3 L'ACQUISITION DE L'ORTHOGRAPHE LEXICALE CHEZ LES ENFANTS	148
1.4 BILAN SUR L'APPRENTISSAGE STATISTIQUE DE L'ÉCRIT	150
2. L'ORTHOGRAPHE DU FRANÇAIS DU POINT DE VUE DE LA PSYCHOLINGUISTIQUE.....	152
2.1. LA MAITRISE DE LA MORPHOLOGIE FLEXIONNELLE, UNE COMPÉTENCE COMPLEXE	153
2.1.1. LA MISE EN PLACE DES PLACE DES COMPÉTENCES AVANT LE PASSAGE À L'ÉCRIT	153
2.1.2. LA MISE EN PLACE DES COMPÉTENCES AVEC LE PASSAGE À L'ÉCRIT	154
2.2. LA MORPHOLOGIE DÉRIVATIONNELLE ET L'ACQUISITION DES FINALES MUETTES	160
2.3. CONCLUSION SUR LA GESTION DE LA MORPHOLOGIE.....	163
OBJECTIFS DE LA RECHERCHE PROBLÉMATIQUE.....	166
CHAPITRE 4. PRÉSENTATION DES ENQUÊTES ET DES CHOIX MÉTHODOLOGIQUES	170
1. PRÉSENTATION DE LA DICTÉE D'IVQ : CHOIX DES ITEMS ET PERSPECTIVES D'ANALYSES	173
1.1. LE CHOIX DES ITEMS	173
1.2. NOTRE APPROCHE DE LA DICTÉE D'IVQ	176
1.2.1. LES MOTS DE LA DICTÉE D'IVQ PRÉSENTÉS SELON LEUR INDICE DE CONSISTANCE ET DE FRÉQUENCE.....	177

1.2.2.	LES MOTS DE LA DICTÉE D'IVQ PRÉSENTÉS SELON LEUR SCORE DANS EOLE	183
1.2.3.	LE CHOIX DES MOTS : BILAN	185
1.3.	LES PSEUDO-MOTS DE LA DICTÉE D'IVQ	185
2.	TRAITEMENT DES DICTÉES.....	186
2.1.	CRITÈRES ET ATTENTES DE L'INSEE SUR LA DICTÉE : LE CALCUL DES SCORES	186
2.2.	UN NÉCESSAIRE RECODAGE DU CORPUS POUR UNE ANALYSE DIFFÉRENTE	187
2.2.1.	LE FICHER DES RÉPONSES : UNE PERTE D'INFORMATION POUR OBSERVER LA VARIÉTÉ DES GRAPHIES	187
2.2.2.	VÉRIFICATION DE LA SAISIE, RECODAGE ET CHOIX MÉTHODOLOGIQUES.....	191
2.2.3.	RÉFLEXION MÉTHODOLOGIQUE SUR LA CONSTITUTION D'UN BARÈME	193
3.	ÉVALUER LES CAPACITÉS D'ÉCRITURE D'ADULTES EN DIFFICULTÉ À PARTIR D'UNE DICTÉE : LA NÉCESSITÉ D'UNE ANALYSE PARTICULIÈRE.....	195
3.1.	LE CRITÈRE ORTHOGRAPHIQUE APPLIQUÉ AUX MOTS	195
3.1.1.	UN SCORE POUR L'ORTHOGRAPHE LEXICALE.....	195
3.1.2.	LA PRISE EN COMPTE DE LA MORPHOGRAPHIE	196
3.2.	LE CRITÈRE PHONOGRAPHIQUE APPLIQUÉ AUX MOTS.....	198
3.2.1.	LES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES	198
3.2.2.	LES GRAPHIES APPROXIMANTES	201
3.2.3.	RÉCAPITULATIF : LE CODAGE DES GRAPHIES QUI S'ÉCARTENT DE LA NORME	203
3.3.	LE CRITÈRE PHONOGRAPHIQUE APPLIQUÉ AUX PSEUDO-MOTS	203
4.	LE PROJET « COMPÉTENCES ORTHOGRAPHIQUES DES ADULTES ET CONTEXTES SOCIAUX D'USAGE » 204	
4.1.	OBJECTIFS ET INTÉRÊT DU PROJET POUR L'ANALYSE DE LA DICTÉE D'IVQ	204
4.2.	UNE ENQUÊTE DE TERRAIN.....	206
4.2.1.	MISE EN PLACE DE L'ENQUÊTE ET BILAN	206
4.2.2.	LE PROTOCOLE DE L'ENQUÊTE COMPLÉMENTAIRE	207
4.3.	LE TRAITEMENT DES DONNÉES DE L'ENQUÊTE COMPLÉMENTAIRE	219
4.3.1.	CALCUL DES SCORES SUR LE « RAYON BRICOLAGE »	219
4.3.2.	LA TRANSCRIPTION DES ENTRETIENS.....	220
5.	SÉLECTION DU CORPUS POUR NOS ANALYSES	229
	PRÉSENTATION DES RÉSULTATS ET ANALYSES	232
	CHAPITRE 5. ANALYSE DE 441 DICTÉES DE SCRIPTEURS EN DIFFICULTÉ DANS IVQ 2011	234

1. PRÉSENTATION DU CORPUS « BAS »	236
2. RÉSULTATS DE LA DICTÉE D'IVQ	239
2.1. RÉPARTITION DES SCORES	242
2.1.1. LES DICTÉES QUI ONT UN SCORE DE « 0 ».....	244
2.1.2. RETOUR SUR LA CLASSIFICATION DES ENQUÊTÉS À L'ISSUE DE LA DICTÉE	247
2.2. COMPÉTENCES EN ORTHOGRAPHE LEXICALE : DES ÉCARTS IMPORTANTS D'UN MOT À L'AUTRE 248	
2.2.1. CLASSEMENT DES MOTS DE LA DICTÉE D'IVQ CHEZ LES ENQUÊTÉS « BAS »	249
2.2.2. COMPARAISON DES DIFFICULTÉS ENTRE LES ENQUÊTÉS « BAS » ET DES ÉLÈVES DE CM2 251	
2.3. COMPÉTENCES EN PHONOGRAPHIE OBSERVÉES SUR LES PSEUDO-MOTS ET SUR LES MOTS .253	
2.3.1. COMPÉTENCES EN PHONOGRAPHIE SUR LES PSEUDO-MOTS	254
2.3.2. COMPÉTENCES EN PHONOGRAPHIE SUR LES MOTS.....	270
2.3.3. BILAN SUR LA PHONOGRAPHIE	282
2.4. COMPÉTENCES EN MORPHOGRAPHIE OBSERVÉES SUR LES MOTS.....	285
2.4.1. SCORES DE RÉUSSITE EN MORPHOGRAPHIE FLEXIONNELLE DU NOMBRE	285
2.4.2. LES AUTRES TRACES DE MORPHOGRAPHIE OBSERVÉES DANS LES DICTÉES	288
2.4.3. LES MARQUES DE NOMBRE IMPLICITES	288
2.4.4. ERREURS MORPHOGRAPHIQUES RELEVÉES DANS « LE FIANCÉ SOLENNEL »	290
2.4.5. BILAN SUR LA MORPHOGRAPHIE	297
2.5. ANALYSE DE QUELQUES MOTS CHOISIS POUR LEUR COMPLEXITÉ ORTHOGRAPHIQUE.....	298
2.5.1. SIX MOTS CHOISIS PAR RAPPORT AUX CRITÈRES DE FRÉQUENCE ET DE CONSISTANCE 298	
2.5.2. LE CAS DU « -E » FINAL.....	310
2.6. CONCLUSION GÉNÉRALE SUR LES 447 ENQUÊTÉS « BAS ».....	319
2.6.1. LES DIFFÉRENTS SCORES CHEZ LES ENQUÊTÉS « BAS »	319
2.6.2. DES DIFFÉRENCES INTER-RÉGIONALES DANS LES COMPÉTENCES ORTHOGRAPHIQUES ? 322	
 CHAPITRE 6. COMPÉTENCES ORTHOGRAPHIQUES ET PRATIQUES DE L'ÉCRIT CHEZ LES ADULTES : APPORT DE DEUX ENQUÊTES	326
1. COMPARAISON DES DICTÉES PRODUITES PAR LES ENQUÊTÉS « BAS » ET « HAUT »	330
1.1. RÉPARTITION ET COMPARAISON DES SCORES.....	330
1.2. COMPARAISON DES COMPÉTENCES EN ORTHOGRAPHE LEXICALE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT »	332

1.3. COMPARAISON DES COMPÉTENCES EN PHONOGRAPHIE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT »	334
1.3.1. COMPÉTENCES EN PHONOGRAPHIE SUR LES PSEUDO-MOTS	334
1.3.2. COMPÉTENCES PHONOGRAPHIQUES SUR LES MOTS CHEZ LES ENQUÊTÉS « HAUT »	336
1.4. COMPARAISON DES COMPÉTENCES EN MORPHOGRAPHIE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT »	340
1.4.1. SCORES DE RÉUSSITE.....	340
1.4.2. LES AUTRES TRACES DE MORPHOGRAPHIE OBSERVÉES DANS LES DICTÉES	341
1.4.3. ERREURS MORPHOGRAPHIQUES RELEVÉES DANS « LE FIANCÉ SOLENNEL »	343
1.5. PRÉSENTATION DE LA VARIATION SUR LES MOTS ET LES PSEUDO-MOTS CHEZ LES ENQUÊTÉS « BAS » et « HAUT »	344
1.5.1. VARIATION SUR LES MOTS CHEZ LES ENQUÊTÉS « BAS » et « HAUT »	345
1.5.2. VARIATION SUR LES PSEUDO-MOTS CHEZ LES ENQUÊTÉS « BAS » et « HAUT »	346
1.5.3. BILAN : APPORT DE L'ENQUÊTE COMPLÉMENTAIRE POUR MIEUX CERNER LES COMPÉTENCES ORTHOGRAPHIQUES DE SCRIPTEURS EN DIFFICULTÉ	354
2. COMPÉTENCES ORTHOGRAPHIQUES, PRATIQUES D'ÉCRITURE ET REPRÉSENTATIONS DE L'ÉCRIT DE 87 SCRIPTEURS	356
2.1. RÉSULTATS GÉNÉRAUX DU « RAYON BRICOLAGE ».....	356
2.1.1. RÉPARTITION DES SCORES	356
2.1.2. COMPÉTENCES EN ORTHOGRAPHE LEXICALE SUR LES MOTS DU « RAYON BRICOLAGE »	358
2.1.3. COMPÉTENCES EN MORPHOGRAPHIE OBSERVÉES SUR L'INTÉGRALITÉ DE LA DICTÉE	360
2.2. PRATIQUES D'ÉCRITURE ET REPRÉSENTATIONS DE L'ÉCRIT : UNE HÉTÉROGÉNÉITÉ DE PROFILS ET DE PRATIQUES	364
2.2.1. RAPPORT À L'ÉCRIT ET USAGE DE L'ÉCRITURE	365
2.2.2. LIEN ENTRE PERFORMANCES ET DISCOURS : ENTRE CONTRADICTIONS ET RÉALITÉ	366
2.2.3. BILAN : DES DISCOURS ET DES PRATIQUES CONTRADICTOIRES	375
CONCLUSION.....	378
TABLE DES FIGURES	400
TABLE DES TABLEAUX	402
BIBLIOGRAPHIE.....	406
ANNEXES.....	419
Annexe 1 : Dictée produite par l'enquêté « bas » n° 1612	420
Annexe 2 : Dictée produite par l'enquêté « haut » n° 187	421

Annexe 3 : Classement des graphies des enquêtés « bas » et « haut » relevées dans la dictée d'IVQ 422	
Annexe 3 bis : Classement des graphies des enquêtés « haut » relevées dans le « rayon bricolage » 463	
Annexe 4 : Projet <i>Compétences orthographiques des adultes et contextes sociaux d'usage</i>469	
Annexe 5 : Avis d'examen du Comité du label de la Statistique Publique478	
Annexe 6 : Avis d'examen du Conseil National du label de l'Information Statistique.....481	
Annexe 7 : Courrier envoyé aux enquêtés « haut »483	
Annexe 8 : Protocole de l'enquête complémentaire, consignes aux enquêteurs.....484	
Annexe 9 : Transcription de l'entretien de l'enquêté « haut » n° 187.....492	
Annexe 10 : Dictée produite par l'enquêté « bas » n° 1960.....495	
Annexe 11 : Dictée produite par l'enquêté « bas » n° 2015.....496	

INTRODUCTION

La société d'aujourd'hui, avec la multiplication des supports d'information et de communication qui la caractérise, demande aux scripteurs de maîtriser l'orthographe. Même l'utilisation d'internet requiert des compétences dans ce domaine, à la fois pour effectuer des recherches mais aussi pour se faire comprendre de ses interlocuteurs. Ainsi, dans un contexte où la demande sociale en matière d'écrit explose, avec un degré d'exigence de la maîtrise de l'orthographe qui ne diminue pas, les médias relaient régulièrement les mauvais résultats des écoliers français dans les enquêtes internationales. Les discours de déploration du niveau d'orthographe des français concernent aussi le public adulte, y compris des étudiants et des cadres.

L'enquête Information et Vie quotidienne (IVQ) de l'INSEE s'intéresse aux compétences des adultes par rapport à la maîtrise des fondamentaux (écrire, communiquer, compter). Dans sa dernière édition de 2011, 2037 personnes (soit 16% des enquêtés au niveau national) ont été dirigées vers un module intitulé module « bas » parce que leurs résultats dans le module dit « d'orientation » laissaient présager des difficultés avec l'écrit. Au terme de l'enquête, ce sont 7% des citoyens français qui seraient en situation d'illettrisme et ne seraient donc pas en mesure de lire ni d'écrire un message simple pour un usage du quotidien (rédiger un chèque, comprendre une consigne, etc.).

L'école est alors accusée de ne pas être capable de former des citoyens maîtrisant les fondamentaux : en effet, selon la définition de l'Agence Nationale de Lutte Contre l'Illettrisme (ANLCI), l'illettrisme qualifie la situation des personnes (âgées de 16 ans au minimum) qui ont été scolarisées en France, ce qui est lourd de conséquence pour la société et le regard qu'elle peut porter sur l'institution scolaire. Ainsi, on parle d'illettrisme, pour qualifier la situation au regard de l'écrit, des personnes qui :

« [...] après avoir été scolarisées en France, n'ont pas acquis une maîtrise suffisante de la lecture, de l'écriture, du calcul, des compétences de base, pour être autonomes dans les situations simples de la vie courante. Il s'agit pour elles de réapprendre, de renouer avec la culture de l'écrit, avec les formations de base, dans le cadre de la politique de lutte contre l'illettrisme » (ANLCI, 2003).

La baisse de niveau des élèves en matière d'orthographe, déjà observée il y a une dizaine d'années (Manesse et Cogis, 2007), se maintient et continue même de se renforcer. L'enquête « Lire, écrire, compter » conduite par la DEPP en 1987, 2007 et 2015 montre qu'en vingt ans, les compétences orthographiques des élèves d'une même classe d'âge ont diminué, avec un

peu plus de 50% des erreurs liées à la morphologie flexionnelle (Andreu et Steinmetz, 2016). Au niveau international, la dernière enquête PISA montre que les élèves français se situent tout juste dans la moyenne des pays de l'OCDE, loin derrière certains pays comme l'Allemagne ou le Canada. Par ailleurs, le système éducatif français apparaît comme l'un des plus inégalitaires, avec des élèves très performants et d'autres qui maîtrisent à peine la lecture en fin de scolarité obligatoire (Journée Défense et Citoyenneté, 2019). L'enquête « Lire, écrire, compter » avait également montré l'existence de différences de performances orthographiques selon le milieu social des élèves.

Pourtant, à voir la baisse du nombre d'heures consacré à l'apprentissage de la langue et de l'orthographe, il apparaît, nous le verrons, que les demandes faites à l'école dépassent largement le lire, écrire et compter qui constituaient autrefois les missions principales assignées à l'école. Il n'en reste pas moins que les attentes en matière de compétences à l'écrit sont extrêmement fortes et chacun se doit de répondre à ces exigences de la vie quotidienne et professionnelle : il faut savoir écrire et surtout écrire conformément à la norme orthographique.

Dès lors, l'enseignement de l'orthographe ne se cantonne plus à la sphère scolaire : réintroduit cette année dans les programmes du lycée, il concerne également les étudiants, les demandeurs d'emplois, les salariés peu ou pas qualifiés, les cadres et dirigeants d'entreprise. Les offres de formation sous forme de stages et de modules en ligne prolifèrent. Cependant, une question importante se pose en amont : même en s'en tenant à la composante de l'orthographe parmi l'ensemble des composantes du lire-écrire, quels sont au juste les besoins spécifiques des publics concernés ? Comment adopter un regard plus fin que celui qui consiste à pointer les erreurs et à opposer ceux qui ne savent pas à ceux qui ne savent pas ? Peut-on identifier des niveaux de compétences ? Telles sont les questions à l'origine de la recherche que nous avons menée.

Pour évaluer les compétences orthographiques d'adultes, nous avons exploité une partie du corpus issu de l'enquête Information et Vie Quotidienne, recueilli en 2011 par l'INSEE. Il s'agit des dictées réalisées par les enquêtés repérés en difficultés avec l'écrit suite à leurs résultats au module « d'orientation » de l'enquête. Une demande d'analyse plus poussée des dictées, à l'initiative de Jean-Pierre Jeantheau (chargé de mission pour l'Agence Nationale de Lutte Contre l'Illettrisme), a donné lieu à une allocation doctorale financée par la communauté de recherche académique (ARC 8, « Industrialisation et sciences de gouvernement ») de la Région Rhône-Alpes. L'ANLCI avait déjà procédé à un recodage des données de ce corpus de dictées, mais sans doute, faute de moyens, a cherché un partenariat universitaire auquel nous avons répondu. Toujours à l'initiative de Jean-Pierre Jeantheau, un second projet est venu compléter

le premier et a été rendu possible grâce à un financement offert par la Délégation générale à la langue française et aux langues de France (DGLFLF), sollicitée au titre de sa mission de soutien des actions visant « à développer la maîtrise partagée du français et à lutter contre l'exclusion linguistique ». Il consiste à réinterroger des enquêtés qui ont répondu à l'enquête de l'INSEE en 2011 mais qui n'avait pas produit de dictées (parce qu'ils n'avaient pas été considérés en difficulté avec l'écrit). Les 87 personnes qui ont accepté d'être réinterrogées dans la région Rhône-Alpes¹ ont passé une dictée (augmentée d'un « rayon bricolage ») qui nous permet de mieux interpréter les résultats de la dictée produite par les enquêtés repérés en difficulté avec l'écrit (dans IVQ 2011) et d'évaluer leurs particularités en production écrite.

Notre corpus est ainsi composé de deux échantillons.

Le premier est composé des productions des 441 enquêtés qui sont ceux qui ont produit une dictée dans le cadre de l'enquête IVQ (INSEE, 2011) parce qu'ils ont été considérés en difficulté avec l'écrit suite à leurs résultats au module « d'orientation ». Dans notre travail, nous appellerons ce corpus de scripteurs considérés en difficulté (à l'issue du module d'orientation) : corpus « bas² ».

Le second comprend 87 productions écrites, questionnaires et entretiens produits par une partie des enquêtés de l'enquête IVQ de 2011, qui n'avaient pas été considérés en difficultés avec l'écrit dans l'enquête de 2011 (et que nous avons réinterrogés en 2014). Nous appellerons les enquêtés de cet échantillon, enquêtés « haut », par contraste avec ceux qui ont été dirigés vers le module « bas » et que nous appelons enquêtés « bas ».

Notre thèse s'articule autour de trois parties que nous détaillons pour permettre au lecteur de disposer d'une vue d'ensemble de ce travail :

La première partie est composée de trois chapitres qui posent le cadre théorique de cette recherche. Le **premier chapitre** contextualise le public qui intéresse notre étude : les adultes en difficulté avec l'écrit. Après avoir situé le contexte historique et social de l'illettrisme, nous proposons une observation de ce phénomène sous l'angle de la littéracie, qui définit l'écrit selon la diversité de ses usages, dans une dimension fonctionnelle. Cette vision de l'écriture dans une perspective davantage communicationnelle que prescriptive nous amène à

¹ Ce projet a également été mené en Haute-Normandie, par le laboratoire Dylis de Rouen.

² Nous reprenons le terme « bas » de la terminologie utilisée par les concepteurs du test, en référence au module « bas » destiné aux personnes repérées en difficulté avec l'écrit au terme du module d'orientation de l'enquête IVQ.

présenter quelques enquêtes que nous considérons comme représentatives de l'évaluation des compétences en littéracie de jeunes adultes (dont la sortie de la scolarité obligatoire est proche) et d'adultes. Nous détaillons alors précisément le protocole de l'enquête Information et Vie Quotidienne (IVQ, 2011). Dans ce chapitre, nous tâchons de comprendre en quoi l'évaluation de l'écriture et des compétences orthographiques représente une difficulté dans les enquêtes. Le **deuxième chapitre** est centré sur l'orthographe du français et ses difficultés. Nous montrons en quoi le système orthographique s'inscrit dans une histoire, qui explique à la fois son fonctionnement et la nécessité de son apprentissage. Dans la perspective d'évaluer les compétences orthographiques de scripteurs, cette approche linguistique est essentielle, notamment pour déterminer les critères d'évaluation tout en ayant conscience des difficultés auxquelles sont confrontés les scripteurs. Nous insistons sur la mixité de l'écriture du français et proposons une description linguistique du système orthographique en lui-même et en synchronie, en nous appuyant en particulier sur les travaux du groupe HESO et son plurisystème graphique (Catach et al, 1980) et sur ceux de Blanche-Benveniste et Chervel (1969). Dans le **troisième chapitre**, nous présentons les développements récents de la recherche en psycholinguistique, qui complètent la description linguistique du système en permettant une approche plus fine du niveau de difficulté orthographique des mots du français (avec les notions de consistance et de fréquence). Cette partie est l'occasion de présenter différents outils utiles pour la recherche, que nous avons utilisés pour interpréter les résultats des enquêtés. Enfin, nous terminons par quelques apports de la psycholinguistique sur les processus qui sont à l'œuvre lorsque les enfants apprennent l'orthographe et qui aident à comprendre l'origine des difficultés orthographiques des adultes. Nous nous centrons sur quelques difficultés emblématiques de l'orthographe du français, comme la gestion de la morphologie et les finales muettes.

La deuxième partie propose dans un **chapitre quatre**, une présentation des enquêtes et des choix méthodologiques. Nous exposons dans un premier temps les choix effectués par les équipes de recherche qui ont participé à l'élaboration du protocole de l'enquête IVQ de 2011, ainsi que le traitement des données qui a été fait dans le fichier qui nous a été communiqué par l'INSEE. Nous présentons alors notre position pour l'exploitation de ces dictées d'adultes faiblement littéraciés, notamment en termes de compétences et pas seulement en termes de manques en matière d'orthographe. Ce chapitre est l'occasion d'aborder la difficulté à traiter

des données manuscrites et à valoriser d'autres compétences que l'orthographe normée. Nous présentons pour terminer le second corpus, ainsi que les conditions de son recueil.

La troisième partie est composée de deux chapitres qui présentent nos résultats et analyses. Le **cinquième chapitre** présente notre analyse des 441 enquêtés repérés comme étant en difficulté avec l'écrit, les enquêtés « bas ». Dans un premier temps, nous exploitons les réponses des enquêtés selon une analyse traditionnelle des réponses (en terme d'écart à la norme), axée sur les compétences en orthographe lexicale et morphographiques observées sur les mots puis phonographiques sur les pseudo-mots. Nous complétons ces analyses par d'autres, plus qualitatives, destinées à décrire finement les graphies qui s'écartent de la norme et à mettre en relief les compétences des enquêtés dans d'autres domaines moins couramment analysés. Il s'agit en particulier de valoriser leur capacité à produire des graphies certes non normées, mais efficaces d'un point de vue communicationnel. Dans le **sixième chapitre**, nous présentons les résultats de l'enquête complémentaire menée auprès de 87 scripteurs « haut ». Nous comparons dans un premier temps les résultats obtenus par les enquêtés « haut » avec ceux obtenus par les enquêtés « bas » sur la partie commune de la dictée (partie d'IVQ). Nous présentons ensuite l'apport de ces dictées produites par les enquêtés « haut » pour interpréter les productions des enquêtés « bas » et aussi, plus largement, pour évaluer l'orthographe chez des adultes dans une dictée. Dans un deuxième temps, nous présentons et analysons les résultats obtenus par les enquêtés « haut » sur le « rayon bricolage », qu'ils ont été les seuls à produire, ainsi que les questionnaires et entretiens recueillis dans le cadre du projet complémentaire, que nous avons intégralement transcrits.

En conclusion, nous proposons un bilan des résultats observés sur les dictées issues de l'enquête Information et Vie Quotidienne (2011) et sur les dictées de l'enquête complémentaire, sur le « rayon bricolage » (2014). Nous proposons ensuite des pistes d'amélioration de la dictée d'IVQ, qui sont issues de nos analyses. Enfin, nous terminons par quelques pistes didactiques issues des dernières recherches, qui concernent en particulier l'enseignement de l'orthographe à l'école, là où tout commence, et qui sont éventuellement transférables dans d'autres contextes de formation, notamment auprès d'un public adulte.

CADRE THÉORIQUE

CHAPITRE 1. LA LITTÉRACIE, UN CADRE POUR REPENSER L'ILLETTRISME ET LES COMPÉTENCES SCRIPTURALES DES ADULTES

1. LA LITTÉRACIE, UN CADRE POUR REPENSER L'ILLETTRISME.....	24
1.1. L'ILLETTRISME AUJOURD'HUI : UNE DÉFINITION IMPOSSIBLE ?.....	24
1.2. LES ATTENTES DE L'ANLCI EN MATIÈRE DE POLITIQUE PUBLIQUE.....	31
1.3. LA LITTÉRACIE : UNE APPROCHE RÉCENTE DES COMPÉTENCES À L'ÉCRIT	37
1.4. CONCLUSION SUR L'ÉCRITURE DES ADULTES DANS LE CONTEXTE DE LA LITTÉRACIE	45
2. L'ÉVALUATION DES ADULTES À L'ÉCRIT ET PLUS PARTICULIÈREMENT EN ÉCRITURE	46
2.1. TOUR D'HORIZON DES ENQUÊTES SUR LA LITTÉRACIE RÉALISÉES AUPRÈS D'ADULTES ET DE JEUNES ADULTES.....	48
2.2. L'ENQUÊTE INFORMATION ET VIE QUOTIDIENNE (INSEE, 2002, 2004, 2011).....	64

Avec la multiplication des supports de communication, l'accès à l'information est devenu un enjeu crucial pour les adultes, autant dans la sphère professionnelle que dans la sphère privée. La maîtrise de l'écrit est aujourd'hui une condition incontournable pour être intégré dans la société et la production d'écrits conformes à la norme orthographique, une compétence valorisée. La diversité des supports disponibles met l'accent sur l'utilisation de l'écrit là où il y a quelques décennies, le canal oral, par le biais de la téléphonie, était plus largement utilisé. Pour communiquer, les personnes sont désormais davantage placées en position d'émetteur et la demande sociale dans le domaine de l'orthographe est très forte.

La multiplication des sources d'information ne fait alors que complexifier les tâches du quotidien pour ceux qui ont déjà un rapport difficile avec l'écrit et dans le milieu professionnel, la diversité des compétences requises dans une économie de l'information conditionne de plus en plus les carrières et les salaires. Même dans la sphère privée, sur internet, la maîtrise de l'orthographe conditionne la qualité des recherches et des discussions dans les forums. En cela, une bonne maîtrise des savoirs de base nécessaires à la manipulation d'un message que représentent la lecture, l'écriture, la compréhension orale et le calcul s'avère indispensable.

Fayol (2013) rapporte que l'apprentissage de l'écrit est systématique depuis environ 150 ans en France. Et si l'analphabétisme a quasiment disparu, l'illettrisme subsiste encore, comme le montre l'enquête Information et Vie Quotidienne de l'INSEE (IVQ 2011). Ainsi, malgré un enseignement désormais systématique, précoce et prolongé dans un environnement où abonde l'écrit, un certain nombre de personnes (7% d'après les résultats d'IVQ 2011), ne maîtriseraient pas la lecture et l'écriture dans des situations du quotidien. De nombreuses enquêtes depuis les années 1990 cherchent à chiffrer les populations en difficulté et à identifier les raisons de leur faible maîtrise de l'écrit. Les différences de résultats obtenus selon les enquêtes montrent que la problématique est complexe, avec des chiffres qui nous le verrons, différent beaucoup selon les études. De nombreux travaux mettent en avant l'existence de déterminants sociaux favorables à cette situation (Lahire, 1993 ; Besse, 1995 ; Besse et Guérin-Pace, 2002 ; Guérin-Pace, 2009). Mais le système orthographique du français est également mis en cause, comme nous le verrons dans le chapitre 2 consacré à l'orthographe du français (Catach, 1995 ; Lucci et Nazé, 1979), ce qui est conforté par les travaux effectués en psycholinguistique et qui font l'objet du chapitre 3 de notre travail. Fayol (2008), Fayol et Jaffré (2014) montrent en effet que même les plus experts en écriture commettent des erreurs d'orthographe.

La première partie de ce chapitre, est consacrée à contextualiser le public qui intéresse notre étude : les adultes en difficulté avec l'écrit. Nous commencerons par situer le contexte

historique et social de l'illettrisme, terme « inventé » selon les propos de Lahire (1999) il y a maintenant une quarantaine d'années et dont les contours, au gré des définitions et des actions menées par l'Agence Nationale de Lutte contre l'Illettrisme (ANLCI), n'ont pas cessé d'évoluer. Nous proposons ensuite d'observer le phénomène « illettrisme » à la lumière de la notion relativement récente de littéracie, qui reflète la diversité des usages de l'écrit. Nous terminerons par une brève présentation de ce que représente l'acte d'écrire (et l'acte d'orthographe), en partant des travaux initiés par Dabène (1987).

La deuxième partie de ce chapitre fournit une présentation des grandes enquêtes, de dimension nationale et internationale, qui ont été conduites pour évaluer les compétences en littéracie de deux publics : les jeunes adultes, dont la sortie de la scolarité obligatoire est imminente et les adultes, qui intéressent plus particulièrement notre propos. Nous présenterons plus longuement l'enquête Information et Vie Quotidienne (IVQ, 2011) dont est issu notre corpus et nous terminerons par une présentation des principaux résultats qui ont été publiés.

1. LA LITTÉRACIE, UN CADRE POUR REPENSER L'ILLETTRISME

« Illettrisme : ce terme apparut au début des années quatre-vingts, mis en question par les uns, considéré au pluriel par d'autres, régulièrement contesté, mais tout aussi régulièrement repris [...] » (Barré-De-Miniac, 1997)

1.1. L'ILLETTRISME AUJOURD'HUI : UNE DÉFINITION IMPOSSIBLE ?

1.1.1. BREF RAPPEL HISTORIQUE

La notion d'illettrisme est apparue en France dans les années soixante-dix et a été introduite par l'association ATD Quart Monde (Aide à Toute Détresse Quart Monde) qui a communiqué sur cette question dès 1967. Cette association, au contact des plus démunis, est investie de missions en lien avec la pauvreté et l'exclusion sociale. La France découvre alors qu'une tranche de la population pourtant née et scolarisée en France, ne dispose pas des compétences de base à l'écrit, lui permettant d'être autonome dans le quotidien. Si cette problématique avait été soulevée chez les personnes issues de l'immigration, l'origine de leurs difficultés aurait été facilement reliée à leur parcours de vie. En revanche, il s'agit d'une découverte pour les Français qui ont été scolarisés dans une école de la République alors même que l'école obligatoire et sa démocratisation datent de presque un siècle. Mais si cette scolarisation massive a permis d'enrayer l'analphabétisme, elle n'a pu empêcher l'existence d'un autre phénomène souterrain dont la France n'avait pas pris la mesure. L'association ATD Quart Monde pose alors le terme d'illettrisme pour distinguer les travailleurs immigrés

qualifiés d'analphabètes des Français scolarisés en France mais dont les compétences en lecture et en écriture sont dites « limitées ». Ainsi, le terme « illettrisme » qualifie depuis sa création, ceux qui ont suivi en partie ou en totalité le cycle primaire de l'école française, mais sans être parvenus à y acquérir les compétences de base requises.

En 1983, le rapport Espérandieu « *Des illettrés en France* » (Espérandieu et al., 1984) marque la prise en compte officielle de l'illettrisme par l'État français et dès 1984, la lutte contre l'illettrisme est affichée comme une priorité nationale au conseil des ministres. C'est alors qu'un outil institutionnel prend naissance sous la forme du Groupe Permanent de Lutte contre l'illettrisme (GPLI) qui devient en 2003 l'Agence Nationale de Lutte contre l'illettrisme (ANLCI). L'ANLCI se voit alors confier comme mission prioritaire, la mise en place d'une évaluation rigoureuse pour évaluer l'ampleur du phénomène et en donner une définition précise. En 2003, la publication du cadre national de référence de l'ANLCI « Lutter ensemble contre l'illettrisme » (sur lequel nous revenons plus longuement dans la section 1.2) permet de définir plus précisément l'illettrisme. Il donne une vision complète des différentes composantes de la lutte contre l'illettrisme, en explicitant les principes directeurs, les champs d'intervention et les recommandations. Dans ce cadre, on insiste notamment sur la pluralité des situations et donc sur la pluralité des actions à mener : « [...] parce que les hommes et les femmes en situation d'illettrisme sont de tous âges et vivent dans des contextes très différents. » (ANLCI, Cadre national de référence, 2012 : 1).

Ainsi, le phénomène illettrisme que l'on voulait « passer » dans les années 1980 et dont les moyens se voulaient radicaux, avec des actions fortes et un vocabulaire qualifié de « guerrier » perdure et ses contours, malgré les nombreux travaux qui ont fleuri après la création du GPLI, sont toujours aussi flous. Cette période féconde en réflexions concernant les actions à mener, en enquêtes (Infométrie en 1988, IVQ en 2004 et 2011) et en travaux de recherches dont ceux de Lahire, qui à partir de 1999 s'est intéressé à la question sociale de l'illettrisme, servent aujourd'hui encore de référence. Et il semblerait malgré les tentatives des 20 dernières années, alors même que le phénomène diminue (nous le verrons en étudiant les données chiffrées), que nous ne sachions toujours pas déterminer précisément ce qui se cache derrière des statistiques. C'est dire si la question est complexe.

1.1.2. QUELQUES CHIFFRES ISSUS DE DIFFÉRENTES ENQUÊTES EN FRANCE

Pour déterminer le taux d'illettrisme en France, deux types de recueil de données ont été utilisés depuis les années 1980 : les recensements de la population effectués par l'INSEE et les enquêtes construites pour un objectif en particulier (évaluer les compétences à l'écrit d'une tranche d'âge et déterminer le taux d'illettrisme). D'autres enquêtes (nationales et

internationales) permettent d'évaluer le niveau de littéracie chez les jeunes et chez les adultes (que nous détaillerons dans la section 2).

Nous présentons à présent quelques données chiffrées qui ont été diffusées depuis les années 1980 et qui nous paraissent importantes concernant l'évolution de l'évaluation des compétences chez les adultes.

Dans les années 1980, en 1986-1987, suite au recensement de la population, l'INSEE estime que 9,1 % de la population se trouve en situation d'illettrisme. Une année après, en 1988, l'enquête Infométrie réalisée par le GPLI et destinée à déterminer le taux d'illettrisme en France montre que 20 % de la population serait en situation d'illettrisme. L'écart entre ces deux chiffres alors qu'une seule année sépare les deux enquêtes montre la complexité à évaluer les compétences à l'écrit des adultes et surtout à définir des critères qui permettent de déterminer la proportion d'illettrés. Les différences de résultat entre ces deux enquêtes (soit la première sous-estime le taux d'illettrés, soit la deuxième le surestime) montrent surtout qu'il est difficile d'utiliser des données destinées à un autre objectif (recenser la population française) pour une autre fin (déterminer le taux d'illettrisme).

Dans les années 1990, en 1994 soit 7 ans après le précédent recensement de l'INSEE ayant déterminé que 9,1 % de la population se trouvait en situation d'illettrisme, un nouveau recensement effectué également par l'INSEE annonce 5,4 % d'illettrés. Une diminution de presque 4 points du pourcentage d'illettrés paraît peu plausible, en seulement 7 ans. Une année plus tard, en 1995, IALS, une enquête internationale de l'OCDE (que nous présentons plus en détail dans la section 2.1.1), établit que 40 % des Français ne disposeraient pas de capacités suffisantes en lecture et écriture nécessaires pour la vie quotidienne. La même année, selon le ministère de la Défense, 20 % des garçons incorporés n'accèderaient pas au sens d'un texte explicatif, laissant sous-entendre de faibles compétences en lecture.

Dans les années 2000, ce sont d'abord les tests de lecture réalisés lors des Journées d'Appel à la Défense qui alertent, avec 9,7 % des jeunes âgés de 16 ans qui seraient en situation d'illettrisme. Puis les années 2004 et 2011 marquent un tournant concernant l'évaluation des compétences en littéracie des adultes ainsi que celle de l'évaluation du taux d'illettrisme en France avec l'enquête Information et Vie Quotidienne (IVQ). Réalisée par l'INSEE en partenariat avec l'Agence Nationale de Lutte contre l'Illettrisme, cette enquête porte spécifiquement sur l'évaluation du niveau de littéracie de la population française âgée de 16 à 65 ans. Elle comporte un module inédit conçu pour rendre compte des difficultés à l'écrit que rencontrent les plus faibles niveaux de compétences. Réalisée à grande échelle en 2004

et 2011, l'enquête établit que 9 % de la population se trouverait en situation d'illettrisme en 2004 et 7% en 2011.

1.1.3. UNE DÉFINITION QUI TÉMOIGNE D'UNE RÉALITÉ COMPLEXE

Le cadre national de référence, qui a été défini par l'ANLCI en 2003, pose une définition qui permet de clarifier la distinction entre illettré et analphabète et qui chasse certaines idées reçues concernant le cumul des difficultés – idée selon laquelle un illettré aurait des difficultés dans toutes les composantes de l'écrit (lecture, écriture, compréhension écrite et orale). La définition insiste sur les acquis des personnes en situation d'illettrisme – les illettrés ont été scolarisés, ils ont des connaissances et des compétences – ainsi que sur leur position dans la société – ils ne sont pas nécessairement en situation d'exclusion, même si cette réalité est également présente. Certains travaillent et ont même très bien réussi professionnellement. Ainsi :

« L'illettrisme qualifie la situation de personnes de plus de 16 ans qui, bien qu'ayant été scolarisées, ne parviennent pas à lire et comprendre un texte portant sur des situations de leur vie quotidienne, et/ou ne parviennent pas à écrire pour transmettre des informations simples.

Pour certaines personnes, ces difficultés en lecture et écriture peuvent se combiner, à des degrés divers, avec une insuffisante maîtrise d'autres compétences de base comme la communication orale, le raisonnement logique, la compréhension et l'utilisation des nombres et des opérations, la prise de repères dans l'espace et dans le temps, etc.

Malgré ces déficits, les personnes en situation d'illettrisme ont acquis de l'expérience, une culture et un capital de compétences en ne s'appuyant pas ou peu sur la capacité à lire et à écrire. Certaines ont pu ainsi s'intégrer à la vie sociale et professionnelle, mais l'équilibre est fragile, et le risque de marginalisation permanent. D'autres se trouvent dans des situations d'exclusion où l'illettrisme se conjugue avec d'autres facteurs. » (Cadre national de référence de l'ANLCI, 2003 : 72).

Cette définition reprend certains points déjà évoqués dans la précédente définition proposée par le GPLI³ en 1995, dans une volonté de continuité. Certains éléments ont cependant été écartés de la définition, car la nouvelle insiste davantage sur les acquis plutôt que sur les manquements, ainsi que sur l'intégration des sujets dans la société. Ont ainsi été retirés les

³ Groupe Permanent de Lutte contre l'Illettrisme (GPLI), devenu Agence Nationale de Lutte contre l'Illettrisme (ANLCI) depuis 2000.

éléments relatifs aux exigences minimales, aux causes de l'illettrisme et au rapport à l'écrit des individus :

« Personnes [...] ne maîtrisant pas suffisamment l'écrit pour faire face aux exigences minimales requises dans leur vie professionnelle, sociale, culturelle et personnelle [...] ayant peu ou mal acquis les savoirs premiers pour des raisons sociales, familiales ou fonctionnelles [...] n'ont pu user de ces savoirs et/ou n'ont jamais acquis le goût de cet usage. Il s'agit d'hommes et de femmes pour lesquels le recours à l'écrit n'est ni immédiat, ni spontané, ni facile, et qui évitent et/ou appréhendent ce moyen d'expression et de communication. » (GPLI, 1995 dans le cadre national de référence de l'ANLCI, 2003 : 72).

Cette définition du GPLI de 1995, avait néanmoins fait évoluer celle proposée en 1978 par l'UNESCO, qui définissait une personne illettrée comme « incapable de lire et d'écrire en le comprenant un exposé simple et bref en rapport avec la vie quotidienne. » (citée par Lahire, 1999) alors que cette définition est aujourd'hui associée à celle de l'analphabétisme fonctionnel (UNESCO, 2006).

1.1.4. « LES ILLETTRISMES » : UNE MEILLEURE CONNAISSANCE DES PUBLICS AUJOURD'HUI

Depuis la prise de conscience en France de l'illettrisme et la mesure du phénomène (plus tardif que pour l'analphabétisme), l'évolution des définitions et le nombre de travaux qui décrivent la diversité des situations sociales ont permis de dépasser le lien réducteur qui a souvent été fait entre illettrisme et exclusion sociale (Besse, 1995 ; Villechaise-Dupont et Zaffrans, 2002 ; Filhon, 2014 ; Guérin-Pace, 2009). Guérin-Pace (2009) a notamment travaillé sur ces questions d'après les données relatives aux informations personnelles des enquêtés d'IVQ 2004 recueillies dans le module biographique. L'illettrisme a en effet souvent été abordé en termes de conséquences plutôt que de causes ; celles-ci ayant été sans réel fondement souvent uniquement associées à des situations sociales difficiles. Les données d'IVQ permettent de montrer qu'il existe de multiples formes d'illettrismes qui représentent autant de pratiques de l'écrit et de connaissances à l'écrit différentes :

« [...] il n'existe pas un profil type des personnes en grande difficulté face à l'écrit, mais un ensemble de configurations associant des éléments du parcours à des conditions sociales et économiques contemporaines. » (Guérin-Pace, 2009 : 49).

Les sujets en situation d'illettrisme ont également des parcours différents et occupent des positions très diverses au sein de la société :

« Même pour les personnes les plus en difficulté face à l'écrit, considérées en situation d'illettrisme, les difficultés rencontrées au quotidien et les éléments de sociabilité se conjuguent différemment selon les éléments déterminants des parcours. ».

Sans occulter les difficultés au quotidien qui sont inévitablement rencontrées par les personnes qui maîtrisent mal l'écrit, ce sont les conséquences de ces difficultés qui sont très variables d'une personne à l'autre. Guérin-Pace (2009) rejoint ainsi les conclusions de Besse (1995) et de Lahire (1999) pour qui il n'existe pas « un illettrisme » mais « des illettrismes », dont les causes tout comme les manifestations sont variables. Cette prise de conscience est importante pour la prise en compte de la complexité du rapport à l'écrit.

En revanche, même si les profils sont très variés, Guérin-Pace (2009) montre, à partir des éléments présents dans IVQ 2004, que certaines variables sont néanmoins très souvent corrélées à des difficultés importantes avec l'écrit à l'âge adulte. Elle montre que celles-ci prennent très souvent leur source durant l'enfance et pas uniquement parce que le milieu social est particulièrement défavorisé. Nous présentons brièvement les deux variables qui apparaissent, selon l'analyse réalisée par Guérin-Pace (2009), de manière très nette dans l'enquête.

Le contexte familial durant l'enfance : chez les adultes en difficulté avec l'écrit, l'évocation d'une arrivée précoce de ces difficultés dans l'enfance est relativement unanime chez les enquêtés, souvent dans un contexte familial difficile ou lorsqu'un événement traumatisant est survenu durant l'enfance (décès, violence à l'école, etc.). Ces conclusions rejoignent celles de Fijalkow (1986) qui a établi que des difficultés familiales durant l'enfance ne favorisaient pas les apprentissages scolaires. Ainsi lorsque des événements extérieurs à l'école (dans le cercle familial) ou liés à l'école, notamment les violences scolaires (dont la société prend conscience depuis quelques années avec les campagnes récentes de sensibilisation) viennent perturber les apprentissages et que les compétences de base ne se mettent pas en place au début de la scolarité, il semble plus difficile de combler ce retard à l'âge adulte.

L'absence de pratique de la lecture dans le cercle familial : Guérin-Pace (2009) a également montré que l'absence de pratique de la lecture comme loisir durant l'enfance était un facteur déterminant de la survenue de difficultés d'apprentissage. La pratique de parents en matière de lecture est alors fortement corrélée à celle des enfants ; l'absence de pratique parentale multipliant presque par deux la probabilité des enfants de parents non lecteurs d'être en difficulté dans les apprentissages scolaires. Sur ce point la corrélation inverse ne se vérifie pas. Selon les déclarations des enquêtés, un quart de ceux dont les parents ne lisaient jamais lorsqu'ils étaient enfants étaient des lecteurs réguliers et ne sont pas, adultes, en difficulté

particulière avec l'écrit. Il semblerait ainsi qu'une pratique régulière de la lecture permettrait aux adultes de maintenir certaines connaissances et compétences qui ne sont plus mobilisées une fois sortis du système scolaire (Guérin-Pace, 2009).

Ainsi, alors que dans les années 1980, l'illettrisme était souvent associé à l'analphabétisme, par manque de connaissance des publics concernés et de leurs compétences, son rattachement récent au concept de littéracie représente une évolution importante à la fois pour le repérage des personnes en situation d'illettrisme ainsi que pour la prise en compte de leurs besoins en formation. L'illettrisme recule à en croire les chiffres des deux dernières enquêtes IVQ, notamment grâce à l'effet de génération⁴ sur lequel nous reviendrons et qui explique en partie le recul du taux d'illettrisme (Murat et Rocher, 2016).

Pour lutter contre l'illettrisme, les actions menées se distinguent en fonction de l'objectif fixé et du public ciblé : il s'agit de prévention lorsque l'on s'intéresse aux jeunes âgés de moins de 16 ans et à leurs familles. Après 16 ans, la lutte contre l'illettrisme est surtout conduite par des actions de formation.

Nous l'avons vu dans la définition de l'ANLCI (2003 : 72), les critères qui permettent de considérer qu'une personne se trouve en situation d'illettrisme sont fortement liés à une conception scolaire des acquis qui apparaît en terme de manquement de l'école « *bien qu'ayant été scolarisées* », « *ces difficultés en lecture et écriture peuvent se combiner, à des degrés divers, avec une insuffisante maîtrise d'autres compétences de base* ». L'acquisition de ces compétences de base est alors un objectif des démarches de formation d'adultes. Cette notion de compétences de base n'est pas clairement définie. Elle apparaît dans de nombreux textes, notamment de l'ANLCI. Dans la définition de l'illettrisme, nous venons de le voir, mais aussi dans son référentiel de compétences (que nous présentons dans la section suivante), présenté comme un outil permettant une « *véritable égalité des chances d'accès à la lecture, à l'écriture, à la langue française ainsi qu'aux compétences de base.* » (Cadre National de Référence de l'ANLCI, 2003). En lien avec la notion de littéracie et donc à une dimension fonctionnelle, nous pouvons nous interroger sur les compétences exactes auxquelles ces « compétences de base » renvoient. Pour mieux les cerner, nous présentons dans la partie

⁴ Les enquêtes sur les compétences à l'écrit des adultes (en particulier IVQ et PIAAC) montrent que les 60-65 ans (dernière tranche d'âge évaluée) ont de moins bons résultats que les plus jeunes. Il a été montré que les compétences sont fortement liées au niveau de diplôme (et qu'une part importante des écarts selon l'âge s'explique justement par l'élévation du niveau de diplôme au fil des générations). Dans IVQ, en 2004, les plus âgés étaient nés en 1939. En 2011, la génération née entre 1939 et 1945 est sortie du champ de l'enquête.

suivante, le référentiel de compétences de l'ANLCI, publié en 2003 : « le cadre national de référence ».

1.2. LES ATTENTES DE L'ANLCI EN MATIÈRE DE POLITIQUE PUBLIQUE

« Outil fédérateur, point d'ancrage commun à tous les acteurs, dans tous les territoires de France métropolitaine et d'Outremer, il veut les rassembler, dans le cadre d'une coopération solidaire, pour construire une véritable égalité des chances d'accès à la lecture, à l'écriture, à la langue française ainsi qu'aux compétences de base. » (Cadre national de référence de l'ANLCI, 2003 : 75).

1.2.1. LE CADRE NATIONAL DE RÉFÉRENCE DE L'ANLCI

En 2003, l'ANLCI a publié le « cadre national de référence ». Il s'agit d'un outil dont l'objectif est de poser les fondements de la lutte contre l'illettrisme (et principalement des actions de formation), de définir les différents champs d'intervention et de fournir des recommandations à destination des acteurs engagés dans la lutte contre l'illettrisme. Ce socle dit « fonctionnel » permet notamment de décrire une politique nationale de lutte qui soit basée sur des principes cohérents, communs et connus de tous et qui s'inscrive dans une approche plus globale des sujets et plus fonctionnelle des compétences.

Ce cadre fait suite à l'engagement du Ministère de la jeunesse, de l'éducation nationale et de la recherche, qui en 2002, a inscrit la prévention de l'illettrisme comme une priorité éducative nationale. Nous présentons brièvement ci-dessous, deux missions qui nous semblent caractéristiques des actions menées.

1.2.1.1 LA PRISE EN COMPTE DE TOUS LES PUBLICS, DE LA PRÉVENTION À L'ACTION

Le cadre montre la nécessité d'inscrire les actions dans le temps, avec des actions de prévention qui doivent commencer dès la petite enfance et se poursuivre tout au long de la scolarité et au-delà. Pour cela, le cadre insiste sur l'importance d'une prise de conscience collective de l'illettrisme qui touche les adultes, pour que dès le plus jeune âge, durant la scolarité obligatoire, les différents acteurs aient conscience de l'importance des difficultés d'apprentissage de la langue écrite et de ses conséquences futures. En effet, jusque récemment, les pays développés ayant réussi à vaincre l'analphabétisme pensaient que l'éradication de ce phénomène avait permis de doter la population d'une maîtrise suffisante de l'écrit pour être à minima autonome dans les situations de communications rencontrées dans le quotidien. Or nous savons aujourd'hui qu'il n'en est rien.

1.2.1.2 LA FORMATION TOUT AU LONG DE LA VIE

Le cadre définit les missions de la lutte contre l'illettrisme dont une part très importante est portée par des actions de formation, notamment la « formation tout au long de la vie », pour que chacun puisse développer ses compétences de base (dont la maîtrise de la langue française) et sa culture. Concernant la maîtrise de la langue française, il s'agit d'acquérir des compétences linguistiques pour un usage littéraire de l'écrit :

« la maîtrise des compétences linguistiques qui permettent de bien s'exprimer oralement et d'avoir une pratique autonome de l'écrit dans toutes ses dimensions, aussi bien culturelles que fonctionnelles et professionnelles. » (CNR, 2003 : 12).

Sont ainsi concernées les compétences à l'écrit traditionnelles comme la communication écrite, la lecture, l'écriture ainsi que les compétences en mathématiques (numératie). À ces compétences de base, sont ajoutées d'autres compétences périphériques, comme les compétences cognitives (développement du raisonnement logique, de la capacité à apprendre, etc.) et les compétences nécessaires pour une intégration sociale complète avec l'apprentissage des langues étrangères, la connaissance des outils liés aux technologies de l'information et de la communication, etc.

Nous présentons ensuite le type d'actions dans lesquelles l'ANLCI souhaite que les acteurs s'impliquent particulièrement et portent une vigilance particulière. Nous présentons ces actions en fonction du public cible.

1.2.1.3 LES ACTIONS AUPRÈS DE PUBLICS CIBLES

L'ANLCI précise les actions qui peuvent être menées auprès de publics qui ne sont pas nécessairement pris en considération dans les actions de lutte contre l'illettrisme, parce qu'ils ne sont pas directement concernés. Pour autant, les actions menées auprès des enfants (avant qu'ils ne soient éventuellement repérés pour leurs difficultés avec l'écrit, généralement lorsqu'ils sortent du système scolaire) et des parents, participent activement à la lutte contre l'illettrisme, en terme de prévention.

Les actions menées auprès des enfants et des adolescents concernent principalement celles actions qui peuvent être menées dans le cadre scolaire. Dans ce cadre éducatif, l'ANLCI insiste sur l'attention particulière qui doit être portée en matière de maîtrise de l'écrit et en général de la langue française, sur le repérage des difficultés individuelles et sur la réactivité pour trouver une réponse adaptée aux difficultés détectées. Dans ce contexte, sont également impliqués les professionnels de santé qui interviennent sur les troubles qui peuvent affecter

les apprentissages, notamment la rééducation orthophonique pour traiter les troubles du langage écrit ou oral, ainsi que la prise en charge des troubles sensoriels (par exemple la surdité).

« Le but est de créer, dès l'école maternelle, des conditions d'apprentissage toujours meilleures pour préparer les enfants et les adolescents à mieux maîtriser la langue française et à devenir des lecteurs motivés, actifs et autonomes. » (Cadre National de Référence, 2003 : 25).

Dans son cadre de référence, l'ANLCI précise que ces différentes actions peuvent également être menées dans un cadre moins formel, en dehors du temps scolaire, dans le milieu associatif culturel et éducatif, dans les bibliothèques, etc.

Auprès des tout-petits, il s'agit davantage d'actions de prévention dont les formes variées sont en lien avec la lecture et le langage. L'objectif de ces actions auprès des plus jeunes, avant l'âge de la scolarité, est de favoriser « *un contexte favorable aux apprentissages.* » (CNR, 2003 : 27).

De nombreuses études montrent le lien entre le contexte familial et notamment l'implication des familles et la réussite scolaire. Dans le domaine de la lecture, le rôle des parents est très important. L'ANLCI soutient ainsi les actions qui aident les parents à s'inscrire dans une culture écrite et qui aident ceux qui en sont les plus éloignés à s'impliquer dans la scolarité de leurs enfants.

Enfin, l'ANLCI soutient le développement de réseaux de partenaires locaux qui sont à l'initiative de projets dans le domaine éducatif et souligne la richesse de leur expérience du terrain, qui constitue une aide importante au développement de nouveaux projets.

1.2.2. LES RÉFÉRENTIELS DES COMPÉTENCES DE L'ANLCI

L'essentiel des actions de lutte contre l'illettrisme concerne celles qui sont menées auprès des adultes, qui pour être autonomes dans leur quotidien, ne disposent pas des compétences à l'écrit requises. Il s'agit notamment pour eux d'acquérir un « socle fonctionnel » de compétences, que nous présentons dans cette section. Le terme « fonctionnel » est utilisé pour insister sur le fait que lorsqu'il s'agit d'adultes, il ne s'agit pas simplement d'acquérir des

compétences de base ; celles-ci doivent être en lien avec l'usage que les individus vont en faire au quotidien. Il s'agit donc bien de compétences en littéracie⁵.

Nous présentons dans un deuxième temps, le « Référentiel des compétences clés en situation professionnelle », conçu par l'ANLCI en 2009, à la demande des partenaires impliqués dans la formation professionnelle. Cette demande est née d'un constat révélé dans l'enquête IVQ de 2004 concernant la situation professionnelle des personnes en situation d'illettrisme, dont plus de la moitié est dans l'emploi.

1.2.2.1 UN SOCLE FONCTIONNEL DE COMPÉTENCES

Le Cadre National de Référence publié en 2003 par l'ANLCI, décrit également un « socle fonctionnel » de compétences, qui détermine selon 4 degrés de difficultés, les compétences à l'écrit nécessaires pour la vie quotidienne.

Définir ce socle, est important pour les actions de formation qui doivent se baser sur des attendus finaux, à défaut de pouvoir disposer de compétences de départ. Il existe de nombreux référentiels de compétences établis par les institutions en fonction de leur public et des moyens dont elles disposent. Un travail a été initié par l'ANLCI pour étudier les différents référentiels et proposer un outil commun qui propose une progression vers la maîtrise des compétences nécessaires en matière de littéracie. À ce jour, l'ANLCI a défini quatre « degrés » de compétences que nous présentons ci-dessous (Cadre National de Référence, 2003 : 31) :

- **degré 1** : ce premier degré concerne une maîtrise de l'écrit, au moins relative, avec des compétences minimales qui doivent permettre de satisfaire les besoins de communication, en réception de l'écrit comme en production. Ces compétences se traduisent concrètement par la capacité des individus à : « [...] *se repérer dans l'univers de l'écrit (identification des signes et des mots), dans l'univers des nombres (base de la numération), dans l'espace et dans le temps, de participer à des échanges oraux avec des questions réponses simples, etc.* ».

Les personnes qui sont considérées comme relevant de l'illettrisme sont, selon cette progression, celles qui n'ont pas acquis le 1^{er} degré de compétences.

⁵ Nous présentons la littéracie dans la section 1.3 qui suit. Nous définissons notamment brièvement ce terme bien entériné aujourd'hui, que nous choisissons d'orthographier « littéracie » bien que dans la littérature, on trouve également la graphie « littéracie », témoin de l'emprunt à l'anglais.

- **degré 2** : il s'agit de l'acquisition de compétences fonctionnelles pour la vie courante, qui sont celles qui permettent aux individus, dans un environnement familial : « [...] de lire et d'écrire des phrases simples, de trouver des informations dans des documents courants, de donner et de prendre des informations orales lors d'un entretien, de résoudre des problèmes de la vie quotidienne nécessitant des calculs simples, etc. ».

Les personnes qui ont acquis le degré 2 de compétences, qui sont les compétences minimales pour être autonomes dans le quotidien sont néanmoins considérées en difficultés avec l'écrit, car leurs connaissances restent limitées à un usage courant de l'écrit.

- **degré 3** : ce degré de compétences se distingue du précédent par la capacité des individus à accéder à un écrit issu de situations variées, et non uniquement dans un champ qui est restreint (comme dans les degrés 1 et 2) à un usage courant de l'écrit au sein d'un environnement familial. Concrètement, les personnes qui accèdent à un degré 3 de compétences sont capables de : « lire et d'écrire des textes courts, d'argumenter, de résoudre des problèmes plus complexes, d'utiliser plus largement des supports numériques etc. ».

C'est à partir du degré 3 de maîtrise de l'écrit, que l'orthographe commence à être abordée. Il est en effet spécifié que les personnes qui ont acquis le degré 3 de compétences, se sont davantage approprié les « codes », notamment orthographiques.

- **degré 4** : dans ce dernier degré de compétences, se renforce l'autonomie des individus pour interagir dans une société de l'écrit et de la connaissance. Ce degré de compétences correspond aux attendus de fin de scolarité obligatoire, qui doivent permettre aux personnes de disposer des compétences nécessaires pour évoluer dans la société, s'adapter et continuer de se former.

Malgré une meilleure connaissance des publics et de leurs besoins, l'ANLCI souligne que le nombre de personnes se trouvant en situation d'illettrisme qui s'inscrivent dans une formation est très faible ; constat qui se vérifie également dans les autres pays européens : « le besoin existe, mais la demande est difficile à susciter. C'est pourtant l'une des principales conditions du développement de la lutte contre l'illettrisme. » (Cadre National de Référence, 2003). Presque 20 ans après la publication de ce cadre, ce constat se vérifie-t-il toujours ?

1.2.2.2 LE RÉFÉRENTIEL DES COMPÉTENCES CLÉS EN SITUATION PROFESSIONNELLE

En 2009, à la demande des partenaires impliqués dans la formation professionnelle, l'ANLCI publie un premier « Référentiel des compétences clés en situation professionnelle » (référentiel CCSP) qui a été réédité en 2013, accompagné d'un guide d'utilisation.

Ce référentiel est axé sur la formation des personnes faiblement qualifiées et sur l'insertion professionnelle des jeunes qui n'ont aucune qualification. Il est né d'un constat que l'enquête IVQ de 2004 a permis de mettre en avant :

- 1/ plus de la moitié des personnes qui sont en situation d'illettrisme ont un emploi ;
- 2/ plus de la moitié des personnes qui sont en situation d'illettrisme ont plus de 45 ans ;
- 3/ les difficultés avec l'écrit augmentent avec l'âge.

Il est alors apparu indispensable, avec l'allongement de la durée de l'activité professionnelle et le recul de l'âge de départ à la retraite, de prévenir mais aussi de lutter contre l'illettrisme dans le milieu professionnel en permettant aux personnes qui sont dans l'emploi de ne pas perdre leurs connaissances et de continuer à se former pour évoluer. L'illettrisme est ici envisagé selon une conception moins classique (qui témoigne aussi d'une évolution du regard porté), selon laquelle les difficultés ne sont uniquement le fait de connaissances insuffisamment acquises lors de la scolarité. L'illettrisme peut également survenir suite à un parcours de vie dans lequel des compétences (probablement fragiles), auparavant acquises, n'ont pas été sollicitées et ont été oubliées.

Ce dispositif permet également à ceux qui ne sont pas dans l'emploi, d'y accéder, en bénéficiant de formations axées sur l'emploi : « *Nous avons donc privilégié une démarche très pragmatique, très modeste, qui ne prétend pas dire ce que les personnes devraient savoir mais qui a seulement pour objectif d'aider sur le terrain tous ceux qui veulent mettre en œuvre des formations de base.* » (référentiel CCSP, 2013). Ce référentiel est étroitement lié à la « formation tout au long de la vie », qui avait été définie comme prioritaire dans le cadre national de référence de l'ANLCI (2003).

Le terme « compétences de base » renvoie ainsi à un panel de compétences nécessaires à l'utilisation d'un écrit quotidien dans sa dimension fonctionnelle. Ces compétences sont évolutives, car elles suivent les besoins liés à une demande sociale qui dépassent le simple lire, écrire et compter. Par exemple aujourd'hui, ne pas savoir utiliser un ordinateur est de plus en plus problématique à l'heure où la plupart des démarches administratives s'effectuent via internet. Lorsqu'il s'agit de compétences de base dans le milieu professionnel, on peut imaginer qu'une partie de ces compétences serait propre à chaque métier. Ainsi, en renvoyant à un usage fonctionnel de l'écrit, pour le quotidien ou pour un usage professionnel, avoir acquis les « compétences de base » renvoie au concept de littéracie que nous définissons dans la partie suivante.

1.3. LA LITTÉRACIE : UNE APPROCHE RÉCENTE DES COMPÉTENCES À L'ÉCRIT

La notion de littéracie a été introduite par des chercheurs anglo-saxons qui ont utilisé la graphie « literacy ». On trouve également l'orthographe « littératie » dans la littérature (Fraenkel, 2002 ; Charaudeau et Maingueneau, 2002), bien que celle de « littéracie » semble s'être imposée aujourd'hui. Aussi, dans notre travail nous choisissons la graphie « littéracie ». Par ailleurs, la terminologie officielle publiée dans le Journal Officiel de 2005 puis confirmée dans le Journal Officiel de 2017 (Légifrance, 2017 : 97) est « littérisme », défini comme la « *Capacité d'une personne, dans les situations de la vie courante, à lire un texte en le comprenant, ainsi qu'à utiliser et à communiquer une information écrite* ». Parmi les nombreuses définitions de littéracie et pour préciser celle publiée dans le Journal Officiel (2017), nous retiendrons celle de l'OCDE qui fait référence : « *Aptitude à comprendre et à utiliser l'information écrite, dans la vie courante en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités* » (OCDE, 2000). Nous pouvons nous interroger sur ce que signifie « *utiliser de l'information écrite* » et nous demander si la production écrite et l'orthographe font partie des compétences attendues. Définir un niveau de littéracie minimal pour « *comprendre et utiliser de l'information écrite* » est compliqué car la définition même de ce concept, renvoie à une hétérogénéité de compétences à l'écrit et de supports pour les mettre en œuvre. Cependant, des réunions d'experts à l'UNESCO s'efforcent de normaliser le terme en produisant une définition susceptible d'être utilisée dans les enquêtes nationales ou internationale destinées à évaluer le niveau de « littéracie » des populations. Ainsi, en 2003, l'UNESCO définit la littéracie comme « *l'habilité d'identifier, de comprendre, d'interpréter, de créer, de communiquer et de calculer en utilisant du matériel imprimé et écrit associé à des contextes variables. Il suppose une continuité de l'apprentissage pour permettre aux individus d'atteindre leurs objectifs, de développer leurs connaissances et leur potentiel et de participer pleinement à la vie de leur communauté et de la société toute entière* » (UNESCO, 2009 : 13).

Jaffré (2004 : 31), qui souligne également la polyvalence de ce concept résume la littéracie, comme « *l'ensemble des activités humaines qui impliquent l'usage de l'écriture en réception et en production.* » Cette définition implique donc, pour être compétent en littéracie, de maîtriser l'écriture. Pourtant, depuis les années 1990, les protocoles des enquêtes internationales ne permettent pas d'évaluer ce versant de l'écrit : la capacité d'écriture et les compétences orthographiques ne sont en effet jamais investiguées (Jeantheau, 2015). Pourtant, à l'heure de l'écrit numérique, la maîtrise de l'orthographe reste une compétence attendue socialement et particulièrement discriminante.

Malgré les différentes définitions qui existent, toutes distinguent les compétences en littéracie des attendus plus scolaires de la lecture et de l'écriture. Le concept de littéracie prend en effet en considération la nature du support (lettre, mode d'emploi, publicité, etc.) comme la mise en œuvre d'autant de compétences différentes et donc d'attendus différents. Par exemple, un écrit fonctionnel pourra s'affranchir de codes plus stricts requis en littérature. De même, la finalité ne se limite pas aux compétences en lecture-écriture mais davantage à une utilisation de l'écrit qui permette d'extraire ou de produire de l'information dont la forme sera adaptée à une finalité précise : par exemple rédiger une liste de courses, remplir un chèque, etc.

1.3.1. L'ORTHOGRAPHE DES ADULTES : UN CONFLIT ENTRE COMPÉTENCES ET ATTENTES SOCIALES

Chez l'adulte, on attend que ce qui a été appris à l'école en matière d'orthographe soit abouti. Mais on verra dans le chapitre 2 qui présente le système orthographique du français, que même adulte, « *on sait de l'orthographe* » (Jaffré, 2008) mais jamais l'intégralité de l'orthographe. Par ailleurs, comme le soulignent Lucci et Nazé (1979), être un bon scripteur n'est pas forcément l'être dans tous les domaines de l'écrit, notamment dans le cadre de la littéracie qui considère l'écrit en fonction de la diversité de ses usages. Ainsi, « *personne ne peut être à la fois journaliste, essayiste, écrivain, chercheur, etc. pour ne citer que quelques exemples concernant différentes sphères d'activité, différentes pratiques d'écriture.* » (Lucci et Nazé, 1979 : 12).

1.3.1.1 QUE SAIT-ON DES DIFFICULTÉS ORTHOGRAPHIQUES DES ADULTES FAIBLEMENT LITTÉRACIÉS ?

« L'adulte, surtout s'il est peu cultivé, éprouve des difficultés parfois insurmontables quand il s'agit de 'mettre l'orthographe' ; il craint les pièges (quand il ne les invente pas), applique mal les règles dont il se souvient encore, imagine parfois des subterfuges qui nuisent à la précision de l'expression ou, dans le pire des cas, soit se résigne à une orthographe fantaisiste qui parasite son message, soit se bloque devant l'écrit avant de s'en détourner. » (Lucci et Nazé, 1979 : 17).

Dabène (1987) a amorcé une réflexion sur l'écriture chez l'adulte, en soulignant l'existence de différents types d'écrits, dont la légitimité ne provient pas de sa forme ; l'écrit ne s'identifie pas uniquement à la littérature, « aux belles lettres » comme il l'explique dans la préface de son ouvrage ; mais de son objectif. Le geste d'écrire, tout comme l'oral, s'exerce dans un vaste champ lié aux contraintes de la situation dans laquelle il s'exerce et dans ce cadre, il existe

donc une multitude d'écrits. On distingue les écrits à dominance informelle, qui sont liés à un objectif de communication immédiate (comme la rédaction d'une liste de courses, d'un courrier électronique, etc.) ou ceux dont la forme est davantage contrainte parce qu'ils doivent répondre à des attendus de forme (rédaction d'une lettre administrative, etc.) dont le degré de formalisme peut être plus ou moins important.

Les travaux qui portent sur l'écriture des personnes faiblement littéraciés et sur l'analyse de leurs difficultés en lecture-écriture se sont principalement intéressés à leur rapport à l'écrit et à leurs représentations de l'écriture. Dabène (2000) a montré que même si les relations qu'ils entretiennent avec l'écrit, en réception comme en production, sont souvent compliquées, ils ont des pratiques réelles de lecture et d'écriture. D'ailleurs, cette prise en compte de la diversité des situations, parmi lesquelles la diversité des pratiques de lecture, d'écriture, du rapport à l'écrit et des compétences effectives à l'écrit (dont on sait très peu de choses), Dabène dénonçait en 2001 l'utilisation du terme « illettrisme », notamment parce qu'il permet à la société de généraliser en englobant une population sous un même terme, ayant pour conséquence d'occulter complètement la diversité des situations.

Les nombreux travaux dans ce domaine ont permis d'aborder cette question en termes de liens avec l'écrit et de compétences plutôt qu'en terme de rupture ou de carences (Penloup, 2000, 2006). Ces travaux ont permis d'inscrire le sujet dans un processus personnel d'appropriation de la langue écrite qui permet de rompre avec les discours généralistes sur l'illettrisme des années 1980, qui ne permettaient pas de rendre compte des rapports individuels que les sujets entretiennent avec l'écrit. L'analyse des pratiques effectives de l'écrit, en situation de production, des personnes considérées comme illettrées, a montré qu'elles ont des pratiques d'écriture, très souvent mésestimées parce qu'elles sont considérées comme illégitimes (Besse, 1997 ; Lahire, 1997). Leurs pratiques s'écartent néanmoins des situations d'écriture formelle, dont elles savent qu'elles les placent en situation d'insécurité linguistique, parce que les personnes en situation d'illettrisme sont conscientes de leurs difficultés. Dans une enquête réalisée en 1999, Frier (2002) a recueilli les discours de jeunes sortis du système scolaire et engagés dans différents dispositifs d'insertion. Dans ce contexte, les discours sur l'école convergent bien évidemment vers une incompréhension des attentes scolaires et des activités de lecture et d'écriture qui ont été vécues comme extrêmement contraignantes, tout comme un sentiment de manque de considération de la part des enseignants, qui ne comprenaient pas leurs difficultés. Mais malgré l'existence de difficultés à l'écrit, Frier (2002 : 89) souligne l'existence de « *pratiques braconnières, souterraines, mal connues, souvent inavouées* » qui doivent être valorisées car elles représentent un levier d'apprentissage certain en formation.

Par ailleurs, de multiples contradictions caractérisent les représentations de l'écrit chez les personnes qui ont un rapport difficile avec l'écrit, notamment sur ce que représente pour elles, le fait de bien écrire. De nombreux auteurs rapportent que dans les discours, les éléments qui se rapportent à la norme orthographique sont récurrents : le poids des valeurs transmises par l'école en lien avec les normes prescriptives, la surestimation de la norme orthographique dans les écrits ou encore une tendance à l'hypercorrection (Dabène, 1987 ; Lahire, 1997, Frier, 1997).

Ainsi, dans la lignée de ces auteurs qui ont recueilli des discours et cerné le rapport à l'écrit d'après l'observation des pratiques d'écriture, les travaux de Leclercq (1999, 2003, 2008) soulignent l'importance de modifier le regard porté sur ces publics. Dans une approche holistique du sujet, il s'agit désormais de valoriser les compétences plutôt que de souligner les manquements. Comme le montre Besse (2001), les personnes en situations d'illettrisme conservent de leur passage à l'école des connaissances et des savoir-faire à l'écrit qui constituent une base de connaissances qui sont propres à chaque personne et sur lesquelles il faut s'appuyer en formation, ce qui est conforté par les observations de Frier (2002). En formation d'adultes, les travaux de Leclercq (1997, 2003, 2008) et de Besse (2003, 2004) ont initié un « *renouveau pédagogique* » (Leclercq, 1997) qui permet une meilleure prise en compte des besoins à partir de leur savoir-faire et en lien avec leur rapport à l'écrit.

Néanmoins, il apparaît qu'aujourd'hui encore, le terme « illettrisme » renvoie à l'idée de « non-maitrise », que la société considère comme anormale. Peut-être cela provient-il du fait que c'est d'abord à l'école qu'est initiée l'écriture et qu'en ce sens écrire renvoie implicitement à des contraintes prescriptives dont la principale concerne les prescriptions orthographiques. En effet, dès son entrée dans l'écrit, l'apprenti scripteur est soumis aux normes fonctionnelles qui opèrent au niveau du fonctionnement linguistique de l'écrit (dont la syntaxe, l'orthographe lexicale et l'orthographe grammaticale). À l'âge adulte, s'ajoutent ce que Dabène (1987) qualifie de « *surnormes sociales* » qui régissent l'adaptation de l'écrit aux attentes sociales. On peut se demander si elles ne sont pas la conséquence d'une pression faite sur l'école, lieu où depuis toujours l'orthographe est instituée comme un enjeu majeur et représente déjà une source de tensions pour de nombreux élèves.

En 1987, Dabène soulignait que l'usage du scriptural comporte aussi ce qu'il nomme les « pratiques scripturales ordinaires » et en ce sens « savoir lire et écrire, ce n'est pas seulement comprendre ou traduire du sens à partir des graphies et de la connaissance de leur fonctionnement linguistique ; c'est connaître aussi, et maîtriser surtout, les situations dans lesquelles tel usage de l'écriture est socialement acceptable. » (Dabène, 1987 : 48). Selon les

situations de production, les finalités de l'écrit diffèrent et la focalisation des scripteurs (et des lecteurs) sur l'orthographe n'est pas toujours opportune. Comme le suggérait Dabène (1987 : 219) : « *Il suffirait peut-être que l'enseignement-apprentissage considère les pratiques ordinaires comme un lieu d'activité scripturale à part entière ?* », et que l'école, dans l'objectif de former des citoyens, s'accorde le droit de valoriser des écrits qui sont moins normés, en fonction de l'objectif sous-jacent qui les motive.

1.3.1.2 L'APPORT DE LA RECHERCHE SUR LA CONNAISSANCES DES COMPÉTENCES À L'ÉCRIT DES ADULTES FAIBLEMENT LITTÉRACIÉS

En dépit d'une alphabétisation systématique, prolongée et établie depuis longtemps en France, il subsiste au XXI^{ème} siècle, une population en difficulté avec l'écrit dont les caractéristiques et les compétences à l'écrit sont mal connues. L'apport des dernières enquêtes sur la littéracie (comme IVQ) a permis de rendre compte de la diversité de profils des personnes en difficulté face à l'écrit et de l'inégalité des conséquences que leurs difficultés ont sur leur quotidien (Guérin-Pace, 2009). Et c'est justement le fait que leur profil soit si varié (toutes les tranches de la population sont touchées) et que leur insertion sociale soit si différente (certains sont en activité professionnelle alors que d'autres vivent en marge de la société), qu'il est si difficile de savoir avec précision en quoi consistent leurs difficultés.

Quelques recherches se sont intéressées à ce public, notamment dans le domaine de la psychologie cognitive, mais elles ont été réalisées auprès d'un nombre restreint de sujets et souvent dans un contexte en particulier (jeunes en échec scolaire, dyslexiques, etc.) qui empêche la généralisation des résultats. Eme (2010 : 759) recense les études menées en psychologie cognitive sur des personnes considérées comme illettrées et conclut par : « *Functional illiteracy could, in many cases, be the consequence of dyslexia that has slipped through the diagnostic net in childhood* ». Les résultats recensés par Eme (2010) rapportent néanmoins que plusieurs caractéristiques sont communes aux différentes études, quelle que soit la population interrogée : des connaissances insuffisantes dans le domaine de la phonologie et de la morphologie, de bonnes connaissances en compréhension (meilleures que celles des enfants qui correspondent au même niveau de compétence en lecture) et le fait que pour écrire, ces publics semblent davantage faire appel à leurs connaissances lexicales. D'autres rares données décrivent des profils de performance inattendus. Nous citons par exemple les travaux de Holmes (1996), dont les résultats ne peuvent pas non plus être étendus (au vu de la taille de l'échantillon) mais qui illustrent la diversité des profils de scripteurs en difficulté.

Holmes (1996) présente dans une étude de cas, le profil d'une étudiante qui présente un Quotient Intellectuel verbal moyen et de faibles capacités phonologiques mais qui a obtenu d'excellentes performances orthographiques dans les tests. Cette étudiante apparaît parmi les 20% d'étudiants les meilleurs en compréhension de texte et montre une très bonne maîtrise de l'orthographe. Cette étude interroge sur les relations entre lecture et production écrite, même si ici, le cas semble plutôt pathologique. Fayol (2013) rapporte que d'autres études montrent que certains adultes qui n'ont aucune difficulté en lecture et en compréhension présentent en revanche un niveau de production écrite faible (souvent corrélé avec un passé de dyslexique). Il est également admis qu'il existe une « asymétrie » entre les performances en lecture et celles en orthographe ; cette asymétrie étant aujourd'hui attestée dans tous les systèmes orthographiques : « [...] certains individus se situent de manière extrême sur cette asymétrie : ils lisent parfaitement mais rédigent en produisant de très fréquentes erreurs. » (Jaffré et Fayol, 2014). Les études actuelles en psycholinguistique (Lété et al., 2008 ; Lété, 2018) montrent que la corrélation entre la lecture et les performances orthographiques s'inscrivent plutôt en terme de fréquence d'exposition aux mots (et cette corrélation s'établit donc probablement en lien avec le temps consacré à l'activité de lecture). Les exigences en matière d'orthographe sont en effet bien plus fortes en production écrite qu'en lecture, à cause du nombre de configurations possibles pour transcrire un même phonème. On peut donc être un lecteur efficace mais un scripteur dont l'orthographe pose problème. En revanche, être un lecteur régulier (et cela depuis l'enfance) semble prédisposer à de bonnes connaissances en matière d'orthographe.

Selon le postulat de l'existence d'un lexique orthographique disponible en mémoire, subsiste la question d'un lexique orthographique unique, qui serve autant pour les activités de lecture que pour celles d'écriture.

Holmes (1994) a comparé les productions orthographiques de deux groupes d'étudiants dont le niveau d'orthographe est très contrasté : des bons en orthographe vs des faibles en orthographe mais qui ont les mêmes pratiques de lecture. Les résultats montrent que si les erreurs d'ordre phonographiques ne varient pas entre les deux groupes, les étudiants faibles en orthographe font en revanche davantage d'erreurs sur les formes orthographiques qui présentent une forme orale différente de la forme écrite. Dans une autre étude, Holmes a demandé à des adultes de reproduire sous la dictée des mots complexes puis d'évaluer leurs réponses et enfin de les corriger. Cette étude montre dans un premier temps, que les enquêtés utilisent des formes graphiques qui semblent être stabilisées en mémoire et qu'ils ont reproduits à plusieurs reprises au cours de l'exercice, même lorsque celles-ci sont erronées. La plupart se déclare satisfaits de leurs productions, même lorsque celles-ci sont

erronées. Ils n'ont par ailleurs fait que peu de modifications lorsqu'on leur en a donné la possibilité et ces modifications ont apporté presque autant d'améliorations que d'erreurs (39% d'améliorations vs 36% d'erreurs).

Ces différents travaux vont plutôt dans le sens de l'existence d'un seul lexique stabilisé en mémoire, qui serait constitué de formes stables conformes ou non à l'orthographe normée.

D'autres études comportementales (brièvement présentées par Jaffré et Fayol, 2014) ont utilisé l'amorçage pour déterminer sous quelle forme orthographique les mots sont mémorisés. L'amorçage se produit lorsqu'un individu perçoit, reconnaît, prononce ou écrit un mot plus rapidement parce qu'il a au préalable lu cet item (il s'agit d'un amorçage de répétition) ou parce qu'il a lu un autre item qu'il relie à celui qu'il doit écrire parce que sa forme orthographique ou phonologique est proche (ou même parce que cet item est proche d'un point de vue sémantique). Les résultats de ces études montrent que les items rencontrés au préalable sont mieux restitués, ce qui suggère que les participants utilisent bien un lexique orthographique unique, en lecture comme en écriture.

1.3.2. LES ENSEIGNANTS ET LA NORME ORTHOGRAPHIQUE

Il existe toute une tranche de la population qui occupe un emploi au sein duquel il n'est socialement pas autorisé d'être en difficulté avec l'orthographe, en particulier parce que leur niveau d'étude doit signifier de bonnes compétences dans ce domaine. Par exemple, la société n'accepte pas qu'un enseignant puisse faire des erreurs d'orthographe. Pour autant, ceux qui sont censés être à l'aise dans ce domaine le sont-ils vraiment ? Si l'on reprend les conclusions des spécialistes de l'orthographe, aucun scripteur, même expert, ne maîtrise l'intégralité de l'orthographe du français.

Le scripteur, nous l'avons vu notamment pour la morphologie flexionnelle, a de bonnes raisons d'être perdu. Entre l'application de règles dans des configurations de phrases qui sont complexes et les cas particuliers qui divisent même les puristes (« DE BONS LIVRES, J'EN AI BEAUCOUP⁶ – LU ou LUS ? »), le scripteur ne peut que douter des choix qu'il fait, tout en craignant la « faute » qui le dévalorisera aux yeux de son lecteur (Lucci et Nazé, 1979).

En 2008, Péret, Sautot et Brissaud ont mené une enquête sur 4 ans auprès de futurs professeurs des écoles (d'abord en formation) puis lauréats du concours, en poste. Cette étude avait pour objectif de mettre en lien leurs performances orthographiques avec leurs

⁶ Les graphies qui relèvent d'exemples sont reproduits en majuscules dans le manuscrit, pour les différencier des graphies qui ont été produites par les enquêtés et que nous reproduisons en minuscules.

discours sur l'orthographe lorsqu'ils sont en formation. Dans un deuxième temps, l'objectif était de relier ces observations avec leurs pratiques professionnelles effectives dans le domaine orthographique, lorsqu'ils entrent dans le métier.

Les discours des étudiants concernant leurs souvenirs d'apprentissage évoquent le Bled avec l'acquisition de règles qu'il fallait apprendre par cœur et les dictées pour s'entraîner. Cet exercice représente un souvenir commun aux enquêtés et apparaît souvent comme un exercice ludique. Seuls quelques stagiaires en gardent de très mauvais souvenirs. Il apparaît dans les discours que l'apprentissage de l'orthographe n'est pas un domaine réservé à l'école. En effet, l'implication de la mère de famille (plutôt que le père) qui relaye l'école, le soir ou pendant les vacances est évoquée de manière très fréquente :

« À les entendre raconter qu'ils ont fait la même chose à l'école et dans la famille, il est donc légitime de se demander quelle est, pour eux, la spécificité de l'enseignement de l'orthographe. Est-ce vraiment une affaire de professionnel ? » (Péret et al., 2008 : 3).

Lorsqu'ils ont été questionnés sur leurs difficultés orthographiques, il ressort des entretiens que c'est surtout l'orthographe lexicale qui leur poserait problème alors que l'analyse de leurs copies de concours blanc montre que leurs difficultés portent plutôt sur la syntaxe et la morphographie. Dans ce domaine, les erreurs relevées concernent particulièrement l'omission ou l'adjonction de marques de genre et de nombre. Comme le soulignent Péret et al. (2008 : 4), il existe un décalage entre ce que les étudiants croient être difficile pour eux et ce qui leur pose réellement problème dans leurs copies, ce qui a pour conséquence que les stagiaires *« [...] ignorent dans une même dynamique un domaine de variation important qui les concerne eux et leurs élèves. »* (Péret et al., 2008 : 4).

Enfin, sans surprise, les stagiaires se sentent investis d'une mission dans le domaine orthographique et adoptent une attitude très normative (*« gardien de la norme »*) car ils ont conscience de l'importance sociale accordée à l'orthographe, tout en ayant une connaissance partielle du système linguistique : *« Cela ne doit pas être sans rapport avec une forme de culpabilité fréquente dans ce domaine chez les enseignants, ou a contrario avec des attitudes très normatives qui les protègent d'aléas théoriques qu'ils maîtrisent mal. »* (Péret et al., 2008 : 4). Pourtant, dans leurs discours, ils ne font pas référence à la manière dont le système orthographique s'apprend, ni aux processus cognitifs qui sont à l'œuvre lorsque les élèves apprennent, domaine de la psycholinguistique qui semble leur être relativement étranger.

La dernière partie de l'enquête consistait à observer les pratiques de classe des stagiaires lauréats du concours lorsqu'ils travaillent l'orthographe avec leurs élèves. Les chercheurs

rappellent que cette partie de l'enquête a été difficile à conduire, avec de nombreux désistements qui semblent liés à l'insécurité linguistique dans laquelle se trouvent certains jeunes enseignants lorsqu'ils commencent à enseigner. Cette insécurité les amène à redouter de montrer leur pratique professionnelle, d'autant plus que les chercheurs étaient également leurs anciens enseignants. S'ajoute ici la question de l'image de soi qui s'inscrit en décalage avec l'image qu'ils voudraient peut-être montrer en tant que professionnels et qui est liée avec leurs représentations de l'orthographe, très normative.

1.4. CONCLUSION SUR L'ÉCRITURE DES ADULTES DANS LE CONTEXTE DE LA LITTÉRACIE

Selon Fayol et Jaffré (2008), lire et écrire en respectant l'orthographe ne posent pas les mêmes difficultés. Généralement, en lecture, les lettres renvoient à un nombre restreint de sons alors qu'en écriture, les phonèmes et les graphèmes ne se trouvent pas dans une relation de bi-univocité⁷. Lorsqu'il orthographe, le scripteur est soumis à de nombreuses contraintes qui relèvent du système (par exemple la gestion de la morphologie flexionnelle) ou qui lui sont propres (par exemple ses capacités d'attention et de mémoire au moment où il orthographe) et c'est ainsi que même les plus experts en écriture commettent des erreurs, même si celles-ci diminuent avec l'augmentation du niveau d'expertise (Fayol, 2013).

De nombreuses études montrent qu'il existe des déterminants sociaux favorables aux difficultés à l'écrit rencontrées par les scripteurs (faible niveau culturel des familles, conditions familiales difficiles durant l'enfance, etc.), il est aujourd'hui admis que cela ne peut pas expliquer l'importance du phénomène (Lahire, 1999 ; Guérin-Pace, 2009 ; Fayol, 2013). En effet, si l'apprentissage du langage oral est une prédisposition biologique qui permet son acquisition facilement lorsqu'il n'y a pas de pathologie associée, l'apprentissage de l'écrit en revanche nécessite un enseignement explicite dont la difficulté d'acquisition est aussi en lien avec le degré de complexité du système orthographique dont il dépend. Catach (1995) souligne que dans la plupart des pays qui nous entourent (par exemple en Italie, en Espagne), un enfant de 8 ans peut écrire un texte sans difficultés alors que les petits Français se heurtent à de nombreuses difficultés qui ne sont parfois toujours pas acquises à l'âge adulte. Il y a 25 ans, Catach s'interrogeait sur l'origine des difficultés en orthographe des Français et avait avancé plusieurs hypothèses : une orthographe trop difficile, une orthographe mal enseignée ou enseignée dans de mauvaises conditions, mais ce serait surtout parce qu'elle est « *mal conçue, de la base au sommet, de façon élitiste, maximaliste et dogmatique* » (Catach,

⁷ Se référer à la section 3 du chapitre 3 qui propose une description du système orthographique du français.

1995 : 6). Concernant les conditions de son enseignement, Catach soulignait également que la pédagogie doit être envisagée dans une perspective de réussite :

« Toute mesure d'expérience acquise devrait être une mesure de progrès. Il viendra bien, le temps où disparaîtront la notion de « fautes » et le zéro en orthographe ».

Fayol et Jaffré (2008) se sont également interrogés sur les facteurs qui affectent la nature et la fréquence des erreurs d'orthographe chez les adultes et il a notamment cherché à savoir si les erreurs étaient différentes d'un individu à l'autre ou s'il y avait des similitudes. Leurs conclusions rejoignent celles de Catach, à savoir que le système orthographique contient des zones complexes qui passent nécessairement par un apprentissage et un entraînement régulier. En ce sens, certaines difficultés orthographiques affectent tout le monde. Mais il y a également des personnes pour qui certaines difficultés apparaissent de manière plus forte que chez d'autres personnes : *« Certaines catégories de personnes sont faibles en lecture ou en orthographe. Par ailleurs, un nombre restreint de mots ou d'accords pose problème à tous les individus. »* (Fayol, 2008 : 125).

Pourtant, la faute d'orthographe est encore extrêmement stigmatisante et certaines idées reçues concernant la norme orthographique comme témoignage de l'héritage culturel français perdurent aujourd'hui :

« La tyrannie de la norme orthographique, qui s'exprime surtout depuis le 19^{ème} siècle, résulte de l'influence conjuguée de l'enseignement obligatoire et de la démocratisation sans cesse grandissante de l'écrit. Les siècles précédents ont en revanche connu des usages plus variés qui contrastent avec les idées reçues sur la permanence de la norme. [...] Auparavant la variation était la règle, même dans les noms propres. » (Jaffré, 2008 : 51).

Les manquements en matière d'orthographe sont toujours socialement dévalorisés et encore souvent rattachés à l'image des mauvais élèves, qui n'ont pas fait d'effort à l'école. Les adultes deviennent alors responsables de leurs difficultés en orthographe. En France, on peut avoir une culture générale défaillante mais la société ne tolère pas les carences en matière d'orthographe :

2. L'ÉVALUATION DES ADULTES À L'ÉCRIT ET PLUS PARTICULIÈREMENT EN ÉCRITURE

Depuis les années 1990, la demande d'information des autorités publiques sur l'état des compétences à l'écrit des adultes dans les pays industrialisés s'est fortement accrue dans un contexte de compétitivité et de concurrence sur le marché international du travail. Le

développement des nouvelles technologies joue une part importante dans cette demande. Avec la diversité des supports aujourd'hui disponibles, l'écrit n'aura jamais été autant sollicité. Avec internet (mails, forums, réseaux sociaux, etc.) et la téléphonie (SMS, etc.), l'écrit devient en effet un mode de communication facile, accessible et privilégié.

Si l'écrit constitue aujourd'hui un moyen de communiquer privilégié pour la diversité de ses supports et pour son efficacité (discussions écrites instantanées), il devient une compétence incontournable que les états ont cherché à évaluer chez les adultes. La tâche est complexe, car c'est auprès des enfants que l'écrit a toujours été traditionnellement évalué et les modalités d'évaluation pratiquées ne peuvent pas être transposées chez des adultes sortis du système scolaire. Ainsi, pour définir et comparer les compétences des populations dans le domaine de l'écrit, différentes enquêtes se sont donné pour objectif d'évaluer le niveau de littéracie des adultes (et non des compétences scolaires). Ces enquêtes ont également pour objectif de rendre compte de la proportion de personnes qui se trouvent en difficulté avec l'écrit pour définir le taux d'illettrisme dans les pays évalués.

En France, la comptabilisation des personnes en difficulté avec l'écrit (plus généralement en lecture-écriture) n'est pas une préoccupation récente et est fortement liée à l'histoire de l'alphabétisation. Selon Besse et al. (2009), l'alphabétisation de la population a toujours figuré parmi les priorités nationales et les premiers rapports qui mentionnent la capacité d'écriture des populations datent du 19^e siècle (Furet et Ozouf, 1977). Pour rendre compte des efforts à fournir dans ce domaine, une évaluation systématique des connaissances (parmi lesquelles figurent la lecture et l'écriture) a été mise en place pendant le service militaire. On retrouve (dès 1829) dans le dictionnaire pédagogique de Ferdinand Buisson (1911), le nombre de conscrits sur cent qui savaient au moins lire à leur entrée au service militaire. Les conscrits considérés comme illettrés étaient alors sanctionnés par l'augmentation d'une année de leur service obligatoire. Cette loi militaire de 1872 avait pour objectif d'inciter les jeunes Français à augmenter leur niveau d'instruction (et donc le taux de scolarisation ainsi que sa durée). Un test demeure aujourd'hui encore lors de la Journée Défense et Citoyenneté (Ministère de la Défense, 2019) destiné au repérage des jeunes en difficulté dans le domaine de la lecture (voir section 2.1.2).

Mais comme le montrent Degorre et Murat (2009), l'évaluation des compétences des adultes (en général) dans le cadre d'enquêtes réalisées à domicile est un domaine récent qui reste à investiguer, les évaluations dans le domaine de l'écrit étant traditionnellement réalisées dans le cadre scolaire. Depuis 1985, les enquêtes internationales mises en place et dont nous présentons les principales ci-dessous, ont permis de progresser dans la mesure des

compétences. Nous le verrons, certaines enquêtes dont les résultats ont été remis en question à cause de fortes limites méthodologiques ont par ailleurs permis de construire les dernières enquêtes (comme IVQ pour la France ou LEO pour l'Allemagne) en améliorant la pertinence des critères d'évaluation et le mode de recueil des données : « *La genèse du protocole d'IVQ a été intimement liée aux écueils rencontrés par ces enquêtes et aux expertises qui en ont découlé.* » (Degorre et Murat⁸,2009 : 6).

2.1.TOUR D'HORIZON DES ENQUÊTES SUR LA LITTÉRACIE RÉALISÉES AUPRÈS D'ADULTES ET DE JEUNES ADULTES

Dans cette partie, qui n'est pas exhaustive, nous présentons quelques enquêtes qui nous paraissent représentatives de l'évaluation des compétences en littéracie (de dimension internationale et nationale) auprès de deux types de publics : les adultes et les jeunes adultes.

Nous commençons cette présentation par quatre enquêtes considérées comme pionnières dans l'évaluation des compétences en littéracie des adultes. Tout d'abord YALA⁹ et NALS¹⁰ conduites aux États-Unis, qui sont les premières enquêtes d'envergure nationale réalisées sur les compétences en littéracie des adultes, qui ont ensuite donné lieu à IALS puis ALLS d'envergure internationale. Dans un deuxième temps, nous nous centrons sur les enquêtes réalisées en France, qui représentent selon nous, des enquêtes incontournables (il existe en effet de très nombreuses enquêtes réalisées nationalement ou plus localement pour évaluer la littéracie). Nous présentons Infométrie qui est l'une des premières enquêtes conduites en France ayant permis de déterminer le taux d'illettrisme puis Information et Vie Quotidienne (IVQ), plus récente et dont le protocole n'a pas d'équivalent aujourd'hui en France. Toutes les deux ont été conduites en partenariat avec l'Agence Nationale de Lutte contre l'illettrisme. L'enquête IVQ fera l'objet d'une section (2.2, chapitre 1), car nous présentons plus longuement son protocole ; (une partie des dictées qui constituent le corpus de notre étude étant issu de cette enquête). Nous nous pencherons également sur deux enquêtes conduites auprès de publics plus jeunes : PISA (de dimension internationale) et les tests des Journées de la Citoyenneté (JDC).

⁸ Fabrice MURAT, chercheur et statisticien à l'INSEE a travaillé sur l'évaluation des adultes à l'écrit et sur l'analyse des résultats des enquêtes réalisées par l'INSEE, dont IVQ. Aussi, nous nous appuyons préférentiellement sur ses travaux.

⁹ « Young Adult Literacy Assessment » (YALA).

¹⁰ « National Adult Literacy Survey » (NALS).

2.1.1. DES ENQUÊTES À GRANDE ÉCHELLE

2.1.1.1 LES PREMIÈRES ENQUÊTES INITIÉES AUX ETATS-UNIS

C'est aux États-Unis qu'ont été initiées les premières enquêtes à grande échelle avec « Young Adult Literacy Assessment » (YALA) et « National Adult Literacy Survey » (NALS).

YALA est la première enquête à grande échelle réalisée sur les compétences en littéracie des adultes. Réalisée aux États-Unis à partir de 1985, l'enquête évalue les compétences en lecture et en compréhension de 3600 jeunes adultes âgés de 21 à 25 ans. Les tests avaient pour objectif de mesurer la capacité à comprendre et utiliser de l'information à partir de documents rédigés comme des textes narratifs, des poésies mais aussi à partir de tableaux, graphiques, etc. Des tests de numératie permettaient de vérifier la maîtrise de notions d'arithmétique utiles pour le quotidien : additionner, soustraire, multiplier (Irwin, 1986). YALA est également le premier dispositif d'évaluation qui propose différents niveaux de difficulté dans les exercices, afin de tenir compte du niveau de compétence des enquêtés. Cette prise en compte est liée à l'avancée des travaux en psychologie sur l'évaluation d'adultes qui ont montré l'importance de la prise en compte de certains paramètres pour ne pas biaiser les résultats. Il s'agit par exemple de prendre en compte les différences entre les individus pour ne pas décourager les enquêtés si les tests sont trop durs ou trop faciles. Il s'agit également de prendre en considération l'importance de la dimension psychologique des adultes que l'on souhaite enquêter, en lien avec l'acceptation d'être évalué quand on est adulte (alors qu'en milieu scolaire les enfants sont habitués à l'être) et à la peur du regard lorsqu'on évalue des compétences censées être acquises. YALA est la première enquête qui propose une différenciation des exercices, qui ont été déclinés en différents niveaux de complexité. Par exemple, trouver l'heure d'une réunion dans un agenda (épreuve dite « simple ») ou choisir l'heure de départ d'un bus en fonction de l'heure d'arrivée souhaitée, à partir d'une grille d'horaires (épreuve plus complexe). Après lecture des conclusions, il semble que le test d'écriture ait été succinct (« *Virtually all young adults today demonstrate the ability to sign their name, thus making them literate according to standards applied to information available more than 100 years ago.* ») et que les enquêtés ont davantage été interrogés sur leurs compétences en lecture, en compréhension et en numératie. » (Kirsch et Jungeblut, 1986 : 63). Les résultats encourageants observés dans l'enquête YALA ont donné naissance à l'enquête « National Adult Literacy Survey » (NALS), conduite en 1989-1990 auprès d'un échantillon plus vaste, qui permette de couvrir l'ensemble de la population adulte des États-Unis.

2.1.1.2 DEUX ENQUÊTES INTERNATIONALES PIONNIÈRES : IALS ET ALLS

L'enquête « International Adult Literacy Survey » (IALS) est la première enquête de dimension internationale, conçue à des fins de comparaisons entre les pays. Coordonnée par l'Organisation de Coopération pour le Développement Économique (OCDE), elle a été réalisée dans une vingtaine de pays entre 1994 et 1999 puis renouvelée à partir de 1999 avec l'enquête « Adult Literacy and Life Skills Survey » (ALLS) élargie à d'autres compétences comme la résolution de problèmes (Degorre et Murat, 2009).

Le classement international de la France à l'issue de sa participation à IALS en 1994 était très mauvais et les premiers résultats publiés ont avancé le chiffre de 40 % de Français illettrés : « *Trois Français sur quatre se situeraient à un niveau tout juste suffisant pour obtenir un emploi semi-qualifié* » (Bastyns, 2012 : 26). D'après les résultats de l'enquête, 27% des moins de 26 ans se situeraient dans le niveau le plus bas de littéracie et dans la catégorie des plus faibles lecteurs. Suite à ces chiffres qui laissaient penser que les résultats pouvaient être biaisés, la France s'est retirée de l'enquête et de nombreuses expertises¹¹ ont été demandées pour comprendre l'origine de ces résultats. Celles-ci ont en effet relevé un certain nombre de dysfonctionnements qui mettent en doute la fiabilité de l'enquête, notamment en matière de méthodologie (Besse et Guérin-Pace, 2002). IALS a néanmoins servi d'enseignement pour la mise en place d'IVQ et elle a permis de prendre conscience en France, de certaines faiblesses dans le domaine de l'évaluation des adultes et de leurs compétences en littéracie.

L'un des premiers enseignements a été de prendre conscience de la nécessité de concevoir une enquête nationale, conçue pour le français et qui élimine ainsi les problèmes de traduction et de différences culturelles. D'un point de vue technique, les expertises relèvent en effet une méthodologie inadaptée à une enquête internationale qui visait à comparer les performances entre les pays. Besse et Guérin-Pace (2002) ont révélé des problèmes de traduction et d'inadaptation de certains exercices aux différents pays, laissant subsister un doute sur la comparabilité internationale des données. Blum et Guérin-Pace (2000) reprochent en particulier aux concepteurs de l'étude de n'avoir pas pris en compte la diversité des pratiques sociales de l'écrit à l'échelle internationale, dans un contexte de pluralité culturelle. Par ailleurs, la diversité des compétences a été prise en compte de manière peu avantageuse pour ceux qui ont montré des difficultés au test du test. IALS ne proposait pas de

¹¹ Pour prendre connaissance de ces travaux d'expertise et de contre-expertise, nous renvoyons aux travaux de Dickes et Vrignaud (1995) et Blum et Guérin-Pace (2000).

tests d'orientation mais un « livret préliminaire » (composé de 6 questions¹²) destiné non pas à adapter la suite du test au niveau de compétences de la personne, mais à l'écartier de l'enquête jugée trop complexe pour elle. Ainsi, les enquêtés qui se sont trouvés en difficulté sur les premières questions ont arrêté le test et ont été classés dans le niveau de compétence en littéracie le plus bas. Selon Murat (2008) moins de 5% des Français ont été ainsi écartés de l'enquête.

Les expertises ont également montré un nombre important de non-réponses, qui semble peut témoigner d'un protocole inadapté aux enquêtés. Sur ce point, Besse et Guérin-Pace (2002) critiquent la longueur du test ainsi que l'absence de transition entre les différents exercices, les enquêteurs ayant pour consigne de rester silencieux : « *Il est probable que ce type de consigne ait contribué à renforcer l'atmosphère d'étrangeté déjà associée au caractère singulier de cette enquête* » (Degorre et Murat, 2009 : 8). Enfin, la procédure de correction des réponses pose problème car selon Blum et Guérin-Pace (2000), elle n'a pas envisagé l'intégralité des réponses possibles (et justes), ce qui a conduit à classer comme incorrectes, de nombreuses réponses correctes.

Les résultats des différentes expertises ont justifié le retrait de la France de cette enquête et son refus de participer à l'enquête ALLS, car son protocole a été reproduit à l'identique malgré les critiques qui avaient été adressées.

2.1.2. LES ÉVALUATIONS AUPRÈS DE PUBLICS PARTICULIERS

L'évaluation des compétences à l'écrit est souvent liée à un besoin de connaissance d'un public en particulier. De nombreuses enquêtes ont ainsi un public cible : les adultes en difficulté avec l'écrit pour déterminer le taux d'illettrisme d'une population (Infométrie, IVQ), l'évaluation d'une classe d'âge (PISA, tests de JDC, etc.).

2.1.2.1 L'ENQUÊTE INFOMÉTRIE : DÉTERMINER LE TAUX D'ILLETTRISME

L'enquête Infométrie (1988) est une des premières études quantitatives conduites en France, destinée à déterminer le nombre d'adultes qui se trouvent en situation d'illettrisme. Elle a été menée à la demande du Groupe Permanent de Lutte contre l'Illettrisme (GPLI)¹³ sur un échantillon représentatif de la population française âgée de plus de 18 ans (Hauser, 1988). Dans cette enquête figure une épreuve d'écriture qui consistait à reproduire la

¹² Nous ne connaissons pas le contenu de ce livret préliminaire.

¹³ Aujourd'hui Agence Nationale de Lutte contre l'Illettrisme (ANLCI).

phrase suivante sous la dictée : « *Depuis 1984, la France possède davantage de chaînes de télévision. Pourtant la redevance a baissé de 72 francs.* ». Pour réussir l'exercice, les enquêtés devaient réussir à orthographier 67% des mots (Jeantheau, 2015). Besse (1995) rapporte que cet exercice montre que 17,8 % de la population connaît des « *difficultés plus ou moins importantes à écrire* », 11,5% « *maitrisent mal l'écriture* » et 10,3 % « *maitrisent mal la lecture* ». Enfin, 6,3% de la population française connaîtrait des difficultés importantes pour lire. En conclusion, l'étude avance que 20% de la population (soit 1 Français sur 5) serait en situation d'illettrisme. Hauser (1988) souligne que ce chiffre serait en dessous de la réalité puisqu'il existe toute une frange de la population que l'on ne peut pas atteindre par le biais de sondages et dont on peut estimer très probablement qu'elle se trouve en difficulté avec l'écrit (marginaux, gens du voyage, immigrés, etc.).

Nous présentons dans la section suivante, deux enquêtes qui permettent d'évaluer les compétences d'une classe d'âge : l'enquête PISA (qui permet une comparaison internationale des élèves âgés de 15 ans) et les tests de la Journée Défense et Citoyenneté (JDC).

2.1.2.2 ÉVALUER UNE CLASSE D'ÂGE : L'ENQUÊTE INTERNATIONALE PISA (MEN ; OCDE)

Nous choisissons de présenter l'enquête PISA car elle propose une dimension internationale qui n'a pas d'autre équivalent pour évaluer les compétences en littéracie de jeunes qui sont encore dans le système scolaire mais dont la sortie, pour certains, est imminente. Dans PISA, on ne parle pas encore d'illettrisme ; en effet ce terme qui qualifie les personnes en grande difficulté avec l'écrit n'apparaît pas lorsque les publics évalués sont âgés d'au moins 16 ans, comme dans les tests de la JDC ou dans IVQ. Cela est lié à la définition officielle de l'illettrisme proposée par l'ANLCI dans le cadre national de référence : « *L'illettrisme qualifie la situation de personnes de plus de 16 ans [...]* » (ANLCI, 2019).

PRÉSENTATION DE L'ENQUÊTE ET DU PROTOCOLE

Le Programme international pour le suivi des acquis des élèves¹⁴ (PISA) est une évaluation qui a été créée par l'Organisation de coopération et de développement économiques (OCDE) pour mesurer l'efficacité des systèmes éducatifs. De dimension internationale, elle permet de comparer les performances d'élèves d'une même classe d'âge qui sont issus de différents environnements d'apprentissage. PISA est ainsi la plus grande évaluation en littéracie qui existe dans le domaine de l'éducation : elle évalue les compétences des élèves âgés de 15 ans dans le domaine de la lecture, de la compréhension de l'écrit, des sciences et des

¹⁴ L'acronyme PISA vient de l'anglais « Programme for International Student Assessment ».

mathématiques et la capacité de ces jeunes à mobiliser leurs connaissances pour les utiliser dans des situations proches de la vie quotidienne. PISA a également pour objectif de mieux comprendre l'impact des politiques éducatives de chaque pays.

Cette enquête se déroule tous les 3 ans : la première édition a eu lieu en 2000 dans 32 pays et sa dernière édition a eu lieu très récemment en 2018, dans 79 pays. Certaines épreuves ont été reprises du protocole d'IALS.

En France, l'enquête est organisée par le ministère de l'Éducation Nationale selon le protocole défini par l'OCDE. En 2018, en France métropolitaine et Outre-mer, 252 établissements (publics ou privés sous contrat, collège ou lycée agricole, général, technologique ou professionnel) ont été tirés au sort pour constituer un échantillon représentatif du système éducatif scolaire, avec 30 élèves choisis aléatoirement par établissement. Ce sont ainsi 7000 élèves qui ont passé les tests en 2018.

Le protocole d'évaluation dure 3h30 réparties de la manière suivante :

- 2 heures consacrées aux tests scolaires en compréhension de l'écrit, culture mathématique et culture scientifique ;
- 1 questionnaire de 45 minutes axé sur le milieu socio-culturel de l'élève ;
- 1 questionnaire de 45 minutes axé sur les techniques d'information et de communication.

Dans l'enquête PISA, les compétences à l'écrit concernent l'évaluation de la compréhension de l'écrit (et donc aussi la lecture) : capacité à comprendre, utiliser et interpréter des textes écrits. Les tests ont été réalisés sur ordinateur et la capacité d'écriture n'a pas été testée. Le détail du protocole indique que depuis 2006, le domaine majeur d'évaluation est la **culture scientifique** (la compréhension écrite étant considérée comme un domaine mineur - OCDE, 2015), dont les tests ont été entièrement remaniés depuis l'édition de 2015. Pour la compréhension de l'écrit, les tests n'ont pas évolué depuis 2009.

Pour concevoir le test de compréhension de l'écrit, l'OCDE a retenu la définition suivante :

« Comprendre l'écrit, c'est non seulement comprendre et utiliser des textes écrits, mais aussi réfléchir à leur propos et s'y engager. Cette capacité devrait permettre à chacun de réaliser ses objectifs, de développer ses connaissances et son potentiel, et de prendre une part active dans la société. » (Cadre d'évaluation et d'analyse de l'enquête PISA 2015, OCDE 2018 : 55).

L'objectif de ce test est donc d'évaluer la capacité des élèves à comprendre l'écrit, l'utiliser et y réfléchir. L'enquête utilise des textes d'organisation classique (constitués de phrases et de

paragraphes), de types narratifs, informatifs et argumentatifs, ainsi que des textes qui présentent l'information sous d'autres formes (listes, formulaires, graphiques, schémas). Les épreuves ne sont pas des tests qui permettent d'évaluer les compétences en lecture (les concepteurs du test sont partis du principe que la plupart des élèves âgés de 15 ans ont acquis ces compétences). Ils sont axés sur la capacité à trouver et extraire l'information, comprendre le sens global d'un texte, développer une interprétation et réfléchir au contenu et à la forme d'un texte.

Différentes situations de lecture ont été utilisées :

- lecture à usage privé : extraits de romans, lettres personnelles, biographies ;
- lecture à usage public : documents et avis officiels ;
- lecture à des fins professionnelles : les manuels et rapports ;
- lecture scolaire : manuels scolaires et les fiches d'exercices.

Suite à leur résultat au test de compréhension écrite, les élèves sont classés en 7 niveaux (1A à 6) dont nous présentons pour la compréhension écrite les deux niveaux les plus éloignés :

- niveau 1A : score minimum de 335

« À ce niveau, les tâches demandent aux élèves de localiser un ou plusieurs fragments d'information explicites, d'identifier l'idée principale d'un texte portant sur un thème familier ou de reconnaître l'intention de son auteur, ou d'établir un lien simple entre des informations présentes dans le texte et des connaissances courantes. Dans ces tâches, les informations pertinentes sont saillantes et il n'y a guère, voire pas du tout, d'informations concurrentes. Les élèves sont explicitement orientés vers les éléments pertinents de la tâche et du texte. »

- niveau 6 : score minimum de 698

« Pour mener à bien les tâches de ce niveau, les élèves doivent procéder à des inférences multiples, se livrer à des comparaisons et opposer des contrastes, dans le détail et avec précision. Pour y répondre correctement, les élèves doivent bien comprendre un ou plusieurs textes en profondeur, voire intégrer des informations provenant de plusieurs textes. Dans certaines tâches, les élèves sont confrontés à des idées qui ne leur sont pas familières, en présence d'informations concurrentes saillantes, ou doivent produire des catégories abstraites à des fins d'interprétation. Dans certaines tâches de réflexion et d'évaluation, les élèves doivent formuler des hypothèses ou se livrer à une évaluation critique à propos d'un texte complexe sur un thème qui ne leur est pas familier, compte tenu de nombreux critères ou perspectives, en s'appuyant sur leur compréhension approfondie de notions extérieures au texte. À ce niveau, les tâches de localisation et d'extraction présentent deux caractéristiques : les élèves doivent faire des analyses d'une grande précision et doivent être très attentifs à des détails peu visibles dans les textes. »

Nous présentons à présent brièvement les résultats obtenus par la France pour l'édition 2015 de l'enquête PISA (les analyses des résultats de 2018 ne sont pas disponibles).

QUELQUES RÉSULTATS

En 2015 dans les trois domaines évalués (sciences, mathématiques et compréhension de l'écrit), les résultats des élèves français sont moyens (juste au-dessus de la moyenne des pays de l'OCDE). L'évolution des performances depuis les dernières éditions de 2009 et 2012 n'est pas significative : *« Le système en France est toujours dichotomique : tenu par ses bons élèves dont la proportion est stable et supérieure à la moyenne des pays OCDE, et se dégradant lentement par le bas, avec une proportion d'élèves en difficulté toujours au-dessus de cette même moyenne OCDE. »* (Rapport PISA 2015, OCDE : 2016). Ces résultats font surtout apparaître que le système éducatif français serait l'un des plus inégalitaires parmi la plupart des pays qui font partie de l'OCDE, avec l'ensemble des élèves performants et très performants (niveau 4, 5 ou 6) qui représente environ un tiers des élèves :

- la proportion d'élèves performants¹⁵ en France (plus de 21 % des élèves) est supérieure à la moyenne des pays de l'OCDE (19 %) ;
- la proportion d'élèves très performants¹⁶ en France (8%) est stable et se situe dans la moyenne des pays de l'OCDE ;
- la proportion d'élèves en difficulté¹⁷ est en très légère augmentation en 2015 (22%) et se situe, comme lors des trois dernières enquêtes PISA, au-dessus de la moyenne des pays de l'OCDE.

Concernant plus précisément les compétences évaluées, la France (en 2015) apparaît en milieu de classement pour l'évaluation des compétences à l'écrit, chez les garçons (rang 22 sur 44¹⁸), avec de meilleurs résultats observés chez les filles (rang 18 sur 44), mais bien loin derrière le Canada par exemple (rang 3 pour les garçons et rang 4 pour les filles). Les performances des élèves dans ce domaine sont néanmoins en augmentation par rapport à 2009, plaçant désormais la France au-dessus de la moyenne des pays de l'OCDE. En mathématiques, les résultats pour la France sont moins bons (rang 27 sur 44 pour les garçons et rang 21 sur 44 pour les filles). Pour chaque test, la France se situe à peine au-dessus de la moyenne des 44 pays qui ont fait l'objet d'une comparaison.

Par ailleurs, l'impact du milieu socio-économique sur les performances scolaires est présent dans tous les pays qui ont participé à l'enquête, avec le constat que les élèves les plus défavorisés réussissent moins bien à l'école que ceux qui sont les plus favorisés. En revanche, cet écart de réussite entre les élèves les plus favorisés et les moins favorisés est particulièrement marqué en France et la relation entre milieu socio-économique et performance scolaire, l'une des plus fortes : « *Les élèves issus des milieux les plus défavorisés en France (quartile inférieur de l'indice du milieu socio-économique) ont quatre fois plus de chance que les autres d'être parmi les élèves en difficulté (contre trois fois en moyenne OCDE).* » (Rapport de l'OCDE pour PISA, 2016 : 7).

Ces résultats sont d'autant plus préoccupants que l'OCDE concluait dans les dernières enquêtes que pour améliorer les performances des élèves, c'est la capacité d'un système à

¹⁵ Sont considérés comme performants, les élèves qui ont été classés dans le niveau 4 de compétence suite aux épreuves de compréhension écrite, culture scientifique et culture mathématique.

¹⁶ Sont considérés comme très performants, les élèves qui ont été classés dans le niveau 5 ou 6 de compétence suite aux épreuves scolaires.

¹⁷ Sont considérés en difficulté les élèves qui ont été classés sous le niveau 2 de compétence.

¹⁸ Dans ce classement, seuls les 44 pays ayant participé à l'enquête en 2006 et en 2015 apparaissent dans les graphiques publiés par l'OCDE.

faire progresser les élèves **en difficulté** qui permet de compenser les carences d'un système en général. Or en France, le système éducatif semble particulièrement inégalitaire.

Ainsi, l'édition 2015 de PISA a permis d'identifier quatre « défis » pour la France que nous présentons brièvement :

- « Rendre le système éducatif plus équitable, soutenir les élèves et les établissements défavorisés » en dotant notamment les établissements fréquentés par les élèves issus de milieux socio-économiques défavorisés, de ressources pédagogiques et d'infrastructures de meilleure qualité que celles dont ils disposent actuellement ;
- « Lutter contre l'échec scolaire dès le plus jeune âge » avec pour objet de s'assurer, dès la maternelle, que tous les élèves atteignent au moins un niveau minimal de compétences à la fin de l'enseignement secondaire, avec une attention particulière pour l'apprentissage de la lecture et des fondamentaux ;
- « Améliorer la qualité de l'enseignement et de la transmission du savoir dans le primaire et le secondaire, et revaloriser le métier d'enseignant » en améliorant la formation des enseignants qui seraient ainsi mieux dotés pour faire face aux difficultés des élèves ;
- « Rehausser la qualité et la valorisation des filières professionnelles au lycée » en développant davantage l'apprentissage et d'autres programmes qui valorisent l'expérience professionnelle.

2.1.2.3 LES TESTS DE LECTURE DE LA JOURNÉE DÉFENSE ET CITOYENNETÉ (JDC, 2019)

La Journée défense et citoyenneté (JDC) succède depuis 2011 à la Journée d'appel de préparation à la défense (JAPD). Il s'agit d'une journée de formation à la citoyenneté, organisée par le Ministère des armées, à laquelle tous les jeunes Français doivent obligatoirement participer entre 17 et 25 ans. La participation à cette journée se solde par un certificat de participation qui est un document indispensable pour toute inscription à une épreuve, un concours ou examen organisé par l'état français (permis de conduire, baccalauréat, etc.). Cette journée touche ainsi la quasi-totalité des jeunes Français. 5% des jeunes ne se présentent pas à cette journée et l'on peut légitimement supposer qu'ils représentent une population en difficulté. Néanmoins, ces tests permettent d'évaluer une tranche d'âge de manière plus précise (si on exclut les 5% de jeunes qui ne sont pas testés malgré le caractère obligatoire de cette journée) que dans toutes les enquêtes (y compris IVQ qui comprend les jeunes âgés de 16 ans) qui n'interrogent que des échantillons de population. Conduite tous les ans, cette enquête permet de suivre l'évolution des compétences pour cette

tranche d'âge¹⁹. En 2018, 710000 jeunes ont participé à cette journée. Traditionnellement, l'armée est en effet un observatoire social privilégié de la population masculine (du fait du service militaire) et a pendant longtemps permis de révéler l'analphabétisme et l'illettrisme des jeunes garçons (Bessin, 1999). Cette fonction de « dépistage » a souhaité être conservée lorsque le service militaire a été remplacé par la Journée d'Appel et de Préparation à la Défense (JAPD) devenue JDC en 2011. L'argument pour défendre la légitimité des tests reste celui inchangé d'une « occasion unique » ou d'une « dernière chance » de pouvoir évaluer des jeunes qui échapperont plus tard aux comptages existants dans d'autres institutions (enquêtes de l'INSEE, enquêtes réalisées à l'occasion de formations pour adultes, par exemple l'AFPA²⁰ ou dans le milieu professionnel, par exemple le FAF-TT²¹).

Nous présentons ensuite les tests de la JDC, qui touchent un public à peine plus âgé que celui de PISA, mais qui comportent des épreuves qui sont davantage scolaires (et que nous détaillons plus bas). Les résultats font en revanche apparaître un profil de jeunes considérés comme illettrés.

CONTENU DU TEST DE LECTURE ET DE COMPRÉHENSION ÉCRITE

Lors de cette journée, les jeunes appelés participent à un test conçu par le Ministère de l'Éducation nationale pour évaluer les acquis fondamentaux de la langue française. En réalité, ce sont les compétences en lecture et en compréhension écrites qui sont mobilisées avec deux épreuves de vocabulaire, deux épreuves de compréhension de texte et une épreuve de rapidité en lecture. Nous détaillons ci-dessous les épreuves :

- « automaticité de lecture » : dans cette épreuve, les participants doivent associer un mot et un pseudo-mot qui ont relation d'homophonie (ils se prononcent de la même manière). Cet exercice fait l'objet d'une mesure du temps moyen par réponse, qui apparaît comme l'indicateur de réussite. Sont ainsi évaluées ici, avec une contrainte de temps, les compétences en lecture de mots et la capacité à décoder des items qui

¹⁹ Depuis 2016, des problèmes techniques lors de la passation des tests ont empêché les jeunes de répondre à certains items ; ce qui ne permet pas de comparer les résultats avec ceux qui sont antérieurs à 2016. Les difficultés en lecture sont surestimées depuis 2016 car les réponses absentes ont été considérées comme erronées.

²⁰ L'AFPA (Association pour la Formation Professionnelle des Adultes) a mené une enquête auprès de ses stagiaires en formation, dont l'objectif était de comparer leur rapport à l'écrit avec leurs résultats en formation et leur insertion professionnelle plusieurs mois après leur sortie de formation. Cette enquête présente également le nombre de stagiaires en situation d'illettrisme.

²¹ Il s'agit d'une enquête de terrain menée en 2012 auprès de salariés affiliés au FAF-TT (entreprise de travail temporaire de métropole) dans le but d'obtenir des informations sur le niveau de l'illettrisme de ses salariés.

n'existent pas puis de les associer pour montrer leur similarité de prononciation. La DEPP indique que les performances dans ce test sont très élevées ;

- « connaissances lexicales » : les jeunes sont exposés à une liste de 20 mots et de 20 pseudo-mots. Les items apparaissent à l'écran et sont également lus à l'oral, pour que ne soit évaluée que la connaissance lexicale (et non l'habileté en lecture). Les enquêtés doivent distinguer les mots des pseudo-mots ;
- « traitement complexe » de l'information écrite sous la forme de deux tests :
 - o 1/ extraction des informations dans un programme de cinéma,
 - o 2/ réponse à 20 questions à partir d'un texte narratif court.

À partir des résultats aux différentes épreuves, huit profils de lecteurs ont été définis.

QUELQUES RÉSULTATS

Les derniers résultats publiés en Juin 2019 par la DEPP²², précisent qu'en 2019, un jeune sur 6 aurait une maîtrise fragile de la lecture et plus d'un jeune sur dix serait en difficulté en lecture (soit 11,5% des participants). Parmi eux, la moitié (5,2%) peut être considérée en situation d'illettrisme. Chez ces jeunes « illettrés », les compétences en compréhension de l'écrit seraient « sévères » voire « inexistantes » (DEPP, 2019). Ils sont surtout caractérisés (par rapport à ceux qui ont de faibles capacités de lecture (mais ne sont pas considérés comme illettrés) par un déficit important en matière de lexique et par le fait qu'ils n'ont pas « installé les mécanismes de base de traitement du langage écrit », ici le décodage de mots et de pseudo-mots, ce qui affecte la compréhension de l'écrit à laquelle ils n'accèdent presque pas. Inversement, près de 8 participants sur 10 seraient des lecteurs efficaces.

Les résultats montrent également que les jeunes en difficulté dans le domaine de la lecture et de la compréhension écrite sont davantage représentés dans certaines régions comme les DOM et pour la France métropolitaine, le Nord de la France et les régions proches de l'Île-de-France. Les filles ont montré de meilleures performances que les garçons, ce qui ressort également de l'enquête PISA en 2015.

2.1.3. LA PLACE DE L'ÉCRITURE ET DE L'ORTHOGRAPHE DANS LES ENQUÊTES

2.1.3.1 LA COMPOSANTE ÉCRITURE SOUVENT ABSENTE DES ENQUÊTES

Au terme de son tour d'horizon des enquêtes nationales et internationales sur la littéracie, Jeantheau (2015 : 191) n'a pas trouvé de tests évaluant spécifiquement l'orthographe ou

²² DEPP : Direction de l'Évaluation, de la Prospective et de la Performance.

l'écriture : « *Après une demande d'information auprès des membres du réseau ELINET (European Literacy Network) financé par la commission européenne et de ceux de l'EBSN (European Basic Skills Network), il n'a pas été possible de trouver d'exemples d'enquêtes nationales ou régionales sur la littéracie des adultes comportant une épreuve d'orthographe ou d'écriture de mots [...]* ». En effet, lorsque l'on observe le contenu des enquêtes sur la littéracie (dont les protocoles sont accessibles) depuis les années quatre-vingt-dix, on se rend compte qu'une grande majorité d'entre elles cherchent principalement à mesurer les compétences en lecture alors que dans la construction de l'écrit et dans la vie quotidienne, l'écriture tient une place tout aussi importante. Quelques enquêtes ont essayé d'évaluer l'écriture, mais les épreuves qui demandent de recourir à l'écrit ne permettent pas d'évaluer les compétences orthographiques.

Pourtant, en 1999 les conclusions d'une expertise (UNESCO, ILI à Philadelphie) concernant l'évaluation de la littéracie avaient montré la nécessité d'évaluer également la capacité à écrire et dénoncé le fait que le versant écriture est rarement investigué dans les enquêtes. En France, alors que l'écriture et l'orthographe suscitent depuis longtemps des débats passionnés (Millet, 1990 ; Lucci, Millet 1994 ; Paveau et Rosier, 2008) avec des rectifications orthographiques qui n'ont que partiellement abouti en 1990, la dernière enquête de l'INSEE (2011) ne propose un test d'orthographe que pour une partie des enquêtés ; ceux qui sont repérés en difficulté avec l'écrit dans le module d'orientation de l'enquête. À croire que la majorité des Français n'ont pas de difficulté dans ce domaine puisqu'il n'est pas nécessaire d'évaluer l'écriture (et l'orthographe) de ceux qui ne présentent pas de difficulté particulière avec l'écrit.

D'autres enquêtes proposent un test d'orthographe, c'est notamment le cas de l'enquête LEO (citée par Jeantheau, 2015). Cette enquête a été réalisée en Allemagne en 2010 auprès de 8500 personnes environ et dont l'objectif était d'évaluer le niveau de littéracie des Allemands âgés de 18 à 65 ans. Comme l'enquête IVQ (voir section 2.2), le protocole de cette enquête permet de repérer les plus bas niveaux de compétence à l'écrit pour établir un portrait précis des difficultés dans ce domaine avec des tests spécifiques. En revanche, dans les épreuves qui demandent de recourir à l'écrit, ce sont les capacités d'écriture et non l'orthographe qui sont évaluées. Les enquêtés devaient par exemple corriger une phrase fautive ; on imagine donc que les compétences évaluées devaient porter sur le lexique. Une épreuve orthographique est également proposée mais elle n'a été passée que par 650 personnes (du fait de l'utilisation de la méthode des « cahiers tournants »). Il s'agissait de reproduire sous la dictée quatre mots qui faisaient partie d'une blague d'une quarantaine de mots que l'on a fait écouter aux enquêtés au préalable. Les mots ont été comptabilisés comme corrects s'ils étaient bien

orthographiés. Mais le résultat de l'enquête ne fait cependant pas apparaître de score d'orthographe ; la réussite à l'exercice s'est additionnée à la réussite des autres tests. Comme le montre Jeantheau (2015), le nombre de mots dictés pour l'évaluation des compétences orthographiques est trop faible pour envisager une quelconque exploitation statistique. Il faudrait s'interroger aussi sur le choix des mots.

Dans une enquête réalisée au Canada en 2004-2005, l'organisme Statistique Canada a proposé un test d'orthographe consistant à écrire 15 mots dictés par un enquêteur (Jeantheau, 2015). Les enquêtés pouvaient choisir de réaliser le test en français ou en anglais. D'après les résultats, les enquêtés ont obtenu 90% de réponses correctes pour le test en anglais et 79% pour le test en français. Cette enquête nous montre que tester l'orthographe dans une enquête dans laquelle des scripteurs composent dans différentes langues est très compliqué, le niveau de difficulté orthographique étant très variable d'une langue à l'autre et difficilement comparable. Ce qui explique probablement l'absence de ce type d'épreuve dans les enquêtes internationales réalisées à grande échelle.

La carence de tests d'écriture dans les enquêtes nationales peut s'expliquer par une double difficulté liée au recueil et à l'analyse des données. En laissant une trace visible, l'écriture représente une compétence particulière, soumise au regard d'autrui et donc au jugement, ce qui, on le sait, complique le recueil de données auprès des publics les plus en difficulté (Besse et al., 2009). Se pose également la question du « qu'est-ce que bien écrire ? », notamment si l'on veut écarter la définition posée par l'institution scolaire, directement liée à la question de la maîtrise de l'orthographe. Besse et al. (2009) montrent que le critère de maîtrise de l'écriture a été défini de diverses manières mais finalement, les questions soulevées ne semblent pas avoir été résolues et sont toujours d'actualité. Enfin, dans les enquêtes internationales, se pose la question de la comparabilité des performances. Puisque chaque système d'écriture comporte des difficultés qui lui sont propres, il faudrait proposer des tests dont le degré de difficultés orthographique serait semblable.

2.1.3.2 L'ORTHOGRAPHE : UNE COMPÉTENCE COMPLEXE À ÉVALUER DANS UNE ENQUÊTE

D'après la définition de l'UNESCO en 1958, toute personne qui n'est pas capable de « lire et écrire en comprenant un exposé bref et simple de faits en rapport avec la vie quotidienne », est considérée comme analphabète, mais cette définition ne définit pas les normes d'écriture que l'on accepte. La définition de l'analphabétisme a continué à s'enrichir depuis, en intégrant et en définissant les notions de compétences de base puis en distinguant les personnes qui ont été scolarisées des autres avec le terme « illettrisme » (voir section 1.1 du chapitre 1).

La définition de l'OCDE (OCDE, 2000) : « *Aptitude à comprendre et à utiliser l'information écrite, dans la vie courante en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités* » ne permet pas non plus de définir clairement les compétences attendues :

- qu'entend-on par le fait d'utiliser de l'information écrite ?
- La production écrite et la maîtrise de l'orthographe font-elles partie de ces compétences ?

Besse et al. (2009) montrent qu'évaluer la capacité à lire et à écrire chez les adultes soulève de nombreuses questions de plus en plus complexes avec le développement d'une société de l'information où les supports écrits avec le développement du numérique sont de plus en plus variés. Cette augmentation des supports et des modalités d'expression n'ont fait qu'accroître les exigences de la société en matière d'orthographe. Cette compétence demeure alors, aujourd'hui plus que jamais, une compétence attendue et discriminante.

Il semblerait en réalité que le fait de savoir écrire impliquerait de manière implicite le fait de savoir « bien » écrire donc notamment de savoir orthographier. Mais ne serait-ce que s'il s'agit de savoir « bien orthographier », les frontières entre savoir et ne pas savoir ne sont pas non plus définies : faut-il connaître toute l'orthographe ? Existe-t-il des erreurs acceptables ou inacceptables ? Besse et al. (2009) montrent que la notion de « savoir écrire et savoir lire » semble tellement implicitement évidente que dans l'enquête Infométrie (1988), les critères de correction des tests de lecture et d'écriture n'avaient pas été définis. Ainsi, dans l'épreuve d'écriture qui consistait à écrire une phrase sous la dictée (voir section 2.1.2.1), seule la restitution orthographique a été prise en compte, ce qui revient à évaluer l'écriture selon la tradition scolaire du contrôle du rapport à la norme alors même que l'écriture dans la vie quotidienne d'un adulte ne se réduit pas à cette seule question de la maîtrise de la norme orthographique. Comme le rappellent Besse et al. (2009 : 33), « *Lire et écrire servent tout d'abord à entrer en communication avec d'autres personnes* ».

UNE DICTÉE POUR ÉVALUER LA CAPACITÉ D'ÉCRITURE

Nous avons vu que l'évaluation de l'écriture dans les enquêtes sur la littéracie est plutôt rare et lorsque cette compétence est testée, c'est au travers d'un exercice de dictée (au sens de transcrire des mots et souvent également des pseudo-mots qu'un enquêteur dicte). Il ne s'agit cependant pas de la dictée traditionnellement pratiquée dans l'évaluation scolaire, qui consiste généralement à reproduire un texte permettant de vérifier les compétences en orthographe lexicale mais aussi grammaticale, par la réalisation d'accords.

Dans le cadre d'une enquête, le choix du test d'orthographe est lié aux contraintes du recueil de données (maîtrise du temps de passation) et de l'analyse des données lorsque l'échantillon est important (standardisation des informations recueillies pour faciliter le codage). Les tests portent donc souvent sur une sélection de mots et pseudo-mots à reproduire sous la dictée, ce qui ne permet pas de tester la capacité à communiquer par écrit et limite également les phénomènes linguistiques à observer (comme la construction syntaxique ou les compétences grammaticales). Un texte libre quant à lui pose de nombreux problèmes en terme de codage et de comparabilité des réponses ; et le texte dicté ressemble aux exercices de tradition scolaire avec le risque que les publics les plus en difficulté ne se soumettent pas à l'exercice.

Pour choisir les mots à évaluer (et donc les compétences requises pour les transcrire et les orthographier), il existe différents outils pour le français (voir chapitre 3, section 1) qui permettent d'opérer une sélection : mots rencontrés fréquemment mais dont l'orthographe est irrégulière (mots fréquents et peu consistants), mots rares mais réguliers (mots peu fréquents mais consistants), etc. On pourrait ainsi identifier différents niveaux de difficultés, pour tester des niveaux de compétences.

L'épreuve de production écrite présente dans l'enquête IVQ se présente sous la forme d'une dictée de mots et de pseudo-mots que les enquêtés doivent produire de manière manuscrite. Cette modalité d'évaluation présente l'avantage de permettre aux scripteurs de focaliser leur attention sur la tâche demandée. Pour apprécier les compétences orthographiques, l'exercice de dictée est estimé comme optimal car les composantes liées à la situation de production sont restreintes. Ce type d'épreuve revêt donc une dimension orthographique mais aussi une dimension graphique avec la réalisation d'une activité motrice manuelle. Cette deuxième dimension peut représenter un frein à l'exécution de la première. On sait en effet que les enfants ou même des adultes ayant une écriture lente et peu assurée seront moins performants lors de la production écrite et notamment en orthographe car ils doivent mobiliser des ressources en attention et en mémoire de travail très importantes liées au geste graphique (Fayol et Jaffré, 2014). Pour Fayol (2013), l'écriture des lettres demande un développement complexe car il demande de mettre en œuvre trois compétences : la motricité fine des doigts et de la main, la perception et la cognition. Il met notamment en évidence la nécessité d'un long apprentissage pour maîtriser le geste graphique :

« Malgré une pratique quotidienne, les enfants de 10 ans ne sont pas encore parvenus à une automatisation de la production équivalente à celle des adultes : il s'ensuit que la mise en œuvre de l'écriture impose un coût élevé en attention et en mémoire. » (Fayol, 2013 : 48).

L'enfant passe par différentes étapes lors de l'apprentissage de l'écriture. Au début, les tracés sont nécessairement lents et irréguliers car l'enfant doit effectuer un contrôle visuel rétroactif qui vient guider ses gestes. Puis avec le temps, ses gestes s'automatisent grâce à la constitution en mémoire de ce que Fayol (2013) appelle des « *programmes visuels et moteurs* » qui permettent de gagner en vitesse et en régularité. C'est lorsque ses gestes sont automatisés et assurés qu'à l'adolescence, l'écriture peut prendre une dimension plus artistique avec la mise en place d'une écriture plus personnelle. Cependant, même avec l'expertise liée aux années de pratique, l'utilisation des signes graphiques reste couteuse pour la plupart des adultes.

2.2.L'ENQUÊTE INFORMATION ET VIE QUOTIDIENNE (INSEE, 2002, 2004, 2011)

Jusqu'en 2002, la France ne disposait pas d'une enquête nationale pour évaluer à grande échelle les compétences à l'écrit de la population. Différentes enquêtes ont permis d'approcher les compétences de base nécessaires à la communication (Infométrie, tests des JAPD) mais aucune ne permet de dresser un portrait précis du niveau des Français. Ce besoin est devenu particulièrement pressant suite aux nombreuses polémiques suscitées par l'enquête IALS et aux résultats particulièrement faibles de la France. C'est ainsi que l'INSEE a été chargée en 2002 de concevoir une enquête ménage sur les compétences en littéracie et en numératie de la population âgée de 16 à 65 ans, résidant en France métropolitaine : l'enquête « Information et Vie Quotidienne »²³ (IVQ). Son protocole conséquent (que nous détaillons ci-dessous) a été conçu pour mesurer la capacité de la population adulte à communiquer dans les domaines fondamentaux de l'écrit : savoir déchiffrer, comprendre ce qui est lu, savoir écrire, savoir compter. Après avoir testé le protocole en 2002 sur un nombre restreint de personnes, l'INSEE a étendu l'enquête en interrogeant en 2004 puis en 2011 un échantillon représentatif de la population française.

²³ Désormais IVQ.

2.2.1. LE PROTOCOLE D'IVQ : ÉVALUER TOUS LES NIVEAUX DE COMPÉTENCES

L'enquête IVQ vise des compétences de lecture, de compréhension, d'écriture, de calcul qui sont mobilisées dans des situations de la vie courante et liées à l'usage de l'écrit, dans sa dimension fonctionnelle. Elles relèvent donc des concepts de « littéracie » et de « numératie ». Cette enquête a la spécificité de prendre en compte tous les niveaux de compétences, y compris les faibles niveaux (parfois écartés des enquêtes internationales) pour rendre compte du nombre de personnes en situation de grande difficulté face à l'écrit dans leur vie quotidienne. Après avoir déterminé le niveau des Français dans la maîtrise des compétences de base et isolé la proportion de personnes qui se trouvent en difficulté avec l'utilisation de l'écrit dans leur vie quotidienne (et défini le taux d'illettrisme en France), l'enquête IVQ a pour objectif de définir la nature de leurs difficultés en mettant en relation leurs résultats aux tests avec les données recueillies dans le module biographique (Besse et al., 2009).

Le cadre de recueil de données a été conçu pour motiver les personnes interrogées à s'impliquer dans la réalisation des tâches proposées, les mettre en confiance et ainsi limiter le taux d'erreurs lié au contexte d'entretien, avec le souci de se démarquer des références scolaires. L'enquête commence ainsi par un module d'orientation. Depuis IALS, la procédure d'orientation dans les enquêtes est présentée comme un moyen à la fois de limiter les biais liés à la motivation des enquêtés mais aussi de prendre en compte tous les niveaux de compétence. Pour Degorre et Murat (2009), l'engagement de l'enquêté pour répondre aux questions posées constitue un point central pour le recueil de données fiables. L'attention et la concentration sont en effet inhabituellement sollicitées dans les enquêtes réalisées à domicile, car les personnes interrogées ne sont pas rodées à ce type d'exercice. Pour préserver l'intérêt et la motivation de l'enquêté, les tests proposés doivent être adaptés aux personnes interrogées.

Pour rompre avec une tradition d'évaluation dans le cadre scolaire (que chaque enquêté a probablement connu) et pour aborder les compétences d'un point de vue fonctionnel, les tests sont issus de situations concrètes du quotidien comme la lecture d'un programme TV, l'écriture d'une liste de courses. Ils se réfèrent ainsi à des « *conduites de communication relationnelle et sociale de la vie courante* » (Besse et al., 2009). Les supports sont également issus de la vie quotidienne (page de programme de télévision, pochette d'album CD, bulletins d'information, etc.) et présentent ainsi l'avantage d'être des éléments familiers ou au moins connus, qui doivent favoriser l'acceptation des exercices par les enquêtés. Ces choix montrent une volonté de proposer des exercices qui s'articulent selon une vision fonctionnelle de l'écrit : ce n'est pas un examen scolaire que l'on fait passer à ces publics.

L'enquête IVQ est longue : sa durée moyenne, si tous les modules sont passés, est de 1 heure 10, dont environ 50 minutes d'exercice, ce qui sollicite fortement l'attention et la concentration des enquêtés. L'architecture de l'enquête prévoit que les tests correspondent au niveau de l'enquêté et l'orientation ou la réorientation des enquêtés au fur et à mesure du questionnaire a été rendue possible par le mode de collecte assisté par ordinateur.

Une première enquête à visée méthodologique a été réalisée sur une centaine de personnes en 2000 puis en 2002, dans le but de mettre au point le protocole définitif (Besse et Guérin-Pace, 2002). La première enquête (auprès d'un échantillon représentatif de la population française) a ainsi été réalisée en 2004 auprès de plus de 10300 personnes (Degorre et Murat, 2009). Les premiers résultats ont été présentés par l'INSEE à partir de 2003 (Murat, 2004, 2004, 2005). La deuxième édition s'est déroulée en 2011 auprès de 13750 personnes et les résultats diffusés par l'INSEE et l'ANLCI à partir de 2012 (Jonas, 2012 ; Brendler, 2013 ; ANLCI, 2013 ; etc.). Le protocole de 2011 a également été repris (avec quelques adaptations qui ne portent pas sur la forme des épreuves mais plutôt sur des questions de sémantique, par exemple le nom des villes dans les textes) en Martinique (2006), à la Réunion (2007) et en Guadeloupe (2008).

2.2.1.1 PRÉSENTATION DE L'ARCHITECTURE DE L'ENQUÊTE

Le protocole de l'enquête a été construit pour permettre d'orienter les enquêtés vers un panel d'exercices qui correspondent à leur niveau de compétences : ceux qui n'ont pas de difficulté particulière dans le domaine de l'écrit et ceux qui sont en difficulté ne passent pas les mêmes exercices.

Un premier **module « d'orientation »** est destiné à diriger les enquêtés vers un module « bas » (composé d'exercices dits simples) ou « haut » (composé d'exercices dits complexes). Un **module « intermédiaire »** permet, à la suite du test d'orientation, de diriger les enquêtés qui se situent dans un seuil de compétences moyen (et ne permettant pas de déterminer de façon nette si ces derniers relèvent du module « bas » ou du module « haut »). Cette organisation permet de proposer des exercices dont le niveau serait adapté à celui de l'enquêté, évitant ainsi une organisation « linéaire » des compétences qui iraient de la plus faible à la plus forte.

Les tests ont été conçus par différentes équipes de chercheurs. Le module « haut » a été conçu d'après les travaux en psychologie de Tardieu et son équipe (Université Paris 5). Le module « bas », le module « d'orientation » et les exercices de compréhension orale ont été élaborés par l'équipe du PsyEF (Université Lyon 2), dirigée par Besse. L'ANLCI a également participé à la conception du module « bas ». Enfin l'équipe EVA (Université de Rennes et Hôpital Sainte

Anne) dirigée par Charron et Meljac, a constitué le module « numératie » et les questions de calcul du module « d'orientation ».

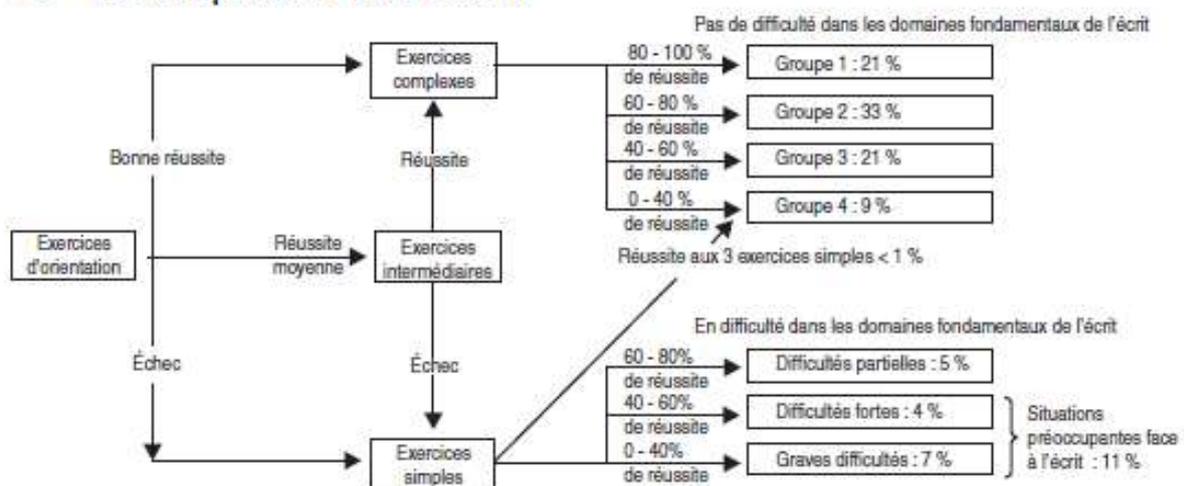
À l'issue de l'enquête, les résultats obtenus déterminent le groupe dans lequel les enquêtés sont classés, au sein de deux catégories :

- **pas de difficulté** dans les domaines fondamentaux de l'écrit : groupes 1 à 4 (1 étant le groupe qui a obtenu les meilleurs scores) ;
- **en difficulté** dans les domaines fondamentaux de l'écrit : cette catégorie se décline en « types de difficultés » (difficultés partielles, difficultés fortes, graves difficultés).

Par ailleurs, les enquêtés orientés vers le module « bas » qui ont obtenu plus de 80% de réussite dans chacun des exercices sont classés dans la catégorie des enquêtés qui n'ont « pas de difficulté dans les domaines fondamentaux de l'écrit » (dans le groupe 4). L'INSEE considère ici qu'il s'agit ici d'une « erreur » d'orientation ; moins de 1% des enquêtés sont concernés.

Voici l'architecture de l'enquête que nous reprenons de la publication de l'INSEE suite à l'enquête de 2011 :

③ Les compétences face à l'écrit



Champ : personnes de 18 à 65 ans en France métropolitaine.

Source : Insee, enquête Information et vie quotidienne, 2011.

Figure 1. Architecture de l'enquête IVQ 2011 (INSEE, 2011)

Dans chaque module « bas », « haut », les enquêtés ont passé différents tests qui correspondent aux compétences de base nécessaires pour communiquer à l'écrit : lecture de mots et de pseudo-mots ; compréhension écrite d'un texte simple ; compréhension orale ; numératie. Dans le module « bas », les enquêtés ont également passé un test d'écriture de mots et de pseudo-mots. Enfin, tous les enquêtés ont passé un **module « biographique »** permettant de recueillir des données à visée sociologique (conditions matérielles de vie,

enfance, vie professionnelle, etc.). Les enquêteurs (et non les enquêtés) ont enfin complété un **module « qualité »** composé de 19 questions destinées à définir les conditions de passation des tests et l'attitude de l'enquêté.

Nous détaillons dans la partie suivante, le contenu de chaque test.

2.2.1.2 LE MODULE D'ORIENTATION ET LE MODULE INTERMÉDIAIRE : ADAPTER LES DIFFICULTÉS DU TEST AUX COMPÉTENCES

L'enquête IVQ avait pour objectif, dès sa conception, de prendre en compte tous les niveaux de compétences dans son protocole et notamment les plus faibles niveaux dont on sait qu'ils sont les plus difficiles à atteindre. Dans les enquêtes, les personnes qui ont de faibles compétences à l'écrit n'ont de module dédié et dans certaines - comme IALS, ils sont parfois carrément écartés du protocole. Si cette pratique peut permettre d'indiquer une proportion de personnes en difficulté, elle ne permet pas de savoir ce que les personnes qui ont montré des faiblesses savent faire. Aussi, dans IVQ, deux types d'exercices sont proposés et pour orienter les enquêtés, deux procédures sont mises en place : un premier module « d'orientation » et un module « intermédiaire » destiné à affiner l'orientation.

LE MODULE « D'ORIENTATION »

L'enquête débute ainsi par un module « d'orientation » qui permet de déterminer le niveau de l'enquêté. Les questions ont été conçues pour être adaptées au plus grand nombre, pour ne pas mettre en difficulté les enquêtés ayant un faible niveau de compétences et ne pas démotiver les autres avec des exercices trop simples. Ce test n'a pas été présenté comme tel (c'est-à-dire dont la réussite conditionne la suite du test avec une orientation vers des exercices simples ou complexes), pour ne pas stresser ni provoquer de sentiment de dévalorisation de la part des enquêtés. Pour garantir le bon déroulement de l'enquête et ne pas entrer en conflit en cas de grande difficulté ou de refus des épreuves, l'enquêteur pouvait interrompre un module pour passer directement au suivant.

Le module « d'orientation » est composé d'exercices d'identification de mots et de pseudo-mots, de compréhension écrite et de calcul qui permettent d'orienter l'enquêté vers un parcours différencié.

Voici quelques données qui permettent de rendre compte du contenu de ce module :

- Épreuve de lecture :
 - mots à lire : FAMILLE – JARDIN – BÊTE – CHAT – MICHEL ;
 - pseudo-mots à lire : BODAR – MOUDAN – VANI – NAGOPIAN – RACITAME.

- Épreuve de compréhension écrite : cette épreuve réutilise un support déjà testé par Besse et al. (DMA²⁴). Les enquêtés devaient répondre oralement à des questions portant sur un texte issu d'un programme TV. Par exemple : « Dans quelle ville se déroule la fête ? », « Quel est l'âge de la personne qui fête son anniversaire ? ».

Ce sont les scores obtenus sur l'identification de mots et pseudo-mots ainsi que sur la compréhension qui conditionnent la suite de l'enquête :

- les enquêtés ayant obtenu moins de 11 points sur 15 à l'exercice d'identification de mots ou moins de 11 points sur 19 à l'exercice de compréhension, sont dirigés vers le module « bas » ;
- les autres sont orientés vers le module « haut ».

Le module d'orientation teste donc les enquêtés sur leurs compétences en lecture : décodage et compréhension. Le protocole d'IVQ prédit donc de possibles difficultés en production écrite pour les personnes qui n'ont pas réussi les tests de lecture et/ou de compréhension (avec des scores inférieurs à 40% de réussite) dans la mesure où les autres enquêtés n'ont pas passé de test d'écriture. Mais avoir des faiblesses en lecture et/ou en compréhension signifie-t-il pour autant ne pas maîtriser l'orthographe ? Et passer avec succès les deux premières épreuves est-il une garantie de maîtrise de l'orthographe ?

LE MODULE « INTERMÉDIAIRE »

Le module « intermédiaire » a été construit pour l'enquête de 2004 pour les personnes qui se situent aux limites du seuil de compétences suite à la première série d'exercices du module « d'orientation ». Il permet de limiter les erreurs d'orientation : par exemple orienter vers des exercices simples une personne qui aurait mal réussi les premières questions de l'enquête (difficulté à entrer dans l'évaluation en général, erreurs liées à de l'étourderie ou à un manque de motivation, etc.) alors qu'elle n'a pas de difficulté particulière à l'écrit, « *le diagnostic étant désormais plus fiable pour les personnes situées autour du seuil de compétence minimal retenu* » (Murat, 2005 : 4). EN 2002, 12% des enquêtés ont été dirigés vers le module « bas » suite au module « d'orientation ». En 2004, ce sont 20% des enquêtés. Cette différence tient selon Murat (2009) à la présence en 2004 du module « intermédiaire ». Ainsi, en 2002, des personnes en difficulté auraient été dirigées à tort vers les exercices complexes du module « haut ».

²⁴ DMA : Diagnostic des modes d'appropriation de l'écrit. Pour plus d'informations sur ce dispositif, se référer à Besse et al. (2004).

Cependant, Degorre et Murat (2009) constatent que le module « intermédiaire » (tel qu'il a été conçu pour IVQ 2004) pourrait être amélioré pour qu'il permette d'affiner réellement les compétences des enquêtés au vue de leur orientation. Dans ce module, l'épreuve de compréhension écrite du module « bas » a été reprise, ce qui interfère avec l'ordre des modules passés ainsi qu'avec la durée totale de l'enquête :

« Les personnes orientées vers le module intermédiaire puis vers le module Haut ont, au final, répondu à un exercice supplémentaire, qui a pu affecter leur concentration et leur performance dans la suite de l'enquête. Les personnes orientées vers le module intermédiaire puis vers le module ANLCl passent des items dans un ordre différent de celles directement orientées vers le module ANLCl, ce qui pourrait altérer la difficulté de l'épreuve, dans la mesure où les questions d'identification de mots, lorsqu'elles sont passées en premier, facilitent ensuite les questions de compréhension de texte. » (Degorre et Murat, 2009 : 26).

Degorre et Murat (2009) préconisent ainsi une refonte de ce module, avec des épreuves pour lesquelles les processus cognitifs mobilisés seraient proches de ceux du module « haut » mais moins difficiles pour que les exercices restent accessibles aux personnes en difficulté avec l'écrit. Ces éléments n'ont pas été pris en considération dans l'enquête de 2011, qui a repris le protocole à l'identique de celui de 2004.

2.2.1.3 LE MODULE BIOGRAPHIQUE ET LE MODULE QUALITÉ

Nous présentons à présent le module « biographique » et le module « qualité » dans deux sous-parties.

LE MODULE « BIOGRAPHIQUE »

L'enquête IVQ a également pour objectif de permettre une meilleure connaissance des publics qui ont des difficultés avec l'écrit (dont ceux qui sont illettrés) et de définir les variables qui sont liées aux compétences en littéracie et en numératie à l'âge adulte. Tous les enquêtés ont donc passé un module « biographique » qui permet, en plus des caractéristiques socio-économiques de la personne interrogée, de collecter des informations sur sa pratique quotidienne de l'écrit.

Les concepteurs du module « biographique » ont considéré que le parcours scolaire n'était pas une variable suffisante pour expliquer les compétences acquises dans les domaines fondamentaux de l'écrit et le protocole très complet prévoit donc d'investiguer d'autres facteurs comme le profil familial, les activités culturelles et sociales, les conditions de vie matérielles ainsi que l'incidence du parcours professionnel. La période de l'enfance, dont on connaît l'importance dans l'acquisition des compétences de base, est finement investiguée

(tant le parcours scolaire que le contexte familial). Ces variables doivent permettre d'expliquer les différences de compétences entre les enquêtés et comprendre ce qui peut influencer sur les difficultés d'apprentissage à l'école.

Le module « biographique » comprend également une série de questions destinées à rendre compte des éventuelles stratégies de contournement utilisées par les personnes qui n'ont pas acquis les compétences de base pour être autonomes dans leur vie quotidienne ou dans leur vie professionnelle. Les résultats montrent en effet que même si la faible maîtrise des compétences de base est corrélée à un taux d'inactivité plus important, pour autant, cela n'interdit pas l'accès à l'emploi, y compris à des emplois qui nécessitent de maîtriser la lecture ou l'écriture (Jonas, 2013). Guérin-Pace (2009) a analysé le versant sociologique de l'enquête IVQ 2004. Elle montre que si certaines variables semblent entraîner davantage de difficultés avec l'écrit à l'âge adulte (difficultés scolaires précoces, pratiques de lecture peu présentes, contexte familial difficile), ce public est également marqué par une diversité de profils en terme d'insertion sociale et professionnelle en lien avec un rapport à l'écrit très différent. Elle souligne notamment que les difficultés avec l'écrit n'ont pas les mêmes répercussions selon les individus.

LE MODULE « QUALITÉ »

Par ailleurs, les enquêteurs ont complété le module « qualité » composé de 19 questions destinées à définir les conditions de passation des tests et l'attitude de l'enquêté. Ce module est destiné à recréer l'atmosphère qui régnait lors de l'enquête et qui peuvent aider à un traitement qualitatif des données. Il doit notamment permettre d'estimer la qualité du recueil des données, en fonction de l'atmosphère de l'entretien qui est décrit. Ces informations doivent également permettre d'améliorer le protocole, car on sait que préserver l'intérêt et la motivation des enquêtés durant l'intégralité du test constitue une variable forte de fiabilité des données.

L'enquêteur a d'abord renseigné une partie concernant le déroulement de l'enquête : le nombre de visites avant de pouvoir interroger l'enquêté, le refus éventuel de répondre aux questions (immédiat ou au cours du test) avec les raisons de ce refus. Les questions suivantes portaient sur :

- l'atmosphère de l'entretien (type de réponse possible : plutôt tendue, froide mais courtoise, agréable, autre) et l'évolution de l'atmosphère (l'enquêté a progressivement montré de l'ennui, de l'agacement, s'est détendu, s'est intéressé, est resté à peu près stable) ;

- la posture de l'enquêté : fatigue, concentration, installation (sur une table, debout, sur un meuble, etc.) ;
- les conditions de passation pouvant interférer sur les réponses : nombre et type de personnes présentes, bruit, intervention pour aider l'enquêté à répondre, etc.

Enfin, l'enquêteur indique si l'enquêté a selon lui joué le jeu (oui vs non).

Nous détaillons à présent le contenu de chaque module.

2.2.2. FOCUS SUR LES COMPÉTENCES TESTÉES DANS LE MODULE « BAS »

Le module « bas » est une originalité de cette l'enquête IVQ puisqu'il teste spécifiquement les compétences d'un public repéré pour ses faibles compétences dans le domaine de l'écrit. Il consiste donc en l'évaluation des personnes qui présentent des difficultés dans les domaines fondamentaux de l'écrit. Rappelons que l'un des objectifs de cette enquête d'utilité publique était d'obtenir une mesure du taux d'illettrisme en France.

Les enquêtés ont ainsi été interrogés sur différentes composantes de l'écrit : lecture, écriture et compréhension écrite. Leur niveau de connaissances dans chaque compétence est évalué par un score de réussite. Le score de réussite s'inscrit alors dans une fourchette de réussite calculée en pourcentages, qui classe les enquêtés en 4 catégories :

- en graves difficultés : moins de 40 % de réussite ;
- en fortes difficultés : entre 40 % à 60 % de réussite ;
- en difficultés partielles : entre 60 % à 80 % de réussite ;
- sans difficulté : plus de 80% de réussite.

Nous détaillons ci-dessous les trois composantes évaluées.

2.2.2.1 LA COMPRÉHENSION ORALE

Pour Besse et al. (2009), la compétence en compréhension orale est déterminante pour évaluer le niveau d'illettrisme afin de ne pas imputer à la lecture ou à l'écriture des difficultés qui relèvent d'autres facteurs :

« Une personne ne maîtrisant pas le français oral présentera de facto des problèmes de compréhension et de production pour la lecture et l'écriture du français et cela qu'elle soit analphabète ou lettrée dans sa langue. Mesurer l'illettrisme c'est donc n'évaluer que la lecture et l'écriture. Ces compétences pourront être établies lorsqu'un niveau minimal de compétence en maîtrise du langage oral (ici, le français : langue dans laquelle s'exprime l'enquêteur Insee), en particulier pour ce qui concerne la compréhension à l'oral, est vérifié. » (Besse et al., 2009 : 35).

À ce sujet, le moyen de recueil des réponses dans les épreuves de compréhension a été revu par rapport à celui de l'enquête IALS. Dans IALS, les enquêtés devaient rédiger leurs réponses alors que le protocole d'IVQ a réduit le recours à l'écrit en recueillant les réponses à l'oral (sauf pour l'épreuve de dictée). L'expertise mise en place suite à la participation de la France dans cette enquête avait notamment montré que l'utilisation de l'écriture pour évaluer d'autres composantes de l'écrit (comme la compréhension) peut biaiser les résultats chez les enquêtés qui ont des difficultés d'expression et de rédaction.

Pour construire les tests, l'équipe de Besse a utilisé un outil mis au point dans le cadre d'entretiens individuels auprès d'adultes en situation d'illettrisme : le Diagnostic des modes d'appropriation de l'écrit (DMA) (Besse et al., 2004). Cette démarche d'évaluation se veut davantage qualitative : il s'agit d'étudier la manière de lire ou d'écrire des personnes interrogées, appelées aussi procédures de lecture ou d'écriture. Pour Besse et ses collaborateurs, *« Le postulat initial est qu'il s'agit plutôt de permettre que se manifestent des compétences chez la personne évaluée (même si ces compétences ne sont pas très nombreuses et diversifiées) plutôt que de sanctionner des insuffisances et des manques (les examens scolaires sont déjà passés par là). »* (2009 : 36). Pour l'enquête IVQ, cette démarche qualitative a été adaptée aux exigences propres aux enquêtes quantitatives.

2.2.2.2 LES COMPÉTENCES EN LECTURE ET EN COMPRÉHENSION ÉCRITE

Le test de lecture porte sur l'identification de mots et de pseudo-mots (sous la forme de mots isolés ou de syntagmes) et sur le signalement de mots (le signalement est la réponse à la question « où est marqué tel mot ? »). Les mots ont été choisis à partir de l'échelle Dubois-Buyse (Ters et al., 1995), selon différents critères (Besse et al., 2009) :

- leur fréquence : seuls des mots fréquents ont été choisis ;
- leur complexité phonographique : des mots réguliers et irréguliers ;
- leur longueur : des mots de trois, cinq et sept phonèmes ;
- des graphèmes fréquents : les mots qui ont été choisis ne comportent que les 45 graphèmes de base décrits par Catach (1980).

Nous présentons ensuite la liste des mots et syntagmes devant être lus ou repérés :

- liste des mots et syntagmes devant être lus :
 - NATURE, DOMICILE, MER SUCCULENTE, PAILLE, SOUPCONNER SES AMIS, LE SOUPIR DE SON FILS, AIGUILLE DU NORD, ADIEU MON CAMARADE, EXAUCE MES DÉSIRES, TIPAK, MON AMI ;
- liste des mots devant être repérés :
 - EXAUCE, AIGUILLE, SOUPÇONNER.

L'enquêté est ensuite amené à lire une jaquette de CD et un texte critique de cet album puis il répond à des questions qui demandent une recherche d'informations.

Suite au test de lecture, les enquêtés obtiennent un score de lecture compris entre 0 et 22, et un score en compréhension écrite compris entre 0 et 23. Ces scores seront pris en considération dans le calcul final (avec la production écrite) pour déterminer si l'enquêté est vraiment en difficulté (si l'enquêté obtient plus de 80% de réussite dans tous les exercices du module « bas », il est considéré comme n'ayant pas de difficulté avec l'écrit). Nous rappelons que pour ne pas être considéré en difficulté avec l'écrit, les enquêtés doivent obtenir plus de 80% de réussite dans **tous** les domaines évalués : un seul score inférieur à 80% et l'enquêté est considéré en difficulté « dans au moins un des domaines fondamentaux de l'écrit » (il n'existe pas à ce niveau de distinction entre ceux qui auraient par exemple de grosses difficultés en production écrite mais aucune difficulté en compréhension et en lecture).

2.2.2.3 LA CAPACITÉ D'ÉCRITURE

Dans les enquêtes à très grande échelle, l'évaluation pratiquée est effectuée à des fins statistiques où la comparabilité des réponses est fondamentale, tout comme l'automatisation de la correction. Toute possibilité d'interprétation dans les corrections peut constituer un biais (nous y reviendrons dans la partie consacrée à la saisie de la dictée dans IVQ). Lorsque l'on cherche à évaluer les compétences en production écrite, ces deux contraintes limitent profondément le type de compétences que l'on souhaite juger (est-ce la capacité à argumenter, à résumer, à développer, à orthographier ?) et donc le type d'épreuve (quels exercices proposer pour recueillir des données comparables et significatives ?). Mais aussi, comme nous l'avons déjà évoqué plus haut, qu'est-ce que savoir écrire ?

Jeantheau (2015) rappelle que dans l'enquête IVQ, la présence de la dictée provient de la volonté de déterminer le taux d'illettrisme d'après la définition nationale de l'illettrisme. Cette définition comprend en effet la capacité à produire des messages écrits. Le module bas propose d'évaluer les capacités d'écriture par l'intermédiaire d'une épreuve de dictée.

L'exercice est présenté comme une liste de courses (composée de mots et de pseudo-mots) dictée par l'enquêteur et que les enquêtés doivent transcrire à l'attention d'un livreur. Le fait que la liste soit dictée, nous le verrons dans la partie consacrée à l'analyse des dictées, pose problème notamment pour les pseudo-mots (certaines propositions peuvent laisser penser que l'item n'a pas été dictée conformément à sa forme phonographique par l'enquêteur). L'enquêté doit écrire sur une feuille de papier libre, ce qui, nous le verrons dans la partie consacrée à l'informatisation du corpus, a posé des difficultés de déchiffrage.

Cette épreuve est inédite : comme on l'a déjà mentionné, à de rares exceptions, seules les compétences en compréhension écrite et en lecture sont habituellement mesurées pour évaluer les compétences en littéracie des adultes. Elle est en tout cas la première à proposer de mesurer la capacité des adultes à produire des mots écrits dans un contexte de communication (transmettre une liste à un livreur) et elle ne comporte pas uniquement des mots sans lien entre eux, comme cela est habituellement le cas. Cette dictée comporte en effet deux syntagmes qui permettent la réalisation de deux accords en nombre. C'est donc principalement la capacité à orthographier qui est ici testée, la capacité à produire de l'écrit dans un but de communication ne pouvant pas être, avec cet exercice, réellement jugée. Seuls les pseudo-mots ont pour objectif de cerner les compétences en phonographie, la graphie de ces items n'ayant comme contrainte que de restituer la forme orale de ce qui est dicté.

Si les élèves sont habitués au fait de passer des exercices décontextualisés dont l'objectif est de mobiliser leurs connaissances, les adultes en revanche, sortis pour la plupart depuis longtemps du système éducatif (avec parfois de mauvais souvenirs de l'évaluation telle qu'elle est pratiquée classiquement) ne sont plus habitués à être évalués. L'évaluation de l'orthographe semble donc relever d'un défi particulièrement important, tant les souvenirs des dictées semblent présents dans l'esprit des adultes, pourtant bien des années après les avoir pratiquées. En effet, selon Degorre et Murat (2009 : 4), « *la dictée par exemple est entourée d'une symbolique où se superposent la crainte de l'erreur et le rapport à l'autorité. Proposer une telle épreuve avec les pratiques scolaires usuelles conduirait à ce que l'enquêteur revête l'image du 'maître', comme figure du juge* ». Le niveau d'acceptation de ce type d'exercice risque ainsi d'être très faible dans le cadre d'une enquête à domicile.

En résumé, le choix de cet exercice peut être discutable car la dictée présente une connotation scolaire et sociale forte (notamment auprès des publics en difficulté avec l'écrit), malgré les précautions prises par l'INSEE (l'exercice n'a pas été présenté comme une dictée mais comme une liste de courses à établir à l'attention d'un livreur). Par ailleurs, les phénomènes linguistiques à observer dans une liste sont limités (par rapport à un texte). Néanmoins, la

dictée présente l'avantage de pouvoir rendre comparables les productions (avec une relative facilité de codage des réponses) et d'approcher la compétence orthographique d'adultes.

À l'issue des différents tests du module « bas », les enquêtés sont répartis en deux catégories :

- pas de difficulté dans les domaines fondamentaux de l'écrit ;
- en difficulté dans les domaines fondamentaux de l'écrit : parce qu'ils ont obtenu, au sein du module « bas », moins de 80% de réussite dans l'une des trois compétences testée. Cette catégorie des « en difficulté » est ensuite déclinée en trois type de difficultés (partielles, fortes, graves). L'INSEE considère alors que les personnes qui ont montré des difficultés fortes ou graves à l'issue du test (c'est-à-dire qui ont obtenu moins de 60% dans l'une des épreuves), sont dans une situation « préoccupante » à l'écrit, et relèvent d'une situation d'illettrisme (si les critères de l'ANLCI comme le fait d'avoir été scolarisé en France sont également présents).

C'est ainsi qu'au terme d'IVQ 2011, l'INSEE a considéré que 7% des français scolarisés en France, quelle que soit leur nationalité, se trouvaient en situation d'illettrisme.

2.2.3. CONCLUSIONS DE L'ENQUÊTE ET PRINCIPAUX RÉSULTATS

L'évaluation des personnes faiblement littéraciées est une particularité de l'enquête IVQ. Pour estimer la fiabilité des données recueillies (et attester la validité du protocole), Besse et al. (2009) ont mesuré l'implication des enquêtés les plus en difficulté dans l'enquête. Pour cela, ils ont analysé le taux de réponses des enquêtés dirigés vers le module « bas » dans l'enquête de 2004. Ils récapitulent, dans le tableau 1 reproduit ci-dessous, le taux de participation des enquêtés avec une distinction faite entre les enquêtés qui ont été orientés vers le module « bas » directement après le module « d'orientation » (« orientation directe ») et ceux qui, situés à la limite du seuil minimal de compétences, ont passé le module « intermédiaire » avant d'être finalement dirigés vers le module « bas » (« orientation indirecte »). Besse et al. (2009) considèrent que les enquêtés dirigés directement vers le module « bas » suite au module « d'orientation » représentent ceux qui sont le plus en difficulté dans l'enquête.

	production écrite	identification de mots et de PM	compréhension de l'écrit	ensemble du module « bas »
orientation directe	77,3%	89,5%	71,5%	81,2%
orientation indirecte	86,9%	98,4%	81,9%	90,3%

Tableau 1. Taux de participation des enquêtés dans les épreuves du module « bas ».

Dans IVQ 2004, 832 personnes ont été directement dirigées vers le module « bas » suite au module « d'orientation » et 1031 personnes l'ont été de manière indirecte. Les chiffres

montrent que les enquêtés « bas » ont répondu à 81,2% des questions quand ils ont été orientés directement et à 90,3% quand ils ont été orientés indirectement. Pour Besse (2009), ces chiffres témoignent d'une implication très importante. Dans chaque épreuve, l'écart d'implication est plutôt constant. Il s'élève à environ 10 points, en faveur des enquêtés qui n'ont pas été dirigés directement et qui s'impliquent davantage (parce qu'ils sont probablement moins en difficulté que les autres). C'est l'épreuve de compréhension de l'écrit qui montre les taux d'implication les moins élevés pour les deux groupes, même s'il est néanmoins supérieur à 70 % pour ceux qui ont été dirigés directement.

Ces chiffres laissent penser que les choix méthodologiques sont convenables puisqu'ils ont permis d'assurer une implication des personnes en difficulté avec l'écrit sur les tâches proposées sur un temps qui soit suffisamment long pour que l'on puisse cerner leurs compétences. Ce degré d'implication montre aussi un niveau suffisant de motivation qui permet d'attester la pertinence des résultats de cette enquête. L'importance du taux de non réponse dans IALS avait en effet été reprochée et imputée à une faiblesse du protocole qui mettait en doute la pertinence des résultats.

Nous présentons à présent brièvement les principaux résultats de l'enquête de 2011 qui ont été publiés par l'INSEE. Ces résultats concernent principalement les variables sociologiques en lien avec les compétences en lecture, en compréhension et en calcul. Nous ne nous étendons pas sur ces publications car celles-ci ne concernent pas le cœur de notre étude qui est axé sur les compétences à l'écrit des enquêtés.

2.2.3.1 PRÉSENTATION DES PRINCIPAUX RÉSULTATS

Nous présentons les résultats publiés suite à l'enquête de 2011 par l'INSEE (INSEE première, Jonas, 2012 ; Murat et Rocher, 2016).

À l'issue de l'enquête, 16% des enquêtés ont montré des difficultés dans les domaines fondamentaux de l'écrit (lecture, compréhension, écriture). Parmi eux, 11% ont des difficultés graves ou fortes dans au moins un des trois domaines évalués : 7% ont été scolarisés en France et peuvent être considérés comme illettrés (selon la définition proposée par l'ANLCI), ce qui représente environ 2,5 millions de personnes âgées de 16 à 65 ans.

Le classement des enquêtés en difficulté proposé par l'INSEE montre la répartition suivante :

- 1% n'a pas pu passer les tests à cause de difficultés trop importantes en français ou d'un niveau de lecture trop faible ;
- 6% auraient une « communication très difficile » (moins de 40% de bonnes réponses dans au moins un des domaines fondamentaux de l'écrit) ;

- 4% auraient une « communication faiblement efficace » (entre 40 et 60% de bonnes réponses dans au moins un des domaines fondamentaux de l'écrit) ;
- 5% auraient une « communication possible » (entre 60 et 80% de réussite dans au moins un des domaines fondamentaux de l'écrit). Pour ces enquêtés, les difficultés se concentrent souvent dans un seul des trois domaines évalués.

Ce serait ainsi une personne sur six, âgée de 16 à 65 ans et résidant en France métropolitaine, qui éprouverait des difficultés (au moins partielles) en lecture, en compréhension et/ou en production écrite. Jonas (2012) souligne également que les difficultés ont tendance à se cumuler même s'il ne s'agit pas d'une règle absolue. En effet, près d'une personne sur deux qui se trouve en difficulté en lecture ou en production écrite, l'est également en calcul ou en compréhension. Par ailleurs, 21% des enquêtés en difficulté en lecture ou en production écrite obtiennent de très bons résultats en compréhension orale.

Chez les enquêtés en difficulté (ceux qui ont obtenu moins de 80% de bonnes réponses dans un des trois domaines évalués), les difficultés sont davantage présentes en compréhension écrite qu'en production écrite et enfin en lecture qui représente le domaine le mieux réussi :

- 15% obtiennent moins de 80% de réussite en compréhension écrite ;
- 11% obtiennent moins de 80% de réussite en production écrite ;
- 5% obtiennent moins de 80% de réussite en lecture.

Des différences de réussite selon le sexe ont également été observées : Jonas (2012) relève que les inégalités entre les hommes et les femmes sont peu marquées en compréhension orale mais qu'elles le sont en calcul, avec de meilleures performances observées chez les hommes dans ce domaine. En revanche, cette tendance s'inverse dans le domaine de l'écrit (lecture, compréhension écrite et production écrite). À l'écrit, les hommes éprouvent davantage de difficultés que les femmes : 34% des femmes ont obtenu plus de 80% de réussite à la dictée contre 24% chez les hommes. En moyenne, 51% des hommes ont éprouvé des difficultés graves ou fortes en production écrite, contre 41% chez les femmes.

Le pays d'origine et de scolarisation des adultes représente également un facteur important de réussite, notamment à l'écrit. Jonas (2012) relève que 61% des personnes qui n'ont pas été scolarisées en France et qui n'ont pas non plus été scolarisées en français éprouvent des difficultés à l'écrit contre 31% des personnes qui n'ont pas été scolarisées en France mais dont la langue de scolarisation (ou la langue maternelle) est le français.

Enfin, les compétences à l'écrit ont tendance à diminuer avec l'âge. Deux facteurs permettent de l'expliquer : le « facteur générationnel » décrit par Jonas (2012), Murat et Rocher (2016) et

le facteur « cycle de vie » décrit par Murat et Rocher (2016). Les résultats d'IVQ (en 2004 comme en 2011) montrent en effet que les plus jeunes ont moins de difficultés dans les domaines fondamentaux de l'écrit et en calcul : selon Murat et Rocher (2016), en 2011, 10% des 18-29 ans ont des difficultés à l'écrit contre 24% des 60-65 ans. Jonas (2012) souligne également que cet effet de génération est visible entre les enquêtes de 2004 et de 2011 (espacées de 7 ans), avec une amélioration moyenne des 18-65 ans. En 2004, la génération née avant 1946 (exclue de l'enquête de 2011) avait montré qu'un tiers était en difficulté avec l'écrit et la jeune génération (ceux nés après 1986), qui était hors du champ de l'enquête de 2004, vient encore renforcer ce fossé des générations en 2011 (qui explique également le recul de 2 points du taux d'illettrisme entre 2004 et 2011).

L'effet « génération » explique en partie l'amélioration des compétences à l'écrit au fur et à mesure des générations. Il est surtout pour Jonas (2012) lié au développement de l'accès à l'enseignement secondaire et donc du niveau moyen d'instruction. L'élévation du niveau de diplômes et de la part de diplômés est en effet fortement liée au niveau de compétences à l'écrit (Murat et Rocher, 2016). Ainsi, le fait d'avoir un niveau d'études supérieures (par rapport à ceux qui n'ont pas dépassé le secondaire) réduit par 7 le risque de se trouver en difficulté avec l'écrit.

L'effet « cycle de vie » concerne plutôt l'idée d'une diminution des connaissances et des compétences au fil des années, du fait du déclin des compétences cognitives avec l'âge (et donc lié au vieillissement) mais aussi parce qu'elles sont peu ou ne sont plus sollicitées dans le milieu professionnel et dans la vie quotidienne. Murat et Rocher (2016) ont comparé les résultats obtenus par une même génération, entre 2004 et 2011. Ils ont constaté que la diminution des compétences à l'écrit (et en calcul), peu visible jusqu'à l'âge de 45 ans, s'accroît ensuite. Ils ajoutent : « *Dans un contexte général de vieillissement de la population, ces questions sont cruciales pour les politiques publiques* » (Murat et Rocher, 2016 : 62).

Enfin, Murat et Rocher (2016) ont relevé un accroissement du lien entre les inégalités sociales et le niveau moyen de compétences chez les jeunes générations : « *Les écarts de compétences entre générations peuvent-ils aussi s'expliquer par des changements des caractéristiques sociales de la population ?* » (Murat et Rocher, 2016 : 78). Différentes études récentes de l'OCDE, dont PISA, montrent en effet que la France est l'un des pays où la réussite scolaire est la plus liée au milieu social d'origine, ce qui en ferait l'un des pays les plus inégalitaires en matière d'éducation. Guérin-Pace (2009), en analysant la partie biographique de l'enquête de 2004, a également montré le lien entre les conditions de vie durant l'enfance et les difficultés à l'écrit à l'âge adulte.

Nous finissons cette présentation des principaux résultats en nous intéressant aux différences régionales qui ont été observées. Nous nous appuyons pour cela sur le travail de Brendler (2013) qui a comparé les 5 régions qui ont bénéficié d'une extension de l'enquête IVQ 2011 : Haute-Normandie, Île-de-France, Nord Pas-de-Calais, Picardie et Provence-Alpes-Côte d'Azur (PACA)²⁵. La proportion d'enquêtés en difficulté avec l'écrit obtenus dans ces régions est assez proche des résultats moyens nationaux (11% d'enquêtés en difficulté grave ou forte à l'écrit) : ils vont de 10% pour la région PACA à 13% en Île-de-France et Nord Pas-de-Calais. Le profil des enquêtés selon les régions varie en revanche fortement : en Île-de-France, plus de la moitié des enquêtés en grande difficulté n'ont pas été scolarisés en France (et ne relèvent donc pas de l'illettrisme). Cette répartition est deux fois plus importante que celle observée dans le reste de la métropole. Dans le Nord Pas-de-Calais en revanche, la plupart des enquêtés en difficulté ont été scolarisés en France et sont en situation d'illettrisme. Les résultats entre les régions sont davantage marqués au niveau de la population des enquêtés dirigés vers le « haut » : 26% des enquêtés « haut » ont réussi plus de 80% des exercices « complexes » en moyenne en France contre 16% dans le Nord Pas-de-Calais. Selon Brendler (2013), le niveau de compétences à l'écrit est fortement lié au niveau d'études. Mais si la proportion d'enquêtés qui ont des difficultés graves et fortes à l'écrit diminue avec l'élévation du niveau d'étude, il apparaît que dans certaines régions, le niveau d'études est moins discriminant : en Haute-Normandie, 26 % des enquêtés qui ont arrêté leurs études après l'école primaire sont en difficulté avec l'écrit, contre 62% en Île-de-France. Comme le souligne Brendler (2013), le niveau d'études est également fortement corrélé au fait d'avoir été scolarisé ou non en France (de nombreux étrangers ont un faible niveau d'étude), variable qui elle-même est corrélée au niveau de difficultés à l'écrit. Enfin, Brendler (2013) décrit un effet « contextuel » (qualifié également d'effet « régional ») selon lequel le fait de résider dans certaines régions augmente la probabilité d'être en difficulté avec l'écrit. Parmi les régions qui ont bénéficié d'une extension d'IVQ, Brendler (2013) a observé cet effet dans les régions Nord Pas-de-Calais et PACA. Il indique qu'une fois prises en compte les variables qui peuvent expliquer des difficultés à l'écrit (pays d'origine, langue de scolarisation, niveau d'études, âge, genre et origine sociale), davantage de personnes sont en difficulté avec l'écrit dans ces deux régions, ce qui permet de penser que d'autres caractéristiques propres à ces régions entrent en compte. Pour Brendler (2013), il peut s'agir du taux de chômage régional davantage marqué dans ces régions : le fait d'être éloigné longtemps du marché du travail semble ainsi avoir une incidence dans la survenue de difficultés à l'écrit. Dans ces régions, la proportion de personnes qui pratiquaient

²⁵ Il s'agit de l'ancien découpage des régions, puisque les extensions d'IVQ 2011 ont eu lieu en 2011.

une langue régionale est également plus importante que dans les autres régions et parmi cette proportion, près d'un quart des personnes sont en difficulté avec l'écrit (et ont un niveau d'études plus faible).

2.2.3.2 BILAN DE L'ENQUÊTE IVQ 2011

L'enquête IVQ de 2011 a montré que si 84% des Français n'éprouvent pas de difficulté avec un usage littéracisé de l'écrit, 16% en revanche ne bénéficient pas de compétences en lecture, écriture et compréhension suffisantes pour être à l'aise dans des situations de la vie courante. Pour 11% des Français, ces difficultés sont graves ou fortes dans au moins un des trois domaines évalués. Si pour certaines personnes les difficultés ont tendance à se cumuler, chez d'autres, en revanche, elles se concentrent sur un des trois domaines évalués. Ainsi, avoir des difficultés en production écrite ne signifie pas nécessairement avoir des difficultés en lecture ou en compréhension : Jonas (2012) montre 21% des enquêtés en difficulté en lecture ou en production écrite obtiennent de très bons résultats en compréhension orale.

L'analyse du module « biographique » en lien avec les compétences révélées dans les tests a par ailleurs montré que certaines variables sont davantage impliquées que d'autres dans la survenue de difficultés à l'âge adulte.

Jonas (2012) montre l'incidence du genre des enquêtés et du niveau d'études sur la survenue de difficultés. Les hommes sont moins à l'aise avec l'écrit que les femmes, notamment en production écrite. Le niveau d'études (notamment le fait d'avoir fait des études supérieures) diminue le risque de survenue de difficultés à l'âge adulte et cette variable est en lien avec celle de l'âge et du pays d'origine des enquêtés. Concernant l'âge, il apparaît que ce sont les générations les plus âgées qui sont les plus en difficulté dans l'enquête. Murat et Rocher (2016) montrent que ce sont également eux qui sont les moins diplômés. Cependant, l'élévation du niveau de diplôme et le développement à l'accès à l'enseignement secondaire tend à estomper l'effet de génération. L'effet « cycle de vie » est également lié à l'âge des enquêtés avec le déclin des capacités cognitives ainsi qu'une diminution des connaissances si les individus ne sont plus en activité professionnelle ou qu'ils n'ont plus besoin de mobiliser certaines compétences dans leur quotidien. Enfin, le pays d'origine et la langue de scolarisation représentent un facteur important de réussite dans les tests, cette problématique étant également en lien avec le niveau d'études puisque l'enquête montre que les personnes issues de l'immigration sont aussi moins diplômées. Enfin une extension de l'enquête IVQ réalisée dans cinq régions a donné lieu à un travail de comparaison régionale réalisée par Brendler (2013). Il montre que les personnes résidant dans certaines régions sont davantage en difficulté avec l'écrit que dans d'autres régions, tout en n'étant pas

nécessairement en situation d'illettrisme. Ainsi, en Île-de-France, plus de la moitié des enquêtés en grande difficulté n'ont pas été scolarisés en France. Par ailleurs, Brendler (2013) décrit un « effet région » qui met en œuvre d'autres variables comme le taux de chômage d'une région ou la pratique d'une langue régionale durant l'enfance (comme dans le Nord Pas-de-Calais).

2.2.3.3 PERSPECTIVES POUR ANALYSER LES COMPÉTENCES ORTHOGRAPHIQUES DE SCRIPTEURS FAIBLEMENT LITTÉRACIÉS

Le module « bas » propose un test de production écrite qui, nous le verrons dans le chapitre 4 consacré à la méthodologie, a été analysé par l'INSEE comme un test d'orthographe. Ce test représente une occasion unique et inédite d'observer les compétences orthographiques de scripteurs qui sont considérés comme faiblement littéraciés. En revanche, pouvoir bénéficier de dictées produites par les enquêtés dirigés vers le module « haut » aurait présenté plusieurs intérêts. Tout d'abord vérifier la pertinence des items de la dictée en observant les résultats obtenus par des enquêtés qui ne sont pas, dans IVQ, considérés en difficulté avec l'écrit. Ensuite, vérifier la validité des choix méthodologiques et des hypothèses : ne pas avoir de difficultés en lecture et en compréhension (qui sont les compétences qui ont été testées dans le module d'orientation) semble impliquer de ne pas avoir de difficultés particulières en orthographe (puisque seul le module « bas » propose un test d'orthographe). Enfin, comparer les données de deux populations différentes pour pouvoir mieux caractériser les compétences de l'une ou de l'autre.

Le poids de l'orthographe dans le module « bas » est important, car la maîtrise de 80% du test orthographique conditionne le résultat final de l'enquête. Un score inférieur à 80% de réussite dans l'un des trois domaines évalués dans le module « bas » plaçait les enquêtés en difficulté dans le domaine de l'écrit en général. Même s'il apparaît qu'évaluer l'écriture dans une enquête n'est pas une tâche aisée, cette mesure qui consiste à n'évaluer que l'orthographe (et particulièrement l'orthographe lexicale) est révélateur du poids de l'orthographe dans la société. Pourtant, nous l'avons vu avec les travaux de Dabène (1987) que la compétence scripturale est large. Le test de production écrite de l'enquête IVQ ne nous montre pas le rapport à l'écriture et à l'écrit des enquêtés. Il nous permet, à partir d'une vingtaine de mots, de montrer comment les enquêtés les moins littéraciés utilisent leurs connaissances orthographiques pour produire des mots (et des non-mots) et pour produire des accords simples. Il est alors possible d'appliquer différents critères selon le poids que l'on souhaite donner à tel ou tel aspect de l'écrit : dans un objectif de communication et d'un usage littéracié de l'écriture, nous proposerons de considérer moins durement les enquêtés qui ne

maitrisent pas ou peu la norme orthographique mais possèdent d'autres compétences. C'est dans cette perspective que s'inscrit notre travail.

CHAPITRE 2. L'ORTHOGRAPHE DU FRANÇAIS ET SES DIFFICULTÉS

1. L'ORTHOGRAPHE : UN PROBLÈME DE SOCIÉTÉ.....	86
1.1. BREF APERÇU HISTORIQUE	87
1.2. L'APPRENTISSAGE DE L'ORTHOGRAPHE À L'ÉCOLE : UNE APPROCHE NORMATIVE.....	89
1.3. LA QUESTION D'UNE RÉFORME	94
1.4. BILAN SUR L'ORTHOGRAPHE ET LA SOCIÉTÉ.....	98
2. PHONOGRAPHIE ET SÉMIOGRAPHIE.....	99
2.1. DE L'ORAL À L'ÉCRIT, DE L'ÉCRIT À L'ORAL.....	101
2.2. L'ÉCRITURE DU FRANÇAIS, UN SYSTÈME MIXTE	102
3. LES GRANDS PRINCIPES DU SYSTÈME ORTHOGRAPHIQUE DU FRANÇAIS : ANALYSE DES GRAPHÈMES	105
3.1. LE « PLURISYSTÈME » DU FRANÇAIS : GRAPHÈMES, SOUS-GRAPHÈMES ET ARCHIGRAPHÈMES (CATACH, 1980).....	108
3.2. L'ORGANISATION DES GRAPHÈMES SELON BLANCHE-BENVENISTE ET CHERVEL (1969).....	114
3.3. LE CHOIX DES SCRIPTEURS : ENTRE RÉGULARITÉS ET IRRÉGULARITÉS.....	119
4. ZOOM SUR UNE DIFFICULTÉ EMBLÉMATIQUE DU FRANÇAIS : LES MORPHOGRAMMES.....	122
4.1. INTRODUCTION SUR LES MORPHOGRAMMES	123
4.2. LES LETTRES MUETTES ET LE PRINCIPE DE DÉRIVATION	126
4.3. LA PRONONCIATION DES FINALES ET LA QUESTION DU « -E » FINAL MUET	130
5. BILAN SUR LES DIFFICULTÉS ORTHOGRAPHIQUES DU FRANÇAIS.....	131

Dans ce second chapitre, nous nous centrons sur l'analyse linguistique de l'orthographe du français. L'objectif est de la décrire pour comprendre son fonctionnement et ses difficultés. Cette approche est essentielle quand on évalue les compétences orthographiques, pour choisir les items sur lesquels porte l'évaluation et pour avoir conscience des difficultés auxquelles seront confrontés les scripteurs, ce qui servira notamment pour déterminer des niveaux de compétence.

Ce chapitre nous permet ainsi de cerner la complexité de l'orthographe du français : la complexité des apprentissages, la complexité des débats passionnés qui entourent l'orthographe et la complexité du système lui-même sont liées.

Dans une première partie, nous envisageons l'orthographe comme ce que l'on peut appeler un problème de société : le système orthographique s'inscrit dans une histoire, il exige un apprentissage scolaire et représente un poids du côté des scripteurs. Parallèlement, il est l'objet de débats concernant la possibilité de le réformer.

Dans une deuxième partie, nous nous centrons sur l'écriture du français en tant que système mixte. La dimension phonographique apporte les fondements de l'écriture du français, dans lequel intervient la sémiographie pour répondre aux impératifs de la communication en fournissant le sens linguistique et en permettant à l'écriture de disposer d'une forme visible et stable.

Enfin, nous proposons une description linguistique du système orthographique en lui-même et en synchronie, en nous appuyant en particulier sur les travaux du groupe HESO sur le plurisystème graphique (Catach et al, 1980) et de Blanche-Benveniste et Chervel (1969).

1. L'ORTHOGRAPHE : UN PROBLÈME DE SOCIÉTÉ

Dans cette partie, nous abordons le système orthographique d'un point de vue historique et socio-culturel, dans la dynamique de ses évolutions et en termes de rapport à l'orthographe des individus et de la société. Cette approche permettra de nous aider à comprendre le fonctionnement actuel de l'orthographe du français et à expliquer les réticences que les Français éprouvent à réformer leur orthographe. Deux facteurs sont en effet constitutifs de l'orthographe française ainsi résumés par :

« C'est en 1^{er} lieu le rapport entre ces deux systèmes fermés que sont un alphabet rigide et un système phonologique en réaménagement perpétuel ; en 2^{ème} lieu, c'est le rapport entre un type d'écriture et le rôle qui lui est imparti par la société dans laquelle il a cours : un facteur interne et un facteur externe. » (Blanche-Benveniste et Chervel, 1969 : 17).

Nous commençons par un aperçu historique et chronologique de la mise en place de l'orthographe du français. Nous ne proposerons pas de panorama précis des nombreuses études historiques existant sur le sujet (Beaulieux, 1927 ; Brunot, 1930 ; Catach, 1995 ; Blanche-Benveniste et Chervel, 1969 ; Cohen, 1973 ; Gak, 1976 ; Lucci et Nazé, 1979 ; Jaffré, 2008 ; etc.). Pour proposer une synthèse, nous utiliserons les écrits de Pellat (2001), Chervel (2008), et de Fayol et Jaffré (2008, 2014). Nous décrivons les ressources graphiques utilisés en fonction des nécessités de l'époque pour adapter l'écriture à partir de l'alphabet latin et s'adapter à l'évolution naturelle de la langue orale.

1.1. BREF APERÇU HISTORIQUE

La langue française et son orthographe possèdent une longue histoire. Les débuts de la langue française se situent à la fin du II^e siècle avant J-C avec l'arrivée des Romains en Gaule. Le peuple gaulois parle le gallo-romain, mélange de latin vernaculaire apporté par les soldats romains et de gaulois. Cette langue qui est l'ancêtre de l'occitan va évoluer avec les invasions germaniques à partir du II^e siècle après J-C pour former le français. L'influence germanique apportée par la prononciation des Francs aura de nombreuses incidences sur l'orthographe du français et provoquera notamment de nombreux changements phonétiques. Parmi les plus marquants, Fayol et Jaffré (2008) soulignent l'impact du déplacement de l'accent tonique qui va modifier le timbre des voyelles avec pour conséquence, à l'écrit, l'ajout d'un « -e » muet en finale des mots, alors que les autres langues romanes utilisent une voyelle audible. Une autre influence germanique concerne les consonnes intervocaliques avec une généralisation de l'usure de la structure phonétique qui aura pour conséquence la multiplication des formes homophones.

La France se trouve alors dans une situation de diglossie car la langue utilisée par l'administration, l'Eglise et plus généralement par l'aristocratie reste le latin quasiment classique, que ne comprend pas le peuple. Progressivement, le gallo-romain devient la langue d'oïl (qui est composée de divers dialectes) et devient la langue de la communication. L'Ordonnance de Villers-Cotterêts (1539) officialise cette situation, promouvant le français comme la langue officielle du droit et de l'administration, en lieu et place du latin et des autres langues du pays (Fayol et Jaffré, 2008).

Pour ce qui concerne la genèse de l'orthographe à proprement parler, les premières productions en français ancien (Serments de Strasbourg en 842 ; Cantilène de sainte Eulalie en 881) montrent que l'écriture était surtout phonographique. Les variations découlaient des manquements de l'alphabet latin pour représenter l'intégralité des phonèmes du français, notamment sur les distinctions ignorées du latin (composé de 24 lettres sans possibilité de transcrire les phonèmes /ʒ/ et /v/ et sans accent) : « *L'inadéquation relative de l'alphabet latin à la réalité phonologique du français explique une partie des complications de notre orthographe* » (Riegel, Pellat et Rioul, 2009 : 63). Pour Jaffré (2008), le manque de lisibilité des textes incite les copistes à la fin du 13^e siècle à chercher des solutions graphiques (en associant une tendance plus sémiographique, pour noter du sens à la phonologie, qui note du son²⁶) pour désambiguïser la langue (notamment dans la graphie des homophones). Avec l'invention de l'imprimerie au 16^e siècle et de la typographie, ce besoin de lisibilité sera accentué et associé à un désir de normalisation. La normalisation du français sera surtout portée par l'Académie française (créée en 1635) qui montre dans les premières éditions du Dictionnaire, une attitude plutôt conservatrice et traditionnelle vis-à-vis de la langue.

C'est dans la 3^e édition du Dictionnaire publiée en 1740, que sont officialisées des graphies déjà présentes dans l'usage (par exemple le remplacement du « s » par un accent circonflexe : FOREST > FORÊT ou l'ajout d'un accent aigu sur les « -e » non muets : BONTÉ), témoignage d'une volonté de l'Académie d'adopter une conduite certes traditionnelle mais qui prenne néanmoins en compte les évolutions issues de l'usage. La 5^e édition du Dictionnaire parue en 1798 ne contient plus de modifications importantes ; l'orthographe telle qu'elle est présentée à cette époque semble convenir à la bourgeoisie qui l'utilise (Jaffré, 2008).

Mais si l'orthographe semble convenir à une élite, elle pose problème pour l'apprentissage de la lecture, notamment parce que les mots contenaient de nombreuses consonnes qu'il ne fallait pas lire (OBMETTRE, NOPCES) et parce que les voyelles pouvaient avoir plusieurs réalisations phoniques (« i » pouvait se lire /i/ ou /ʒ/). Chervel (2008) rapporte qu'au XVIII^e siècle, seul un Français sur deux et une Française sur quatre sait lire et écrire. Or la démocratisation de l'orthographe à partir du XIX^e siècle, qui correspond à une demande importante de sa maîtrise dans la société, inscrit son apprentissage à l'école. Cette mesure soulève alors une nécessité de simplification et donc de rectifications pour alléger le poids de son enseignement. La circulaire Duruy (Ministre de l'Instruction Publique) en 1863, demande que l'on arrête « *d'accabler l'esprit des enfants [...] de ces analyses prétendues grammaticales*

²⁶ Nous définissons les termes de phonologie/phonographie et de sémiographie dans la section 2 de ce chapitre.

qui sont pour eux des hiéroglyphes indéchiffrables [...] » et relève « [...] *les bizarreries, les contradictions, les erreurs, les survivances de fausses subtilités* » (cité par Jaffré, 2008 : 112) de l'orthographe du français.

1.2. L'APPRENTISSAGE DE L'ORTHOGRAPHE À L'ÉCOLE : UNE APPROCHE NORMATIVE

Selon Chervel (2008), c'est au sein de l'Education Nationale qu'a été soulignée la nécessité d'enseigner l'orthographe et la complexité de mener à bien cet enseignement. Dès 1880, les activités d'écriture sont mises en place à l'école primaire, l'enjeu de l'école étant de former des citoyens qui maîtrisent la langue nationale. Les écrits demandés sont très formatés et doivent présenter des formules stéréotypées ainsi que des normes linguistiques bien établies qu'il faut appliquer. Comme le formule Garcia-Deban (2012 : 186) « *Les écritures scolaires obéissent ainsi à un modèle explicite et stable qu'il s'agit de reproduire* ». Ainsi, jusqu'en 1972, les écrits valorisés à l'école ont des caractéristiques linguistiques et stylistiques bien précises. On enseigne « *un français scolaire* » régi par des normes prescriptives et ces attentes dépassent le cadre de l'école. En effet, une fois l'apprentissage de l'écriture et donc de l'orthographe inscrite dans les programmes scolaires dès l'école élémentaire, la maîtrise orthographique se trouve de plus en plus valorisée socialement et devient en quelques années une compétence discriminante « [...] *l'orthographe française s'empare de l'école primaire et la soumet peu à peu à sa tyrannie.* » (Chervel et Manesse, 1989 : 37). Dès la fin du XIX^e siècle, le moindre examen de la fonction publique exige de solides connaissances dans cette discipline et certains concours comportent des dictées souvent très complexes. Le Brevet des collèges (créé en 1947) devient le garant d'une bonne maîtrise de l'orthographe et n'est quasiment plus qu'un examen d'orthographe (Chervel, 2008). Son obtention est un facteur de promotion sociale ; or d'après Chervel (2008), en 1879, lors de la première session du Brevet (alors Brevet élémentaire), 55% des élèves sont ajournés à l'écrit et parmi eux 48% ont obtenu une note éliminatoire à la dictée. Au début du XX^{ème} siècle, les inspecteurs primaires mais aussi le linguiste Ferdinand Brunot dans sa lettre ouverte au ministre de l'Instruction Publique, ont ouvertement accusé l'orthographe d'être « *le fléau de l'école* » (Brunot, 1905). Pourtant, bien plus tard, entre 1933 et 1957, le certificat d'études primaires et l'entrée au collège sont toujours conditionnés par un examen sélectif basé sur une dictée (Fayol et Jaffré, 2014).

Au fil des années, il apparaît que l'apprentissage de l'orthographe à l'école et le statut social lié au développement de son usage (avec le développement des nouvelles technologies) entretiennent un rapport paradoxal (Fayol et Jaffré, 2014). Ainsi, alors que la société exige une maîtrise de l'orthographe qui devient une compétence de plus en plus discriminante, l'école

consacre de moins en moins de temps à son apprentissage. En 1965 le rapport de la Commission Beslais (créée à l'initiative du ministère de l'Éducation Nationale) souligne le manque de temps dont l'école primaire dispose pour former ses élèves sur les aspects formels de la langue écrite et déplore une baisse du niveau des élèves dans cette discipline. Diverses tentatives de modification de l'orthographe ont alors vu le jour pour alléger le poids de son enseignement et simplifier son usage au quotidien mais aucune n'a entièrement abouti. Seules certaines suggestions proposées en 1965 dans le rapport Beslais se retrouvent dans les Rectifications adoptées par l'Académie française en 1990 (Jaffré, 2008).

Déjà dénoncée dans les années 1970, l'insuffisance du nombre d'heures consacrées à l'enseignement de l'orthographe à l'école primaire n'a eu de cesse de s'accroître. Dans les programmes de 2002 au cycle 3, parmi les douze heures consacrées à la langue française et à l'éducation littéraire et humaine, seul un créneau d'une heure et demie est dédié à l'apprentissage de la grammaire, de la conjugaison et de l'orthographe (Bulletin officiel de 2002 : 4). Malgré cela, l'objectif du cycle des approfondissements est de préparer l'élève à sa future vie de collégien et l'on peut lire dans les programmes de 2002 qu'en fin de cycle primaire, les élèves doivent avoir acquis les connaissances de base en vocabulaire, grammaire et orthographe :

« La maîtrise du langage et de la langue française constitue le premier pôle. Elle doit être, pour chaque enseignant du cycle 3, la priorité des priorités et une préoccupation permanente. Aucun élève ne doit quitter l'école primaire sans avoir cette assurance minimale dans le maniement du français oral et du français écrit qui permet d'être suffisamment autonome pour travailler au collège. » (Bulletin officiel de 2002 : 64).

À partir des années 1980, de nombreux travaux en linguistique voient le jour et amènent des descriptions plus objectives du fonctionnement de l'orthographe du français permettant de remettre à mal certaines idées reçues selon lesquelles l'orthographe fonctionnerait avec un ensemble de règles plutôt régulières où se maintiennent quelques exceptions (Fayol et Jaffré, 2014). C'est également dans ces années-là que les instructions officielles annoncent l'objectif de lutter contre l'échec scolaire, ce qui a pour conséquence de diversifier les situations d'écriture proposées à l'école, mais « *Les situations de production écrite scolaire sont calquées sur les situations sociales de communication écrite des adultes cultivés sans que soit véritablement prise en compte la spécificité de la communication scolaire* » (Garcia-Debanc, 2012 : 187). L'accent est mis sur l'aspect communicationnel des écrits, mais pour autant, le poids de la norme orthographique pèse toujours aussi lourd dans les attendus scolaires. Dès 1997 Catach rappelle que l'école primaire, face au décalage entre la difficulté d'apprentissage du français et le peu d'heures dont elle dispose pour l'enseigner, a été contrainte de repousser

certaines acquisitions en langue à plus tard et donc au collège et dénonce qu'en conséquence : « *Il est positivement impossible, pour toutes sortes de raisons, au meilleur enseignant du monde d'apprendre toute l'orthographe française à un enfant dans le temps qui est imparti à l'école primaire* » (Catach, 1997 : 36). Catach (1997) pour qui le nombre d'heures dédié à l'enseignement de la langue écrite n'est pas le seul problème, regrette que l'on enseigne l'orthographe aux enfants avant qu'ils n'aient compris comment le système de la langue fonctionne et elle préconise donc de passer plus de temps à expliquer le système linguistique du français avant de se focaliser sur certaines acquisitions (comme les notions de genre et de nombre). Elle rappelle que les études faites dans les années 1980 montrent qu'un enfant qui entre au collège maîtrise environ 45% de l'orthographe et de la grammaire française ; ce qui est loin des préconisations des programmes. L'enseignement secondaire a donc une mission importante en matière d'acquisition de la langue. Les programmes des IO pour le lycée de 2010, quelles que soient les filières, insistent sur la maîtrise de la langue (axée sur un enseignement du vocabulaire, de la syntaxe et des formes de discours) mais son enseignement est principalement axé sur l'étude d'œuvres littéraires et il n'est pas fait mention dans les programmes d'un enseignement spécifique de la grammaire et de l'orthographe :

« [...] il s'agit de consolider et de structurer les connaissances et les compétences acquises, et de les mettre au service de l'expression écrite et orale ainsi que de l'analyse des textes. [...]. Poursuivant l'effort qui a été conduit au cours des années du collège, le professeur veille à ce que les élèves possèdent une bonne maîtrise de l'orthographe » (MEN, BO spécial n°9 du 30 septembre 2010).

Il faudra attendre les nouveaux programmes du lycée de 2019 pour que la maîtrise de la langue soit un objectif clairement annoncé,

« Les nouveaux programmes donnent une place nouvelle à la maîtrise de la langue [...]. Le travail sur la langue occupe désormais, au lycée, une place fondamentale aux côtés de l'étude de la littérature »,

et que toutes les composantes de l'écrit figurent au programme, en lien avec l'étude de la littérature :

« Les compétences de compréhension et d'expression et les connaissances linguistiques sont complémentaires ; elles se nourrissent et s'éclairent mutuellement : une connaissance des principes de l'orthographe, de la grammaire et de la conjugaison rend l'expression plus sûre et, inversement, la possession d'un vaste vocabulaire ou l'aisance à bâtir des phrases sont renforcées par le regard réflexif que la grammaire porte sur les discours » (MEN, programme de français de seconde générale et technologique 2019 : 3).

Avant 2019, les connaissances grammaticales et orthographiques étaient donc censées être acquises en fin de collège et si cela n'était pas le cas, c'était souvent dans l'enseignement supérieur que ressurgissaient les difficultés en langue ou lors de l'insertion dans la vie professionnelle (lettre de motivation, etc). Il apparaît alors clairement que si certains apprentissages (SVT, physique, histoire-géo, etc.) peuvent être oubliés à l'âge adulte parce que ni le quotidien ni la profession exercée ne nécessitent qu'on y fasse appel, l'orthographe en revanche ne peut être oubliée sans que cela porte préjudice au scripteur qui se verra alors sanctionné par la société. Comme le soulignent Lucci et Nazé dans un ouvrage au titre provocateur *« Enseigner ou supprimer l'orthographe »* publié en 1979 *« [...] le problème de l'orthographe est de nature non seulement linguistique et pédagogique mais aussi sociale »* (1979 : 23). Dans cet ouvrage vieux de 40 ans, les auteurs soulignaient déjà l'importance accordée à l'orthographe à l'école puis par la société qui cultive *« un mythe orthographique »* (expression que nous retrouvons aussi chez de nombreux auteurs). Cette expression reflète l'idée selon laquelle cette dernière serait une sorte de graal qu'il suffirait de posséder pour que l'ensemble des capacités à l'écrit soient maîtrisées (lexique, syntaxe, expression écrite) alors que l'orthographe n'est qu'une des composantes de l'écrit. Lucci et Nazé (1979) rapportent les propos de stagiaires peu scolarisés qui souhaitent améliorer leurs performances à l'écrit et qui, pour cela, ont demandé des dictées au formateur pour *« pouvoir écrire bien »* à l'issue du stage, c'est-à-dire *« écrire sans fautes »*. Toujours selon Lucci et Nazé, il serait même courant que les parents d'enfants scolarisés en primaire aient pour habitude de faire une dictée à leurs enfants le soir après l'école lorsqu'ils estiment qu'ils n'ont pas atteint le niveau dans ce domaine. Cette pratique familiale de la dictée faite par la mère (le plus souvent) qui vient relayer l'école, a été évoquée dans l'enquête menée par Péret, Sautot

et Brissaud (2008) auprès d'étudiants qui préparent le CRPE²⁷. L'apprentissage de l'orthographe n'est donc pas un domaine réservé à l'école : la pression sociale semble telle que même dans la sphère privée, les parents (et les grands-parents²⁸) se sentent eux-mêmes investis d'une mission dans ce domaine.

Nous verrons dans les discours des enquêtés interrogés dans le cadre de l'enquête complémentaire que le poids accordé à l'orthographe est toujours très lourd (presque 40 ans après la parution de l'ouvrage de Lucci et Nazé). Les années d'apprentissage de l'orthographe ont effectivement laissé des traces. Lucci et Nazé rapportaient déjà en 1979 que « [...] *mal ou bien, on apprend à l'école l'orthographe du français. L'objet obstacle, à dominer, et si difficilement que la vie d'écolier en est définitivement affectée.* » (1979 : 7). Ainsi, pour ceux qui ont eu des difficultés, les souffrances restent vives des années plus tard et pour les autres, c'est un conservatisme dépassant l'objectivité qui témoigne de leurs années d'apprentissage de l'orthographe. Le « *mythe orthographique* » devient alors le « *malaise orthographique* ». Dans notre enquête, peu d'enquêtés ont déclaré avoir oublié ces années d'enseignement de l'orthographe. Comme le soulignent Lucci et Nazé « *Tout scripteur / lecteur investit son propre passé dans l'orthographe. [...] La sensibilité de chacun à son orthographe, qu'il veuille la conserver telle qu'il l'a si patiemment apprise ou la modifier pour tant de labeur dépensé, s'origine dans son passé et ses débuts dans la vie* » (1979 : 10).

Après leur passage au lycée (et parfois dans la vie active), c'est souvent dans l'enseignement supérieur que ressurgissent les difficultés en langue des étudiants, qui sont parfois les premiers surpris du poids accordé à leurs compétences en orthographe. Nous avons abordé cette question en amont de cette thèse en nous intéressant au public des étudiants et à la question de la continuité entre l'enseignement primaire, secondaire et l'université (Blondel, 2011, 2012). Les étudiants dont on exige qu'ils fassent des efforts dans ce domaine, sont parfois stigmatisés par des enseignants agacés de devoir intervenir sur les compétences orthographiques de leurs étudiants, et se retrouvent pour beaucoup dans une situation d'insécurité linguistique particulièrement mal vécue. Selon Jaffré et Fayol (2008), la prise en charge de ce problème a longtemps été rejetée par l'université : « [...] *les professeurs de l'enseignement supérieur s'agacent des déficiences orthographiques de leurs étudiants, mais ils ne jugent pas de leur ressort de l'enseigner* » (Jaffré et Fayol, 2008 : 208).

²⁷ CRPE : Concours de Recrutement au Professorat des Écoles.

²⁸ Voir l'analyse des discours sur l'acquisition de l'orthographe dans l'enquête complémentaire (Chapitre 5 « résultats »).

À cela peut s'ajouter un tabou, nourri par le fait que les carences en matière d'orthographe ne sont plus censées subsister à un niveau d'études supérieur au baccalauréat et par le sentiment que les connaissances devraient être acquises depuis longtemps. Pourtant, l'université a décidé de ne plus ignorer les nouveaux besoins de ses étudiants et la mise en place de dispositifs (comme le plan « réussite en licence » mis en place à la rentrée 2008), semble avoir amorcé cette nouvelle prise de conscience. Des dispositifs se développent aujourd'hui pour mieux accompagner les étudiants au niveau de l'orthographe en particulier et des compétences rédactionnelles en général²⁹.

1.3. LA QUESTION D'UNE RÉFORME

Dans cette partie consacrée à la réforme aujourd'hui, nous ne dresserons pas la liste des nombreuses tentatives de réformes et des débats qui les ont accompagnées :

« La volonté de réformer l'orthographe française, volonté qui traduit l'existence d'un « malaise orthographique », se manifeste depuis toujours ; et toute l'histoire du réformisme orthographique est jonchée de projets avortés, de tentatives interrompues, de ravaudages laborieux ou, dans le meilleur des cas, d'aménagements limités et rarement cohérents. » (Lucci et Nazé, 1979 : 9).

Il nous semble cependant important de bien avoir en tête cette question de la réforme : c'est au travers des débats actuels qu'il faut considérer les changements envisageables, au vu de notre système orthographique mais également et surtout vis-à-vis des usagers plutôt réfractaires. Le débat sur la proposition graphique NÉNUFAR en remplacement de NÉNUPHAR en est un bon exemple avec des réticences liées notamment au poids de la trace écrite et à l'importance du visuel, mais aussi à des craintes et une vision alarmiste sur la disparition généralisée du graphème « ph ». De violentes campagnes d'opposition aux réformes ont été relayées dans la presse, notamment dans *Le Monde* (30, 31 décembre 1990), avec un comité constitué pour l'occasion (le « Comité Robespierre ») qui demandait « *la guillotine morale du mépris contre les technocrates sans âme et sans pensée qui ont osé profaner notre langue* ». Les mêmes types de discours ont également eu lieu lorsque la suppression partielle des accents circonflexes a été adoptée en 1991, parmi lesquelles celui de Bernard Pivot.

²⁹ Voir par exemple les travaux de Boch et Frier (2015) dans « *Écrire dans l'enseignement supérieur : des apports de la recherche aux outils pédagogiques* ».

1.3.1. LES ENJEUX ET DIFFICULTÉS D'UNE RÉFORME

L'épineuse question d'une réforme de l'orthographe l'était déjà il y a 50 ans lorsque Peytard (cité par Lucci et Nazé, 1979) écrivait :

« [...] pour beaucoup [l'orthographe] représente l'essentiel du langage, qu'elle recouvre et domine, autoritairement. Aussi, rien ne peut être fait ni dit de l'orthographe - et cela précède toute réforme, toute méthode - si d'abord le phénomène « orthographe » n'est pas situé, démythifié, dédramatisé ».

Il semble ainsi qu'aucune réforme ne pourra aboutir tant que les représentations de l'orthographe chez les usagers n'auront pas évolué et que l'orthographe ne sera pas considérée comme étant prioritairement un objet de communication. Lucci et Nazé (1979) vont même jusqu'à se demander si l'orthographe est tout simplement « réformable ». Les travaux de la « Table ronde sur la structure de l'orthographe française » (Catach, 1973) organisée par l'équipe de recherche du CNRS dirigée par Nina Catach, ont montré que seule une réforme restreinte sur un nombre limité de phénomènes aurait la chance d'aboutir, à moins d'une peu probable révolution.

Pour Blanche-Benveniste et Chervel (1969) en revanche, la nature même de l'orthographe nécessite une refonte complète, sous peine de voir s'installer des phénomènes de réactions en chaîne (si la réforme était partielle) qui ne pourraient être résolus. Ils donnent l'exemple de la généralisation du digramme « an » qui note /ã/ et de ses conséquences :

- CENT > CANT et GENS > GANS = mais cette transformation ne peut subsister sans aménagement puisque la réalisation phonique n'est plus respectée (CANT = /kã/ et GANS = /gã/);
- il faudrait alors remplacer le graphème à l'initial : CANT > SANT et GANS > JANS ;
- mais la graphie SANT (pour le nombre) serait aussi l'homonyme de la forme verbale : IL SENT > IL SANT.

Blanche-Benveniste et Chervel (1969) concluent que la mise en place de ce type de réforme nécessiterait d'abandonner deux principes importants constitutifs de notre orthographe : la différenciation graphique des homonymes et le respect de l'étymologie.

Quoi qu'il en soit, il apparaît que cette volonté de réformer l'orthographe ainsi que les nombreux changements dans la manière de l'enseigner à l'école traduisent bel et bien l'existence d'un malaise. Lucci et Nazé (1979) se sont interrogés sur l'existence de ce malaise : est-il dû à l'orthographe elle-même, à ses enjeux dans la société et donc aux usagers eux-mêmes ou à la manière dont elle est enseignée ? L'explication se trouve probablement dans

toutes ces réponses. Pour de nombreux linguistes, une réforme profonde de l'orthographe et de sa pédagogie est nécessaire car l'orthographe ne devrait pas être si difficile à maîtriser. Elle ne devrait pas non plus être si sélective et stigmatisante :

« Si l'orthographe française devenait simple, cohérente et logique, elle serait vite démythifiée et perdrait, dès l'école, une grande partie de son importance ; elle cesserait d'être un instrument de sélection et l'apprentissage de la langue, nous l'avons vu, s'en trouverait amélioré ; quant à l'adulte peu scolarisé, il retrouverait, peut-être, une certaine sérénité face à sa langue et à son emploi et pourrait reporter tout son intérêt sur l'élaboration de ses messages écrits et sur leur contenu. » (Lucci et Nazé, 1979 : 26).

1.3.2. UNE ORTHOGRAPHE PLUS « ÉCONOMIQUE » ?

Selon les propositions avancées pour réformer l'orthographe, il semble que deux niveaux se dessinent. En premier lieu, apparaissent des propositions permettant d'alléger le poids de son enseignement en supprimant certaines formes graphiques qui n'ont plus d'utilité aujourd'hui et qui sont accusées de surcharger la mémoire des usagers. Il s'agit par exemple des pluriels des noms en « -x » qui ne sont que des variantes du cas général en « -s » (Fayol et Jaffré, 2014). Une réforme plus radicale apporte l'idée d'une orthographe plus « économique ». La notion d'économie de l'orthographe est régulièrement utilisée lorsque sont abordées des thématiques comme la complexité de l'orthographe du français ou la question d'une réforme. Cette notion suggère de recourir à des graphies qui permettraient au lecteur d'identifier plus rapidement les mots et au scripteur de gagner également en efficacité. Mais nous pouvons nous interroger sur les limites de telles graphies, entre une écriture purement phonétique qui ne transcrirait que la forme orale sans tenir compte des règles de distribution et de position spécifiques au français (pour CERISE, dans les dictées de notre enquête, cela pourrait donner *serise*, *cerize*, *sriz* ou même *criz*) ou une écriture que l'on pourrait qualifier de phonographique, qui respecterait les règles graphiques (la forme *criz* serait exclue au profit de *serise* par exemple).

Nous avons retrouvé ces graphies dans les dictées IVQ et nous avons été confrontée à des choix parfois difficiles :

- la graphie *criz* ou *sriz* permet-elle de retrouver la forme CERISE qui à l'oral peut être prononcée /səriz/ ou /sriz/ ?
- *rhum* peut-il être lu /rym/ (comme HUME se lit /ym/) ou renvoie-t-il davantage à la forme orale /rɔm/ comme la forme « -um » de PÉPLUM ou MAXIMUM ?

Ces exemples nous ont amenée à distinguer les graphies dites phonétiques de celles qui sont phonographiques pour expliquer des compétences à l'écrit différentes chez les scripteurs qui les utilisent (voir chapitre 4 qui présente notre méthodologie). Et il nous est apparu que les graphies les plus simples comme *sriz* pour /seriz/ ne sont pas nécessairement les plus aisées pour le lecteur. Jaffré et Fayol (2008) ont en effet montré que les graphies qui seraient économiques du point de vue du scripteur le seraient moins pour le lecteur et qu'il s'agirait alors de trouver dans la graphie, un équilibre entre des besoins contradictoires : les graphies phonétiques permettraient au scripteur de communiquer plus rapidement et avec moins d'effort (geste graphique, temps, fatigue, etc.), mais pour le lecteur en revanche, la tâche se complexifie :

Exemple de graphies que l'on pourrait qualifier de phonétiques recueillies dans les dictées IVQ : *sl* pour SEL ; *libreri* pour LIBRAIRIE ; *sol à nel* pour SOLENNEL.

1.3.3. UN ATTACHEMENT PARFOIS IRRATIONNEL À L'ORTHOGRAPHE

Il n'est pas possible d'envisager le système orthographique et ses évolutions de façon isolée, en dehors de ses usages, du côté des scripteurs comme des lecteurs et il faut prendre en compte aussi le rapport à l'écrit. Or, l'orthographe fait l'objet d'un attachement parfois irrationnel.

Lucci et Nazé (1979) soulignent que les plus conservateurs à l'égard de l'orthographe (qu'il s'agisse de scripteurs dits « moyens » ou de « lettrés ») invoquent souvent le caractère esthétique du français pour justifier le maintien de ses graphies et les exemples choisis pour étayer cet argument jouent davantage sur les jeux de mots qu'ils n'illustrent le caractère esthétique du français : ainsi *nouillorque* pour NEW-YORK fera évidemment sourire parce qu'il rappelle la graphie *nouille* et son signifié.

Tout scripteur a pu se dire qu'une graphie n'est pas « jolie » lorsqu'elle n'est correctement orthographiée. Dans notre enquête, certains scripteurs interrogés sur la graphie des pseudo-mots ont justifié leur choix par des arguments d'ordre esthétique. De même certains puristes, selon Lucci et Nazé, avancent que l'orthographe ne peut changer sans altérer la signification même des mots qui sont les témoins d'un héritage riche. Le côté irrationnel prend ici toute sa signification en touchant au mysticisme. Car nous l'avons vu, le plus souvent, les graphies ne portent plus les marques du passé : on peut observer que l'on n'écrit plus *doubter* ni *eschole*. Dans l'enquête de Péret, Sautot et Brissaud (2008), il transparait des discours des stagiaires qu'ils se sentent investis d'une mission dans le domaine orthographique et ils adoptent ainsi une attitude très normative (« *gardien de la norme* ») car ils ont conscience de l'importance

et de la valeur sociale qui est accordée à l'orthographe. En même temps, l'analyse de leurs écrits montre qu'ils ont une connaissance partielle du système linguistique : « *Cela ne doit pas être sans rapport avec une forme de culpabilité fréquente dans ce domaine chez les enseignants, ou a contrario avec des attitudes très normatives qui les protègent d'aléas théoriques qu'ils maîtrisent mal.* » (Péret et al., 2008 : 4). Concernant la question d'une réforme, leur discours est ambigu : tout en admettant sa nécessité, ils sont plutôt contre et leurs arguments sont ceux qui sont rapportés dans la littérature depuis les années 1990 (dont principalement celui de la permanence, avec l'idée que les changements n'apporteraient que des confusions). Ils ne sont par ailleurs que mal (voire pas) informés des rectifications de 1990, et certains, lorsqu'ils les connaissent, pensent qu'il ne s'agit que de propositions. Péret et al. (2008) rapportent aussi le cas d'une stagiaire ayant appris l'orthographe avant 1990 et qui applique, sans le savoir, l'orthographe rectifiée de 1990.

1.4. BILAN SUR L'ORTHOGRAPHE ET LA SOCIÉTÉ

Depuis l'origine, l'orthographe du français (et son enseignement) est le point d'ancrage d'enjeux et de passions considérables. Il suffit de voir le nombre important de réformes proposées, qui n'ont jamais abouti ou qui n'ont apporté que de modestes changements. Et pourtant la complexité de l'orthographe du français est avérée, car nous l'avons vu, son histoire explique que dès le départ elle est la source de nombreuses incohérences, à cause du nombre limité de lettres latines disponibles pour transcrire tous les phonèmes de la langue orale. Il en découle que sa maîtrise totale se révèle peu probable :

« [...] personne ne peut se vanter de sortir indemne de la célèbre dictée de Mérimée. Il faudrait pour cela connaître l'orthographe académique de tous les mots du dictionnaire et apprendre chaque année tous les mots nouveaux ; mots dont la forme orale ne laisse pas toujours préjuger la forme écrite. » (Blanche-Benveniste et Chervel, 1969 : 12).

Connaître l'orthographe aujourd'hui, c'est être capable de choisir la bonne graphie parmi un ensemble de formes possibles et maîtriser, lorsque les aspects lexicaux ont été surmontés, tous les aspects grammaticaux. L'orthographe grammaticale constitue une difficulté majeure de l'orthographe du français, notamment parce que la plupart des marques de morphologie sont soit homophones (le /e/ de CHANTÉ, CHANTAIT), soit silencieuses (le « -s » de FROMAGES pour /fRomaz/ ou le « -nt » de JOUENT pour /zu/)³⁰.

³⁰ Nous développons dans la section 2 du chapitre 3, les apports de la psycholinguistique concernant l'acquisition de la morphologie flexionnelle et dérivationnelle.

En matière d'orthographe, les préjugés sont anciens et bien implantés avec par exemple l'immanence d'une orthographe éternelle (Jaffré, 1992) ou la supposée baisse du niveau en orthographe avec le prétendu âge d'or de l'orthographe évoqué par Manesse et Cogis (2007). Si c'est à l'école primaire que revient principalement la formation des élèves en langue écrite, il semble que certaines acquisitions sont nécessairement repoussées. Mais comme le dénoncent Manesse et Cogis (2007 : 211) « *Peut-on en effet exiger des adolescents d'aujourd'hui qu'ils maîtrisent aussi bien la langue qu'autrefois alors qu'ils disposent de moins de temps pour la mémoriser et qu'ils ont davantage de matières à étudier ?* ». Se pose alors la difficulté d'une optimisation de l'enseignement de l'orthographe au vu des attentes à la fois de l'école mais surtout de la société et des contraintes de temps qui pèsent aujourd'hui sur l'enseignement avec des programmes chargés, liés aux nouvelles matières qui suivent l'évolution de la société comme les nouvelles technologies, les langues vivantes, etc.

Enfin, l'attitude irraisonnée qui va jusqu'à la passion chez certains usagers face au mot constitue également un frein important à toute réforme et le rejet des arguments d'ordre esthétique au profit de choix plus fonctionnels semble loin de constituer une évidence. Du côté des jeunes enseignants, le résultat de l'enquête de Péret, Sautot et Brissaud (2008) est plutôt édifiant : en décalage entre leur pratique de l'orthographe et leurs discours, ils tiennent des discours normatifs tout en ne possédant que des connaissances parcellaires du système linguistique et une connaissance limitée des processus cognitifs qui sont nécessaires à son apprentissage. Ces différents éléments les placent, lorsqu'ils commencent à enseigner, dans une insécurité linguistique : « [...] *la reconfiguration du rapport à l'orthographe des élèves qu'ils furent, en rapport à la norme des enseignants qu'ils deviennent se heurtent à quelques obstacles* » (Péret et al., 2008 : 8).

2. PHONOGRAPHIE ET SÉMIOGRAPHIE

Nous avons vu dans la partie précédente que la complexité de l'orthographe a des conséquences sociales importantes. Nous abordons à présent la complexité d'un point de vue linguistique pour comprendre le fonctionnement du système orthographique.

L'écriture du français repose sur le principe alphabétique : les lettres de l'alphabet servent à transcrire les sons et l'orthographe comporte une large part phonographique. Cependant, les correspondances entre phonèmes et graphèmes sont loin d'être transparentes comme ce serait le cas si à un graphème correspondait un phonème et inversement, comme c'est le cas dans d'autres langues (par exemple le finnois ou le roumain - Jaffré, 2008). Cette notion de transparence à laquelle s'oppose celle d'opacité est développée par Jaffré (2004, 2008). Nous y reviendrons plus loin (dans la section 2.2.1). Ces notions reflètent l'idée que différents

facteurs viennent perturber la relation entre phonèmes et graphèmes : un seul graphème qui correspond à plusieurs phonèmes, un seul phonème auquel correspondent plusieurs phonogrammes ou encore l'ajout d'éléments qui ne sont pas phonographiques mais dont la présence s'explique par l'histoire ou l'étymologie. Ainsi, un nombre non négligeable de graphèmes, pour la plupart sans lien avec l'oral ou avec lequel ils ont un lien ténu, renvoient à une signification linguistique dans laquelle intervient la sémiographie (exemple TANT, TEMPS, TAON, etc.). De fait, l'écriture du français mobilise donc un système mixte, à la fois phonographique et sémiographique : les informations apportées par le matériau écrit ne sont pas uniquement liées à la manière d'oraliser ce qui est écrit, mais comportent une dimension visuelle qui permet de donner à l'écriture une forme visible et stable.

Avant de continuer, nous définissons brièvement ce qu'est l'écriture. Dans le langage courant, la notion d'écriture est largement polysémique. Dans le TLFi (2018), le terme « écriture » ne possède qu'une seule entrée mais comporte de nombreuses acceptions qui regroupent toutes les formes d'activités scripturales :

- le code et sa composition « *Système de représentation graphique* », « *Ensemble des caractères d'un système de représentation graphique* » mais aussi « *Ensemble des formes particulières de chacun des caractères dans un système de représentation graphique* », « *Façon d'orthographier* » ;
- son action avec le geste : « *Manière personnelle de tracer les caractères* », « *Action de rédiger de la correspondance* », « *résultat de cette action* » ;
- et ses différentes significations selon les domaines : dans la littérature par exemple, l'écriture est définie comme : « *Action de composer un ouvrage littéraire* », « *Manière de s'exprimer par écrit* », en Droit « *Preuve écrite* », etc.

Au-delà de l'activité scripturale, Catach (1980) définit de manière très large l'écriture comme toute trace pouvant être reçue par le canal visuel, porteuse d'une signification et permettant de représenter le monde. Cette définition pose cependant le problème de la frontière entre graphie et écriture, par exemple pour un enfant qui trace des formes : à partir de quand peut-on parler d'écriture et non plus de dessin ?

Dans nos analyses, notre position suivra celle de Millet (1990) qui adopte une définition plus restreinte en considérant l'écriture comme la représentation du verbal, impliquant un lien étroit entre le signe graphique et le signe phonique : « [...] nous considérons qu'il y a écriture quand il y a reproduction de la parole. Nous postulons donc par là même un lien étroit entre le signe graphique et le signe phonique et au niveau syntagmatique un lien étroit entre la chaîne parlée et la chaîne écrite. » (Millet, 1990 : 71). En ce sens, l'écriture permet de représenter

graphiquement la langue et assemble pour cela des unités graphiques qui représentent des unités phoniques grâce à des conventions. Dans son usage, l'écriture implique en effet que celui qui produit et celui qui reçoit disposent d'un code commun. Fayol et Jaffré (2014) distinguent alors l'écriture, qui est un procédé technique et sa mise en œuvre avec la constitution d'une norme permettant de la pérenniser : l'orthographe.

2.1. DE L'ORAL À L'ÉCRIT, DE L'ÉCRIT À L'ORAL

Depuis Saussure, l'orthographe du français est décrite comme un système graphique complexe caractérisé par de nombreux décalages entre la forme orale et la forme écrite. Saussure dénonçait déjà au début du siècle une orthographe complexe faite de « *graphies irrationnelles* » éloignées de l'oral, au point de proposer d'adopter une écriture phonologique (Saussure, 1971 : 51). Considérant l'écriture comme un système secondaire n'ayant pour vocation que de représenter l'oral, il s'attaquait particulièrement au fait de prendre l'écrit pour la norme « *l'image pour le modèle* » (ibid.) Et lorsque les phonèmes ne sont pas représentés par un seul et même signe « *[Dans] l'orthographe du mot français 'oiseau' où pas un son du mot parlé /wazo/ n'est représenté par son signe propre* », alors pour Saussure l'écriture « *n'est pas un vêtement mais un travestissement* » (Saussure, 1971 : 52). Mais cet exemple, comme le souligne Jaffré (1992), n'illustre que la complexité du système phonographique du français où plusieurs phonogrammes peuvent être associés à un même phonème et s'éloigne donc d'un système de correspondance idéal entre l'oral et l'écrit. De plus, les phonogrammes peuvent prendre une valeur sémiographique et dans la liste établie par Catach (1980), on peut observer que de nombreux graphèmes possèdent cette double fonction (phonographique et sémiographique).

Les auteurs qui se sont penchés sur l'orthographe et le français écrit s'accordent également à souligner la distance importante entre langue orale et langue écrite. Blanche-Benveniste et Chervel (1969) le formulent de la manière suivante : « *En fait, tout se passe comme si notre écriture révélait une langue différente de la langue parlée* ». Pour Millet (1990), qui s'appuie sur les travaux de Blanche-Benveniste et Chervel (1969), l'analyse de l'orthographe française met en évidence deux systèmes qui conduisent à conclure qu'il existe une langue orale et une langue écrite, même si celles-ci entretiennent d'étroits rapports qui « *permettent, après une accoutumance plus ou moins longue, de passer aisément de l'une à l'autre, de 'traduire' donc.* » (Blanche-Benveniste et Chervel, 1969 cités dans Millet, 1990 : 78). Elle signale qu'à force de « *travestissement* », la langue écrite devient une « *forme d'expression particulièrement inadéquate à la transcription d'une seule et même forme de contenu* » (Millet, 1990 : 75). Ainsi, la notion de signe linguistique se complique : en référence à F. de

Saussure, on considère que le signe linguistique se compose d'un signifiant, qui est la réalisation phonétique individuelle de l'image acoustique et d'un signifié qui représente la face conceptuelle ou représentation mentale. Or, pour définir le décalage entre « langue écrite » et « langue parlée », Blanche-Benveniste et Chervel (1969 : 18) présente des mots au double visage : un visage oral /fam/ et un visage écrit qui s'écrit FEMME. On a bien affaire au même mot mais à deux réalisations.

Il est donc délicat de poser l'existence de deux langues : ce sont plutôt deux versants d'une même langue. Nous adoptons donc la position suivante définie par (Millet, 1990 : 77) : « *Nous pensons que les langues qui connaissent une écriture peuvent – devraient serions-nous tentée de dire - être dotées de deux systèmes d'expression sans que soit remis en cause le caractère unique de la forme du contenu.* ». Ainsi les spécialistes de linguistique de l'écrit en français sont d'accord pour affirmer la nécessité de deux approches complémentaires : une approche oraliste, selon laquelle l'écrit est à envisager principalement par rapport à l'oral et une approche autonomiste qui met en évidence la spécificité de l'écrit et le fait que le ramener à la transcription de l'oral ne suffit pas³¹.

Pour Catach (1980), ce sont les compétences du lecteur-scripteur qui déterminent le degré d'autonomie du signe graphique par rapport au signe phonique, ce qui permet de justifier, dans nos analyses, que les graphies qui utilisent le code phonographique au détriment de l'orthographe, doivent certes être placées à un niveau différent d'expertise mais correspondent néanmoins à une compétence qu'il faut analyser comme telle. Une analyse particulière qui prenne en compte l'oralité sera indispensable pour vérifier que la graphie, en tant que transcription phonographique, permet effectivement d'accéder à la forme orale d'origine.

2.2. L'ÉCRITURE DU FRANÇAIS, UN SYSTÈME MIXTE

2.2.1. LE PRINCIPE PHONOGRAPHIQUE

Le code phonographique représente l'ensemble des correspondances entre phonèmes et graphèmes. Pour transcrire l'oral, la relation idéale entre le son et la lettre que Chervel (2008) nomme « biunivoque », proposerait un système phonographique dont la sémiographie serait totalement absente. Pour Jaffré (2008 : 31) : « *L'orthographe idéale serait alors celle qui comporterait autant de graphèmes que de phonèmes, ou de syllabes.* ». Mais cet idéal

³¹ Voir Klinkenberg 2014 pour une analyse récente et très détaillée de ces positions : https://www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01397.pdf).

graphique ne se présente jamais. Dans certaines orthographe, le nombre d'unités graphiques est même bien plus important que le nombre d'unités phoniques et ce décalage constitue alors un facteur important d'irrégularité d'encodage des langues. Parmi les langues romanes, certains systèmes se rapprochent d'un idéal graphique : par exemple l'espagnol qui possède 29 graphèmes pour 25 phonèmes (Jaffré, 2008). À l'inverse, certains systèmes ont une relation inégale entre phonèmes et graphèmes, comme l'anglais qui compte selon Jaffré (2008), 561 graphèmes pour 41 phonèmes.

Le système orthographique du français ne possède pas non plus de correspondance biunivoque de type : un graphème pour un phonème, un phonème pour un graphème (comme c'est le cas avec l'alphabet phonétique international) mais dispose d'un nombre important de graphèmes par rapport au nombre de phonèmes, avec 130 graphèmes pour 36 phonèmes selon Catach et al. (1995). Dans le cas du français, les relations entre phonèmes et graphèmes peuvent être compliquées dans les deux sens : un même phonème peut être transcrit par des graphèmes différents, comme le /s/ qui peut être oralisé ainsi dans SI, AUSSI, CERISE, ÇA, ATTENTION, SIX, etc., mais réciproquement, un même graphème peut correspondre à des réalisations orales différentes (et y compris avoir une valeur muette), comme « g » dans GARAGE et DOIGT.

Pour Jaffré (2008), l'écriture d'une langue serait complètement régulière si chacun de ses phonèmes disposait d'une graphie propre différente de toutes les autres « *le plus important est que le nombre de phonogrammes soit aussi limité que possible.* » (Jaffré, 2008 : 85). Lorsque le nombre d'unités graphiques est plus grand que le nombre d'unités phoniques, cela représente un premier facteur d'irrégularité de l'écriture d'une langue, avec pour conséquence de rendre l'encodage plus difficile que le décodage. Cette situation découlerait de la genèse des orthographe, car avec le temps, certaines distinctions phoniques disparaissent alors que les graphies perdurent, créant ainsi un surplus graphique. C'est par exemple le cas des phonogrammes « o » et « au » qui n'ont plus de rôle phonographique distinct aujourd'hui (ORAGE, ATOMOBILE, etc.). Pour qualifier les orthographe qui ont un système phonographique régulier, avec une relation entre phonèmes et graphèmes la plus biunivoque possible et un nombre de phonogrammes limité, Jaffré (2008) parle de « transparence » : « *Une orthographe est d'autant plus transparente que sa phonographie est régulière et que les unités phoniques (syllabes ou phonèmes) et les unités graphiques sont dans une relation d'égalité [...]. L'orthographe idéale serait alors celle qui comporterait autant de graphèmes que de phonèmes, ou de syllabes.* » (Jaffré, 2008 : 31). Pour l'auteur, les orthographe les plus transparentes sont également celles qui font l'objet d'aménagements réguliers, en lien avec l'évolution de l'usage. Du fait de sa polyvalence orthographique élevée,

l'orthographe du français peut être considérée comme irrégulière. Cette idée de polyvalence orthographique, autrement dit le fait que les graphèmes du français répondent à plusieurs grands principes orthographiques sera détaillée dans l'analyse des graphèmes présentée dans la section 3 du présent chapitre.

2.2.2. ENTRE PHONOGRAPHIE ET SÉMIOGRAPHIE : LA NOTION DE MIXITÉ

Traditionnellement, on distingue les écritures dites « idéographiques » comme le chinois, où chaque mot est représenté par un signe et les écritures dites « phonographiques » qui reproduisent les phonèmes de la langue parlée grâce à un système alphabétique ou syllabique. Cependant, Lucci et Nazé (1979), tout comme Catach (1986) soulignent qu'en réalité la plupart des écritures sont mixtes : certains signes de l'écrit considérés comme idéographiques en chinois ont également une valeur phonique et le français utilise aussi largement le principe sémiographique en plus de son système alphabétique ; pour Jaffré et Pellat (2008 : 11) « *toute orthographe forme par définition un système mixte* » où deux dimensions coexistent pour noter du sens et du son. Selon les orthographes, le poids accordé à l'une ou à l'autre dimension diffère et aucune orthographe ne propose la configuration idéale d'un signe linguistique représenté par un graphème pouvant à la fois noter le contenu et la manière d'exprimer ce contenu à l'oral. Pour définir le poids accordé à la phonographie et à la sémiographie dans les orthographes du monde, Jaffré (2008) divise les orthographes en deux tendances sémiographiques qu'il a nommées « *sémiographie mineure* » et « *sémiographie majeure* ». La première tendance concerne les langues dans lesquelles l'homophonie-hétérographie est limitée. La deuxième tendance est surtout présente dans les orthographes anciennes comme le français, qui ont dû inventer différents procédés pour pallier l'insuffisance phonographique. Les langues réputées pour leur grande accessibilité, aussi bien en lecture qu'en écriture (comme le finnois, le turc ou le vietnamien) le sont parce que presque toutes les variations morphologiques sont prises en charge par la phonographie. La représentation sémiographique du français est justement ce qui fait sa spécificité et sa complexité, si on la compare à d'autres langues romanes, notamment avec la distinction écrite des homophones. L'analyse sémiographique des orthographes permet de comprendre pourquoi certaines sont plus difficiles à apprendre que d'autres et l'orthographe du français, nous l'avons vu, dispose d'une solide base phonographique et d'une sémiographie complexe. L'écriture du français n'est donc pas seulement une transcription des phonèmes. Pour Blanche-Benveniste et Chervel (1969), le choix entre les différentes possibilités de graphies qu'offre l'orthographe du français est toujours basé sur la considération du contenu : « [...] *toute une partie de notre code graphique, celle qui est fondée sur ce qu'on a appelé le « visage*

du mot », est foncièrement idéographique. » (Blanche-Benveniste et Chervel, 1969 : 31). Il est à noter que c'est aussi le terme « idéographique » qu'utilisent Lucci et Nazé (1979). Cependant, nous ne conserverons pas cette terminologie car elle nous semble ambiguë. Le terme « idéographie » laisse penser que les graphèmes, en plus ou en dehors de leur dimension phonographique, seraient porteurs d' « idées ». C'est parfois vrai, comme avec les morphogrammes que nous traiterons plus loin, mais on sait aussi que certains graphèmes sont relativement arbitraires, voire inexplicables. Nous préférons parler de sémiographie pour mettre l'accent sur le fait qu'il y a dans l'orthographe un principe visuel qui permet d'accéder au sens, en plus d'un principe de transcription de l'oral. Nous rejoignons donc Blanche-Benveniste et Chervel (1969 : 31) sur l'importance du « *visage du mot* ». Nous reviendrons ci-après sur ce principe sémiographique en nous appuyant notamment sur Jaffré et Pellat (2008) et Jaffré (2008). Nous détaillerons également la notion d'idéographie chez Blanche-Benveniste et Chervel (1969) quand nous exposerons leur analyse de la valeur des graphèmes (section 3.2).

En résumé, toutes les orthographes sont un compromis entre phonographie et sémiographie et toutes relèvent donc d'un système mixte (Jaffré, 2008). Si la phonographie permet de fournir « *l'infrastructure de l'orthographe à l'aide d'un inventaire clos d'unités qui correspondent d'une façon plus ou moins régulière aux formes sonores de la langue.* » (Jaffré et Pellat, 2008 : 12), la sémiographie intervient alors pour fournir une forme visible aux signes de la langue, en fournissant des formes qui « *existent d'abord et avant tout pour être lues et reconnues de la façon la moins ambiguë possible.* ». Pour le français, c'est sa sémiographie qui fait sa spécificité et aussi une partie de sa complexité, si on la compare à d'autres langues romanes, notamment avec la distinction des homophones qui lui a demandé de faire appel à différents procédés distinctifs. Pour le scripteur, c'est la compétence orthographique qui lui permettra d'accéder à la dimension sémiographique de l'orthographe, après être parvenu à associer des unités graphiques à fonction phonogrammique.

Nous développons ensuite l'idée de polyvalence orthographique du français en détaillant l'analyse des graphèmes, autrement dit le fait que les graphèmes du français répondent à plusieurs grands principes orthographiques.

3. LES GRANDS PRINCIPES DU SYSTÈME ORTHOGRAPHIQUE DU FRANÇAIS : ANALYSE DES GRAPHÈMES

Pour éclairer notre étude sur la capacité d'écriture chez les adultes, nous proposons à présent, après avoir mis en évidence le fonctionnement général de l'orthographe du français, une description du système orthographique basée sur les graphèmes. En 1988, Pellat a proposé

un inventaire des différentes définitions du graphème, terme dont l'apparition est, selon lui, tardive. Il serait apparu pour la première fois dans la littérature en 1907, dans un article publié par le linguiste Polonais Baudoin de Courtenay.

Pendant longtemps, le terme *graphème* a été considéré comme synonyme du mot *lettre* et le graphème comme l'équivalent graphique du phonème. Dans cette optique, le point de vue adopté est celui de l'oral et la façon d'envisager l'écrit est subordonnée à l'oral. Pour Pellat (1988), il a fallu attendre les années 1970 pour qu'un déplacement s'opère vers l'étude des spécificités de l'écrit et pour qu'une véritable réflexion sur le fonctionnement des systèmes d'écriture voie le jour et que le graphème soit considéré comme l'unité de base de l'écrit (Blanche-Benveniste et Chervel, 1969 ; Gak, 1976, Catach et al., 1980). L'inventaire des définitions qu'il propose met au jour quatre types de définition du graphème, dont la dernière nous intéresse plus particulièrement. Le graphème y est considéré comme « [... une unité graphique polyvalente, dont le rôle varie selon les systèmes d'écriture. » (Pellat, 1988 : 133). Cette approche fonctionnelle de l'orthographe est décrite par le groupe HESO (1980, 1995) à partir de la notion de plurisystème orthographique, qui selon lui est une bonne façon de représenter l'imbrication des différents niveaux linguistiques en jeu dans le fonctionnement de l'écriture du français, en y incluant les unités significatives (par exemple les morphogrammes), qui donnent à voir du sens. Ce système est donc mixte et cette notion de mixité des systèmes d'écriture a été développée par un linguiste comme Jaffré, avec deux principes à l'œuvre : la phonographie et la sémiographie (Fayol et Jaffré, 2014, 2016).

Nous nous appuyerons sur les travaux de Catach et collaborateurs, mais aussi sur ceux de Blanche-Benveniste et Chervel (1969) et notamment sur les valeurs des graphèmes qu'ils proposent, en lien avec la chaîne sonore. Selon la définition de Pellat (1988 : 133), Blanche-Benveniste et Chervel (1969) considèrent que « *le graphème, simple ou composé, représente le phonème* ». Leur organisation des graphèmes prend en considération la place qu'ils occupent dans le mot et leur valeur est donc liée à leur environnement à l'intérieur du mot. Ces deux approches du graphème serviront de support pour expliquer les difficultés rencontrées par les scripteurs dans la dictée d'IVQ.

Au préalable, il nous semble intéressant de commencer par quelques considérations rapides sur le code abordé du point de vue du scripteur et du lecteur, au niveau de l'encodage et du décodage.

AU PRÉALABLE : ENCODAGE ET DÉCODAGE

Nous l'avons vu, le code phonographique représente l'ensemble des correspondances entre phonèmes et graphèmes. Pour Catach (1995), le français comporte environ 130 graphèmes

pour 36 phonèmes ; pour Jaffré (2008) ce nombre peut même aller jusqu'à 175 unités. Cette divergence sur le nombre de graphèmes dépend de la méthode de comptage utilisée. Pour Jaffré (2008), cette divergence qui fait apparaître une large fourchette (comprise entre 130 et 175 unités) pour une trentaine de phonèmes, est révélatrice de la complexité de l'encodage du français.

Par exemple, doit-on considérer que dans ALCOOL, on a un dédoublement du graphème « o », avec une valeur simple, ou un graphème spécifique « oo », que l'on peut considérer comme relevant de graphèmes marginaux, mais qui se distingue de « o » comme « ç » se distingue de « c » ? À cela il faut ajouter que la suite « oo » peut se lire de deux manières, si l'on compare les mots ALCOOL et COOL : dans le second cas il s'agit bien sûr d'un anglicisme, mais le problème qui se pose au lecteur montre bien la difficulté pour analyser les graphèmes et notamment compter le nombre de graphèmes différents à l'intérieur du système orthographique.

Le fait que l'orthographe du français dispose de beaucoup plus de graphèmes que de phonèmes pose en fait davantage de difficultés en production qu'en réception « *Les analyses statistiques montrent que l'orthographe française est régulière à 96% dans les correspondances graphèmes-phonèmes. Elle ne l'est en revanche qu'à 71% pour les correspondances phonèmes-graphèmes.* » (Jaffré, 2008 : 115). Le scripteur doit donc souvent choisir entre des possibilités graphiques concurrentes et ce choix est d'autant plus important que de nombreux graphèmes ont des fonctions distinctives ou morphologiques. Pour sélectionner le bon graphème, le scripteur doit faire appel à différents processus qui varient en fonction de son âge et de son niveau de compétences.

D'un point de vue psycholinguistique, Lété (2008) rejoint les constats des spécialistes du français écrit : il souligne que les connaissances en phonologie ne suffisent pas pour être performant en lecture et en écriture. Les compétences en morphologie ont un impact important sur l'apprentissage du français, en production comme en réception et ce sont les marques flexionnelles qui posent particulièrement problème à l'écrit (marques de nombre silencieuses, formes morphologiques homophoniques comme MANGE vs MANGENT, AMI vs AMIE, etc.). Différentes observations sur l'impact de la fréquence d'occurrences de certaines formes ont été faites. On peut citer une surgénéralisation de la flexion « -s » pour marquer le nombre (ILS MANGE-S, ILS ONT MANGÉ-S), en lien avec la fréquence de certaines marques ou le fait que les connaissances finales muettes semblent mieux réalisées lorsque le mot possède un dérivé (le d de BAVARDD serait mieux réalisé que celui de BUVARDD). Nous présentons dans la section 2 du chapitre 3 (consacré aux apports de la psycholinguistique) différents travaux

consacrés à l'acquisition de la morphologie flexionnelle (Thévenin et al., 1999 ; Pacton et al., 2002 ; Fayol et Miret, 2005) et de la morphologie dérivationnelle (Pacton et Casalis, 2006 ; Béguin et al., 2010). Ainsi les connaissances flexionnelles et dérivationnelles peuvent permettre de lever certaines incertitudes quant à la graphie de finales muettes (BAVARDDGE > BAVARDD ; GRISE > GRISS), même si nous le verrons, l'apport de ces connaissances sur la réussite orthographique est à nuancer.

Ces notions de flexion et de dérivation, ou encore de morphologie et d'homophonie sont au cœur de l'analyse des difficultés d'encodage et de l'analyse du code lui-même, comme nous allons le voir à présent en traitant de l'approche linguistique de l'orthographe du français et en particulier de sa polyvalence : quelle(s) information(s) apportent les graphèmes, si ce n'est pas seulement la relation entre graphème et phonème ?

3.1. LE « PLURISYSTÈME » DU FRANÇAIS : GRAPHÈMES, SOUS-GRAPHÈMES ET ARCHIGRAPHÈMES (CATACH, 1980)

Le groupe HESO (Catach, 1980), conçoit l'orthographe du français comme faisant partie d'un « plurisystème » dans lequel trois éléments indissociables - le phonème, le graphème et le lexème - fonctionnent en synchronie. Ces trois éléments, dont l'unité graphique de base et de description est le graphème, entrent eux-mêmes en combinaison syntaxique pour former la phrase.

Dans la lignée des travaux du groupe HESO, nous partageons également la définition du graphème (que nous détaillons ci-dessous), comme l'unité de base permettant la description du code orthographique, mais aussi la description des difficultés des scripteurs et des erreurs d'encodage.

En tant qu'unité fonctionnelle de l'écrit, Catach décrit le graphème de la façon suivante :

« Le graphème représente la plus petite unité distinctive et /ou significative de la chaîne écrite, composée d'une lettre, d'un groupe de lettres (digramme, trigramme), d'une lettre accentuée ou pourvue d'un signe auxiliaire, ayant une référence phonique et/ou sémique dans la chaîne parlée. » (Catach, 1980 : 16).

Le plus souvent, le graphème transcrit donc un phonème (« s » qui représente /s/), mais il peut également être muet et avoir une fonction grammaticale (le « -s » de TABLES) ou avoir une fonction lexicale (le « -s » de gris).

3.1.1. CATÉGORIES DE GRAPHÈMES

Au sein de la structure complexe que constituent les graphèmes, Catach et al. (1980) classent ces derniers en trois catégories : les phonogrammes, les morphogrammes et les logogrammes.

1/ Les **phonogrammes** représentent le stock des graphèmes qui correspondent directement aux phonèmes et sont donc chargés de transcrire les sons. Dans cette zone centrale du système (dans laquelle on retrouve un noyau essentiel d'archigraphèmes³²), les signes s'assemblent selon leurs variantes positionnelles (formes et valeurs diverses qu'ils peuvent prendre selon leur position). Toujours selon Catach et al., l'orthographe du français est phonogrammique à 85%, ce qui lui offre une étonnante stabilité « [...] *les fondations de notre écriture sont bel et bien phonétiques ou plutôt phonogrammiques : 80 à 85% des signes d'un texte quelconque sont chargés en français de transcrire les sons.* » (Catach et al., 1980, 28). En revanche, les phonogrammes ne rendent pas compte des unités de l'écrit qui n'ont pas de lien avec l'oral. C'est pour cela que le principe phonogrammique se double d'un principe morphogrammique.

2/ Les **morphogrammes** sont les représentations graphiques des morphèmes, qui sont les plus petites unités graphiques dotées d'un sens lexical et d'un sens grammatical. Les morphogrammes fournissent donc des informations grammaticales et lexicales : désinences, flexions verbales, préfixes, suffixes, etc. Les informations qu'ils apportent sont parfois absentes de l'oral (par exemple le « -s » du pluriel nominal dans LES FROMAGESS ou le « -t » de FORTT). Selon Jaffré (2005), leur forte présence dans l'orthographe du français en fait une de ses principales spécificités mais aussi sa difficulté.

3/ Les **logogrammes** ou figures de mot, selon la définition du groupe HESO, correspondent à une zone annexe qui apporte un nouveau complément d'information de type sémiographique. Leur fonction est en effet de « *donner une image spécifique à certains mots, afin d'aider à la reconnaissance rapide de leur sens* » (Catach et al., 1995 : 262). Ainsi, le rôle des logogrammes lexicaux dans le système est de distinguer certains homonymes qui sont homophones (SEL vs SELLE vs CELLE) ou d'identifier un rôle grammatical dans le cas des logogrammes grammaticaux (A vs À, CES vs SES). Le graphème logogrammique représente alors un signe distinctif qui permet, à l'écrit, de lever une ambiguïté qui existe à l'oral et il établit ainsi un rapport de type sémiographique entre le signifiant écrit et son signifié (ex. les lettres muettes « h », « e », « x » qui permettent de distinguer HOUÉ vs HOUX vs OU et

³² Cette notion est définie dans la section 3.1.1.2 de ce chapitre.

l'accent qui permet de distinguer OU vs OÙ). Les logogrammes sont souvent des monosyllabes ou des mots très fréquents dans lesquels la graphie ne fait qu'un avec le mot dont on ne peut la dissocier (SEPT, THYM, COING).

4/ Catach et al. (1980) distingue enfin un autre ensemble composé de lettres qui ne jouent plus de rôle dans le système graphique du français : les **consonnes doubles dites graphiques**, les **lettres latines** ou **grecques**, par exemple le « h » et le « m » de HOMME. Ici, le critère fonctionnel ne permet pas de justifier toutes les lettres utilisées, car certaines lettres étymologiques ou historiques ainsi que certaines consonnes doubles ne jouent plus de rôle dans le système graphique. Ainsi, elles ne répondent à une certaine compétence, que chez le lecteur expert ou averti (connaissance du latin ou de l'histoire). Le maintien de graphies dites grecques dans les mots courants comme le « h » dans THÉÂTRE ou RHUME n'a plus d'explication aujourd'hui ; c'est ce que Catach (1980) appelle le « *reliquat peu fonctionnel et non fonctionnel* ». Ces lettres viennent compléter le système et leur suppression n'engagerait aucune des trois fonctions que remplissent les graphèmes. Pour Cogis (2005 : 51) ces lettres appartiennent à

« une zone floue, instable, voire anarchique. On n'en finirait pas d'énumérer les exceptions, les anomalies, concernant des graphies incorporées pour des raisons aujourd'hui sans intérêt, et qui ne subsistent que comme traces des choix graphiques, plus ou moins heureux, ou comme erreurs et incohérences malheureuses d'une époque antérieure »

Il est à noter que les heurs et malheurs proviennent aussi de l'évolution actuelle de la langue au contact des autres comme pour l'exemple de la graphie marginale « oo » de ALCOOL, que nous avons citée plus haut et qui s'oppose avec la même suite « oo » dans COOL ou IGLOO.

Ainsi pour Catach et al. (1980), les phonogrammes, les morphogrammes et les logogrammes se complètent au sein d'un « plurisystème » qui relève d'un système mixte permettant de noter du son (principe phonographique) et du sens (principe sémiographique).

Dans l'analyse des graphies que nous proposons (voir le chapitre 5 qui présente les résultats), nous observons principalement ce que Catach (1980) appelle la zone des phonogrammes. Nous analysons en effet les choix graphiques opérés par les enquêtés lorsque qu'ils ont dû choisir un graphème parmi d'autres pour transcrire de manière acceptable un phonème (par exemple « o », « au », « eau », etc. pour /o/). Nous abordons aussi la zone des morphogrammes avec les quelques marques présentes dans la dictée : la morphographie flexionnelle avec le pluriel nominal de FROMAGES et le pluriel verbal de SENTENT ainsi que la morphographie dérivationnelle sur FORT.

Nous retenons également de Catach (1980) que le graphème est une unité intéressante plutôt que la lettre (qui est un élément constitutif du graphème) pour partir de la forme orale et observer sa correspondance graphique. Une approche lettre par lettre, coupée de l'oral et de la langue, pose plusieurs problèmes : comment classer la graphie « oi » qui correspond à deux phonèmes ? Ou encore dans SOLENNEL, comment expliquer le graphème « e » pour transcrire le phonème /a/ indépendamment du graphème « n » qui suit ? À l'inverse, dans PAYS, c'est « a » (associé à « y ») qui transcrit un phonème /e/.

C'est en particulier pour ces différentes raisons que le groupe HESO a introduit la notion d'archigraphème que nous présentons à présent.

3.1.2. L'ARCHIGRAPHÈME

Pour Catach (1980), les archigraphèmes sont les formes maximales et stables des graphèmes correspondants dont ils sont une des réalisations en tant que représentants de la série, sans distinction liée à leur position distributionnelle. L'archigraphème a ainsi une réalisation autonome qui neutralise le graphème. Les archigraphèmes peuvent aussi être représentés par des digrammes ou des trigrammes du français (« on », « ou », « oin », etc.). Leur rendement linguistique moyen est ainsi très élevé (80 à 90% des phonogrammes correspondants selon les textes). Il existe cependant une « entorse » linguistique à la règle de fréquence avec le cas du « s » et du « z », tous deux archigraphèmes alors que le « s » intervocalique note /z/ dans 90% des cas.

Cet alphabet minimal constitué de signes fonctionnels (ou noyau graphémique) comporte pour Catach (1980) 33 graphèmes qui « [...] peuvent être considérés comme les graphèmes de base du français, suffisants pour couvrir les besoins de communication immédiate d'un scripteur français débutant (code de communication minimal) » Catach (1980 : 60).

En voici la liste (Catach, 1980) :

- pour les voyelles : A / AN, E, I/IN/ILL/Y, O/ON, OI/OIN, U/UN, EU, OU ;
- pour les consonnes : P/B, T/D, C/G, F/V, S/Z, X, CH/J, L/R, M/N, GN.

3.1.3. HIÉRARCHISATION DES GRAPHÈMES SELON CATACH (1995) : L'IMPORTANCE DE LA FRÉQUENCE

Selon Catach (1979), le noyau graphémique se compose d'archigraphèmes qui représentent eux-mêmes une classe de graphèmes et de sous-graphèmes.

Pour distinguer l'importance de certains graphèmes (graphèmes de base) des autres (sous-graphèmes et exceptions), Catach (1979) propose plusieurs critères d'identification dont le

plus important serait le degré de fréquence. Le degré de fréquence permettrait ainsi de distinguer les graphies fréquentes de celles qui sont moins fréquentes et qui sont d'un intérêt limité. Par exemple, parmi les sous-graphèmes qui font partie des formes les moins fonctionnelles, Catach (1979) cite le graphème « ph » pour la transcription de /f/.

Cette hiérarchisation des graphèmes amènerait à une hiérarchisation des erreurs : « *Il vaut mieux que l'enfant écrive foto pour photo que phontaine pour fontaine.* » (Catach, 1979 :15). Dans le cadre de notre travail, nous verrons avec les dictées examinées que pour le terme PHARMACIE, les enquêtés ont choisi préférentiellement le graphème « ph » peu fréquent, même lorsque les propositions des enquêtés étaient non conformes aux attentes orthographiques (par exemple la graphie *pharmassi*). Le graphème « ph » a ainsi été plutôt réussi, alors que l'on pourrait s'attendre à un usage du « f », chez des scripteurs qui présentent des difficultés avec l'écrit. Il faudra donc ajouter d'autres facteurs explicatifs que celui de la fréquence des graphèmes développé par Catach pour rendre compte des performances des enquêtés : la fréquence lexicale joue aussi un rôle important et nous y reviendrons dans le chapitre 3, à la lumière de travaux récents en psycholinguistique qui utilisent les notions de consistance et de fréquence.

En 1995, Catach et al. se sont appuyés sur une analyse de la fréquence des phonèmes et des graphèmes dans le lexique³³ pour proposer une répartition du système graphémique du français qui prenne en compte le niveau de complexité lors de l'apprentissage. Selon ce classement, trois niveaux de difficulté apparaissent :

Le **niveau 1** comporte 45 graphèmes : les 33 archigraphèmes auxquels sont ajoutés 12 graphèmes considérés comme les graphèmes de base qui sont les plus faciles à acquérir parce qu'ils sont les plus fréquents. Catach et al. (1995) précisent que la connaissance de ces 45 graphèmes est attendue à l'école, en fin de CE1 (en fonction des programmes qui étaient en vigueur en 1995). Parmi ces graphèmes, certains n'ont pas de concurrents : « d » pour /d/ alors que d'autres ont déjà des graphies concurrentes mais dont la liste est limitée ici : « o », « au », « eau » pour /o/.

Le **niveau 2** (ou niveau intermédiaire) est constitué de 70 graphèmes, avec 25 graphèmes supplémentaires qui devraient être acquis à la fin de l'école élémentaire.

³³ L'évaluation chiffrée des différentes zones du « plurisystème » graphique provient d'un sondage statistique réalisé en 1972 à partir de différents textes (Catach, 1980).

Le **niveau 3** auquel s'ajoutent des sous-graphèmes, comprend l'ensemble des graphèmes du système orthographique du français (environ 130).

À titre d'exemple, nous présentons dans le tableau 2, comment selon Catach et al. ont décliné les archigraphèmes « S » et « Z ». Les pourcentages indiquent la présence du graphème correspondant (par exemple « s » et « ss » notent /s/ dans 69% des cas).

niveau 1 (graphèmes de base)		niveau 2 (niveau intermédiaire)		niveau 3 (sous-graphèmes ³⁴)	
/s/	/z/	/s/	/z/	/s/	/z/
« s » : DANSE « ss » : BOSSE	« s » ³⁵ : ROSE	« t » + i : NATION	<i>pas de graphème de niv. 2</i>	« sc » : SCIENCE « sth » : ASTHME « x » : SOIXANTE	« x » : DEUXIÈME « zz » : MEZZANINE
69%	90%	3,3%			
« c » : CI « ç » : ÇA	« z » : ZÉRO				
26%	10%				

Tableau 2. Répartition des archigraphèmes « s » et « z » en trois niveaux selon Catach et al. (1980).

Cette classification en trois niveaux montre que, quel que soit le phonème considéré, le système orthographique du français offre toujours plusieurs possibilités de transcription graphique. Ainsi, même pour le phonème /y/ qui se transcrit « u » dans 100% des cas (niveau 1), la réalisation graphique se révèle très variée puisque l'on trouve, outre « u », des variantes avec diacritique (« û » : SÛR, MÛR et « ü » : AIGÛE, AMBIGÛE) et les combinaisons « eu » (IL A EU). Le scripteur se trouve ainsi confronté à des choix, comme l'est le lecteur quand il doit décoder (« eu » /y/ dans IL A EU ou /ø/ dans ADIEU). En revanche, certaines graphies exceptionnelles comme la transcription des phonèmes /e/ et /i/ dans PAYS (qui se trouve dans la dictée d'IVQ) semble se trouver en dehors de cette classification. En effet, dans cet item, le « a » se lit /e/ et le « y » se lit /i/.

Nous n'utilisons pas cette classification des graphèmes proposée par Catach et al. (1980) pour classer les graphèmes de la dictée d'IVQ et déterminer des niveaux de difficulté dans les items (ou seulement à titre exceptionnel pour souligner une difficulté en particulier). Ce choix est

³⁴ Les graphèmes qui comportent un astérisque sont des graphèmes considérés comme « exceptionnels » car ils possèderaient moins de 5 à 6 unités lexicales (plus les dérivés). Catach et al. (1980) précisent que cette liste n'est pas exhaustive. Ces sous-graphèmes n'ont pas fait l'objet de mesures statistiques pour indiquer leur pourcentage d'apparition.

³⁵ Graphème « s » placé à l'intervocalique.

principalement guidé par le fait qu'il existe aujourd'hui des outils développés récemment dans le domaine de la psycholinguistique pour les besoins de la recherche, qui sont plus actuels et plus complets car ils se basent sur des corpus de grande envergure. Manulex en particulier proposent des indices de fréquence associés à des indices de consistance, qui fournissent des informations permettant une analyse de la composition des mots selon différents critères. Par ailleurs, Catach et al. (1980) précisait que la fréquence des graphèmes est davantage fournie à titre indicatif : « *Il ne s'agit là, soulignons-le, que d'une évaluation, effectuée sur un certain nombre de textes et de mots. Le décompte des phonogrammes demanderait à être affiné, ainsi que celui des autres parties.* » (Catach et al., nouvelle présentation de la version de 1980, 2016 : 58).

Catach (1980) souligne ainsi les difficultés spécifiques liées à la notation de certains phonèmes, notamment /ã/ noté à part égale « an » ou « en » et /ẽ/ qui serait graphié majoritairement « in » et « ain » mais qui peut également s'écrire « im », « aim », « ein », « eim », « un », « yn » ou « ym ». De plus, la plupart des consonnes ont une double réalisation, à cause de la possibilité de doubler cette consonne ou d'ajouter une lettre muette finale, ce qui modifie le rendement phonogrammique des consonnes.

Cependant, même si, comme nous l'avons vu, le système graphique du français ne présente pas l'idéal de relation biunivoque entre le phonème et sa graphie qui faciliterait son apprentissage, il offre néanmoins une certaine correspondance entre les unités phoniques et les unités graphiques. « *Notre 'plurisystème' est une orthographe phonologique (à plus de 80%), dans une certaine mesure morphologique (avec une concentration extrême des 'morphogrammes grammaticaux', bien moindre des 'morphogrammes lexicaux') et enfin en partie distinctive (les distinctions des homophones grammaticaux étant de loin plus informatives que celles des homophones lexicaux).* » (Catach, 1980 : 65). L'objectif de décrire le plurisystème consiste à poser qu'il existe bel et bien des principes en jeu et qu'on peut les mettre en évidence. C'est dans cette même perspective que se situe l'approche des valeurs du graphème chez Blanche-Benveniste et Chervel (1969).

3.2. L'ORGANISATION DES GRAPHÈMES SELON BLANCHE-BENVENISTE ET CHERVEL (1969)

Blanche-Benveniste et Chervel (1969) dégagent une organisation des graphèmes qui prend en considération la place qu'ils occupent dans le mot. Leur valeur est donc liée à leur environnement à l'intérieur du mot. Les rapports univoques qu'ils entretiennent sont ainsi décrits en termes de valeur de base, valeur de position, valeur auxiliaire, valeur zéro et digramme. Les exceptions aux lois générales du code, constituées de mots qui ont un

comportement « aberrant » sont classées d'après le contenu du mot et définies en sous-codes et en zones marginales.

Ainsi, dans les monosyllabes, les graphèmes prennent souvent des valeurs exceptionnelles. Par exemple dans SIX et DIX, le « x » final est conforme au code lorsque ces mots sont suivis d'un autre nom (valeur zéro dans SIX GARÇONS et valeur /z/ dans SIX ENFANTS) mais il prend exceptionnellement la valeur /s/ en fin de syntagme (ILS SONT SIX), exception que l'on retrouve dans SOIXANTE.

Les adverbes aussi constituent un sous-code. Ainsi le graphème « e » prend la valeur /a/ dans les adverbes en « -emment » (ex. RÉCEMMENT), phénomène qui n'apparaît que dans des cas isolés sur les noms : dans FEMME qui possède la même figuration avec le redoublement du « m » ou dans SOLENNEL (mot de la dictée d'IVQ), POÊLE (dont la prononciation diffère selon les régions : /pwal/, /pwɛl/).

Ces fonctionnements phonographiques isolés montrent que le code a laissé des sous-codes se développer et cette organisation des graphèmes selon Blanche-Benveniste et Chervel (1969) nous aide à expliquer des phénomènes particuliers que nous retrouvons dans la dictée d'IVQ (la graphie de SOLENNEL et de PAYS).

3.2.1. LES VALEURS DU GRAPHÈME SELON BLANCHE-BENVENISTE ET CHERVEL (1969)

Nous détaillons à présent les différentes valeurs du graphème décrites par Blanche-Benveniste et Chervel (1969).

Les valeurs de base : pour Blanche-Benveniste et Chervel, si tous les graphèmes (sauf « j » et « v »³⁶) sont susceptibles d'avoir deux valeurs ou plus, tous ont pour valeur de base, la valeur qui est liée au minimum de contraintes. Selon ce principe, la valeur de base de « c » est /k/ plutôt que /s/. En effet, sa valeur /s/ est limitée au cas où il est suivi de « e » ou « i » et contient donc davantage de contraintes (d'où la notion de « valeur de position »).

Une écriture idéale, constituée d'un rapport biunivoque de la lettre et du son, ne comporterait ainsi que des graphèmes à valeur de base. Mais la complexité de l'écriture du français est également révélée dans la redondance de certains graphèmes qui possèdent la même valeur de base. Les auteurs relèvent ainsi les graphèmes « v » et « w », « i » et « y », « c » « k » et « q » et enfin « x » qui associe les valeurs de base de « k » et « s ».

³⁶ Pour Blanche-Benveniste et Chervel (1969), les graphèmes « j » et « v » n'ont qu'une valeur de base parce qu'il s'agit de graphèmes récents.

Valeur de position et valeur auxiliaire : le système d'écriture du français fonctionne aussi avec des règles positionnelles qui font qu'un graphème peut changer de valeur en fonction de son environnement. Ainsi, certains graphèmes ont des valeurs secondes qui sont conditionnées par leur position :

- « s » dont la valeur de base /s/ prend la valeur de position /z/ lorsqu'il est entouré de voyelles et lorsqu'il est en position de liaison (LES ENFANTS) :
 - par exemple « s » qui a une valeur de base /s/ dans SOLENNEL (mot de la dictée d'IVQ) vaut /z/ en tant que valeur de position lorsqu'il est placé entre deux voyelles comme dans CERISE (mot de la dictée d'IVQ) ;
- « c » dont la valeur de base /k/ qui devient /g/ entouré de voyelles, comme dans SECOND et ses dérivés ;
- « x » qui a pour valeur de base /ks/ devient /gz/ entre un «-e » et une autre voyelle : EXAMEN et vaut également /s/ dans certaines finales : DIX ainsi que /z/ à l'intervocalique : DIXIÈME ;
- « c » et « g » qui ont pour valeur de base /k/ et /g/ prennent comme valeur de position /s/ : CERISE, /ʒ/ : GENRE devant les graphèmes « e » et « i » :
 - par exemple si le graphème « g » a comme valeur de base /g/ (« g » à l'initiale de GARAGE), il devient /ʒ/ en valeur de position devant « e » et « i » (comme dans le 2^e « g » de GARAGE) ;
- « t » qui a pour valeur de base /t/ devient /s/ : PARTIEL devant « i » ;
- « e » dont la valeur de base est /ə/ prend dans certaines positions la valeur /e/ ou /ɛ/ : lorsqu'il est suivi de deux consonnes en syllabe fermée, il se prononce /ɛ/ : RESTE tout comme en syllabe fermée : MER, alors qu'il sera prononcé /e/ en syllabe ouverte : ÉTUDE.

Ainsi, certains graphèmes permettent de sélectionner une valeur de position : lorsque le graphème qui permet de sélectionner la **valeur de position** n'est pas prononcé (comme le « e » dans GEAI), il possède alors une **valeur auxiliaire**. Dans cet exemple, la valeur de position du « g » dépend donc de la valeur auxiliaire du « e ». Alors qu'ils ne sont pas prononcés, les graphèmes qui servent d'auxiliaires ont ainsi pour rôle de modifier la valeur d'un graphème voisin :

- par exemple le « u » de GUÉRIR qui rend au graphème « g » sa valeur de base ou le « c » de EXCÈS qui permet au « x » d'être prononcé /ks/ et non /gz/ comme dans EXISTER ;

- mais cette fonction peut également être assumée par un diacritique, par exemple la cédille dans GARÇON.

Dans le système d'écriture du français, le graphème « e » a un rôle important d'auxiliaire lorsqu'il est placé en finale des mots : « *Le e introduit un élément de régularité fondamentale, car il impose uniformément la prononciation de la consonne précédente [...]. Il est évidemment tout à fait indispensable pour la transcription, à la finale du mot, des phonèmes /v, z, s / [...].* » (BBC, 1969 : 139). Nous retrouvons ce phénomène dans les items TOMATEe, FROMAGEe, CERISEe de la dictée d'IVQ.

Enfin, le système d'écriture du français prévoit que même dans une position précise qui prenne en compte le graphème qui suit, il existe souvent plusieurs solutions graphiques pour représenter un même phonème : par exemple « an » ou « en » à l'initiale et avant le graphème « t » comme dans ANTI et ENTRER.

3.2.2. LES DIGRAMMES

Certains phonèmes³⁷ ne peuvent pas être transcrits par un graphème unique et on a alors recours à des digrammes (juxtaposition de deux lettres) ou trigrammes (pour trois lettres). Si deux digrammes : « oi » et « gn » permettent de noter deux phonèmes : respectivement /wa/ et /ɲ/, pour Blanche-Benveniste et Chervel (1969), d'autres groupes de graphèmes ont une valeur analogue. Ainsi selon cette conception, dans PAYS (mot de la dictée d'IVQ), le « y » apparaît à la fois comme un constituant de digramme pour former avec le « a » le phonème /e/, et un graphème qui vaut /i/.

Si certains digrammes sont indispensables pour la transcription des phonèmes (ex. « ou » pour /u/, « on » ou « om » pour /ɔ̃/), d'autres peuvent être considérés comme redondants puisque leurs valeurs peuvent être assumées par des graphèmes simples : par exemple « ai » (« é »), « au » (« o »), « ei » (« è »), « ph » (« f »).

Il existe enfin des associations de lettres formant des cas exceptionnels, comme le « um » de PARFUM pour transcrire le phonème /œ̃/ (dont la distinction à l'oral a tendance à disparaître) ou le « en » de SOLENNEL (mot de la dictée d'IVQ).

³⁷ Il s'agit des phonèmes /ã/, /ɛ̃/, /ɔ̃/, /œ̃/, /ø/, /u/, /j/.

Ainsi, les digrammes ne forment pas un groupe homogène, ils sont des « *palliatifs* » qui permettent de compenser un système alphabétique lacunaire, par la possibilité d'augmenter les combinaisons phonographiques :

« Tout se passe comme si le système graphique du français, incapable de s'élever à des règles générales, cherchait à parer au plus pressé en résolvant quelques problèmes de détail, sans parvenir à les résoudre tous » (Blanche-Benveniste et Chervel, 1969 : 142).

3.2.3. LA SYLLABE GRAPHIQUE ET LA VALEUR ZÉRO

L'étude des valeurs des graphèmes et des digrammes permet d'énumérer les procédés dont dispose le français pour transcrire les phonèmes. Mais certains faits, comme la répartition des graphèmes à valeur zéro (qui ne sont pas prononcés) ne peuvent pas être expliqués au niveau du graphème, mais au niveau de la syllabe. « *Les valeurs muettes des graphèmes ne peuvent évidemment être justifiées dans le cadre de la relation phono-graphique. Seul leur emplacement dans la syllabe relève d'une explication de ce type.* » (Blanche-Benveniste et Chervel, 1969 : 150).

Ainsi BOULEVARD dispose d'un découpage en syllabes graphiques (BOU-LE-VARD = 3 syllabes) et en syllabes phoniques (BOUL(E)-VARD = 2 syllabes). Le « e » ne joue pas ici un rôle de valeur auxiliaire. Il s'agit d'une valeur zéro du « e » dans le mot qui ne peut donc s'expliquer que par sa position dans la syllabe.

Pour Blanche-Benveniste et Chervel, les graphèmes muets ont une place de prédilection à la fin de la syllabe, après la voyelle si la syllabe est de type C-V comme RIE dans ÉPICE/RIE ou après la consonne si la syllabe est de type C-V-C comme TUREE dans CONFITUREE. Les graphèmes muets sont également présents à l'initiale (ou après la lettre initiale de façon moins courante comme dans RHUME) et dans les monosyllabes, entre la voyelle et la consonne (ex. TAQN). Quatre phénomènes sont alors particulièrement importants dans l'écriture du français :

- le « h » initial dans les mots non savants (ex. HOMME) ;
- les consonnes doubles (ex. POMME) ;
- les consonnes finales (ex. parmi les mots de la dictée d'IVQ : FORTT, PAYSS, TROISS) ;
- le « e » à valeur zéro (ex. parmi les mots de la dictée d'IVQ : PHARMACIE, ÉPICERIEE, LIBRAIRIEE).

Pour les graphèmes à valeur zéro, il n'y a plus de correspondance entre phonème et graphème. Pour Blanche-Benveniste et Chervel, c'est alors par l'intermédiaire des graphèmes à valeur zéro qu'une partie de l'orthographe grammaticale et de la composante

sémiographique est notée. Ces graphèmes muets s'expliquent en effet par un fonctionnement de la langue qu'ils nomment « idéographique » :

« L'absence d'un code phono-graphique au niveau du syntagme autorise, à la finale du mot, une certaine 'fantaisie' qui ne saurait gêner la lecture. C'est ici que l'idéographie prend le relais en attribuant aux graphèmes muets un rôle tout à fait spécial : celui d'établir la liaison entre mot simple et mot dérivé [...], et, de façon plus générale, de marquer le rattachement du mot à un paradigme. » (Blanche-Benveniste et Chervel, 1969 : 151).

Le principe idéographique apporte ainsi dans l'écriture un principe d'unité entre les éléments d'un paradigme grammatical (ex. le « -nt » de la troisième personne du pluriel dans ILS SENTENT) et entre les mots d'une même famille lexicale (ex. le « s » final de PAYS - mot de la dictée d'IVQ - que l'on retrouve aussi dans PAYSAGE, DÉPAYSER). Nous adopterons dans la suite de notre travail, la terminologie « sémiographie ».

3.3. LE CHOIX DES SCRIPTEURS : ENTRE RÉGULARITÉS ET IRRÉGULARITÉS

Nous avons vu que l'orthographe, du fait de sa polyvalence orthographique élevée, est considérée comme « opaque », « irrégulière » ou encore « profonde » (Jaffré, 2008 ; Jaffré et Pellat, 2008). Selon une terminologie empruntée à la psycholinguistique, elle est aussi peu « consistante » (Lété et al., 2004) comme nous y reviendrons plus loin. Autant de termes qui laissent entrevoir les difficultés que rencontre le lecteur et davantage encore le scripteur. La sélection des phonogrammes, nous l'avons vu, nécessite en effet un traitement qui est très rarement uniquement phonographique et l'écrit doit faire appel à différents procédés pour pallier l'insuffisance phonographique, dont la plupart se trouve dans la sémiographie.

La répartition des graphèmes (Blanche-Benveniste et Chervel, 1969 et Catach, 1995) montre cependant un certain nombre de constantes, avec quelques séquences décrites comme obligatoires :

- des graphèmes qui suivent les règles de distribution : par exemple le choix du digramme « om » vs « on » devant la consonne « p » comme dans COMPTER (malgré quelques exceptions comme EMBONPOINT) ;
- ou l'impossibilité de trouver certains graphèmes dans certaines positions : par exemple le graphème « q » qui est toujours suivi du graphème « u » pour former la séquence « qu » sauf en finale de certains mots : CINQ, COQ vs QUART, BARQUE³⁸.

³⁸ Des exceptions à cette règle de séquences de graphèmes non attestées (comme « qa ») décrite par Blanche-Benveniste et Chervel (1969) apparaissent aujourd'hui sous l'effet de termes empruntés. Ainsi la présence du

Lorsqu'ils sont confrontés à l'écriture, certains écoliers prennent rapidement conscience que certaines graphies ne sont pas conformes au français. Il existerait donc des régularités dans l'orthographe qui seraient apprises intuitivement (Lucci et Nazé, 1979). Blanche-Benveniste et Chervel (1969) soulignent à ce sujet que les scripteurs ont conscience de certaines régularités sous certaines conditions :

« [...] comment se fait-il que tout sujet français ait conscience d'une impossibilité de qa- initial et non pas de um- initial ? Cela tient au fait que la séquence qa n'existe à aucun endroit du mot, alors que la séquence um s'y rencontre ailleurs qu'à l'initiale : humiliation, parfum. [...] il pèse sur qa- une interdiction formelle : c'est un niveau supérieur de contrainte. » (Blanche-Benveniste et Chervel, 1969 : 120).

Ce niveau supérieur de contraintes a été observé dans les dictées recueillies dans l'enquête IVQ : nous n'avons en effet relevé que très peu de séquences qui n'existent pas, même chez les scripteurs les plus en difficulté (par exemple *qharmacie* pour Pharmacie ou *fiyncie* pour FIANCÉ).

Les dictées de notre enquête sont en effet un matériau précieux pour analyser cette expérience de la distribution des graphèmes notamment chez les scripteurs en difficulté. Parmi les règles inhérentes au français, nous reviendrons sur celles que les items de la dictée nous permettent d'observer pour voir si les scripteurs les ont bel et bien intériorisées en tant que régularités :

- le redoublement de voyelles est très peu fréquent en français, sauf pour « é » (CRÉÉ, RÉÉDITION), « o » (dans ALCOOL³⁹, mot de la dictée d'IVQ), « u » (CONTINUUM) et « i » dans certaines formes verbales (comme RIIONS) ;
- les consonnes « h, j, k, q, v, w, x » refusent la gémination ; on peut penser que d'autres, à l'inverse, s'y prêtent particulièrement bien, comme « l, m, t » et nous pourrions voir ce qu'il en est lorsque des scripteurs en difficulté produisent des consonnes géminées erronées sur un item (par exemple *tommate* ou *tomatte*). Seul l'item SOLENNEL dans la dictée nous permettra d'observer la restitution (attendue) de consonnes géminées, bien que, nous l'avons vu, cet item pose d'abord le problème de « -e » ou « -en » pour transcrire /a/ ;

segment « qa » dans QATAR, QATARI, devient (même si cette structure est rare), un candidat possible pour la réalisation graphique de /ka/.

³⁹ L'étymologie du mot ALCOOL explique la présence peu fréquente de la séquence « oo » pour /ɔ/. Il s'agit d'un emprunt de l'arabe (AL KHÔL) devenu ALCOHOL avant que l'Académie Française n'adopte la graphie moderne ALCOOL en 1835. Adoption de la graphie moderne par l'Académie (TLFI).

- aucun mot ne propose de redoublement de consonnes à l'initiale et en finale. Cette disposition est réservée à des mots empruntés à l'anglais notamment (EXPRESS, BUSINESS, etc.) ;
- à l'initiale, toutes les consonnes peuvent faire groupe avec une autre consonne sauf « j », « m », « w », « x », « z ». Dans les dictées, nous verrons lors de l'analyse des graphies non conformes aux attentes orthographiques (et qui ne respectent pas non plus la forme phonographique attendue), que même les scripteurs les plus en difficulté ont très peu associé de consonnes qui ne peuvent l'être en français⁴⁰.

Pour décrire ces phénomènes, Blanche-Benveniste et Chervel (1969 : 121) introduisent la notion de « *moule graphique* » qui se distingue de certaines séquences de graphèmes que l'on ne rencontre que dans quelques mots (voire dans un seul mot) : « *Ces graphies uniques [...] sont isolées dans le système, face à un certain nombre de moules graphiques bien établis.* ». Sans prendre le cas de séquences de graphèmes non attestées, Blanche-Benveniste et Chervel (1969) montrent que les moules graphiques s'appliquent également pour des graphies plus fréquentes (par ex : en finale, un « -u » après une voyelle est plus souvent suivi d'un « -x » que d'un « -s »). Nous retenons de ce terme de « moule graphique » l'idée que certaines suites de lettres sont plus fréquentes que d'autres à l'écrit et que cela influe très probablement le scripteur lorsqu'il n'a pas enregistré la forme orthographique attendue et qu'il doit choisir entre plusieurs combinaisons de graphèmes concurrentes. Il paraît alors très probable que la fréquence des associations de graphèmes dans le lexique du français joue un rôle important dans le choix final. Nous vérifierons cette hypothèse lorsque nous analyserons les choix graphiques des enquêtés sur les mots qui ne restituent pas la forme orthographique attendue (voir chapitre 5, section 2.3).

Concernant la prise en compte de la fréquence des graphèmes lorsque les scripteurs sont confrontés à des choix graphiques concurrents, nous retenons que la classification de Catach (1995) a entre autres mérites celui d'utiliser la fréquence des graphèmes dans les unités lexicales mais qu'elle ne prend pas en compte l'utilisation effective du système par les scripteurs. L'apport de la psycholinguistique pour rendre compte de la difficulté orthographique se révèle intéressant, comme nous le verrons plus loin avec les notions de fréquence et de consistance.

⁴⁰ Voir les chapitres 5 et 6 qui présentent les résultats de notre étude.

4. ZOOM SUR UNE DIFFICULTÉ EMBLÉMATIQUE DU FRANÇAIS : LES MORPHOGRAMMES

Comme nous l'avons vu, l'orthographe du français fonctionne donc comme un système mixte, au sein duquel cohabitent deux principes : le principe phonographique, qui est la représentation phonique des unités graphiques et le principe sémiographique par lequel les unités de l'écrit notent également des informations de l'ordre du sens. La sémiographie permet de fournir une forme visible et stable des unités de la langue. Elle complète la phonographie qui, en fournissant un inventaire figé des formes sonores de la langue, ne pourrait suffire à la transcription d'une langue en constante évolution (Jaffré, 2000). Elle est indispensable pour pallier l'insuffisance phonographique et pour distinguer à l'écrit des formes que le contexte situationnel permet de distinguer à l'oral (contexte extralinguistique, gestuelle, etc.). Pour Blanche-Benveniste et Chervel (1969), le choix d'une distinction graphique pour distinguer les homophones lexicaux lorsque le contexte suffit à accéder à la compréhension crée un phénomène de redondance :

« Lorsque l'orthographe apporte de surcroît une différenciation, comme entre les deux formes /amãd/, elle est redondante, puisque les syntagmes « manger une amande » et « payer une amende » suffisent à discriminer les signifiés. » (Blanche-Benveniste et Chervel, 1969 : 195).

Pour différencier à l'écrit les homophones et lever les ambiguïtés (notamment pour différencier les homophones lexicaux), l'orthographe du français a utilisé différents procédés qui font sa complexité ; et Fayol et Jaffré (2014) soulignent que les choix effectués sont parfois discutables. Blanche-Benveniste et Chervel (1969) rapportent que pour procéder à un classement des signifiés, une tradition de lexicographes a eu recours à l'étymologie mais rappelle que :

« ce classement, qui dépendait de connaissances étymologiques souvent mal établies, tend à être abandonné aujourd'hui. [...] selon une opinion encore répandue, l'orthographe aurait pour mission de rappeler ces distinctions lexicales héritées du passé. » (Blanche-Benveniste et Chervel, 1969 : 193).

En effet, l'orthographe confond parfois sous une même forme deux étymons différents :

- ex. le DÉ À COUDRE (= DIGITALEM) et le DÉ À JOUER (DATUM) s'orthographient de la même manière alors que sont différenciés les descendants d'un même étymon, par exemple DESSIN et DESSEIN (= DESIGNATUM) ou CONTER et COMPTER (= COMPUTARE).

Par ailleurs, Blanche-Benveniste et Chervel (1969), relèvent que l'homophonie ne possède pas de procédé graphique particulier qui permette de différencier les formes et l'on peut observer que le français a recours à toutes les possibilités :

- la présence d'un graphème muet (ex. BAN vs BANC) ;
- l'opposition entre deux graphèmes muets (ex. PORT vs PORC) ;
- l'opposition entre consonne et consonne double (ex. CANE vs CANNE) ;
- le recours à un diacritique (ex. SUR vs SÛR) ;
- l'utilisation de deux graphies très différentes (ex. CAR vs QUART).

Enfin, le recours à une distinction graphique pour différencier les homophones lexicaux n'est pas systématique, ce qui rajoute encore de la difficulté : le français ne distingue par exemple pas graphiquement le mot POÊLE selon qu'il s'agisse de la cheminée ou de l'ustensile de cuisine. Inversement, c'est parfois l'oral qui propose une distinction phonique à un mot qui n'a pas de distinction graphique dans le cas des homographes non homophones comme pour FILS (/fil/ vs /fis/) ou COUVENT (/cuv/ vs /cuvã/). Pour les jeunes apprenants, la diversité des procédés utilisés pour distinguer (ou non) les formes homophones à l'écrit sont autant de difficultés qu'il leur faudra surmonter par l'apprentissage.

Ainsi, certains choix graphiques qui se sont figés pour devenir la norme ont alors rendu son orthographe contraignante. Nous détaillons ici les principales difficultés qui relèvent de la dimension sémiographique du français : les morphogrammes et les lettres muettes (sous-partie dans laquelle nous traiterons la question de l'homophonie).

4.1. INTRODUCTION SUR LES MORPHOGRAMMES

Le rapport inégal entre phonèmes et graphèmes dans l'orthographe du français, et qui en fait sa difficulté, peut s'expliquer par la morphologie écrite. Au sens traditionnel et étymologique, la morphologie est décrite comme « l'étude de la forme des mots » et concerne presque tout ce qui concerne le fonctionnement interne des mots. La composante morphologique de la grammaire du français distingue alors :

- la morphologie flexionnelle, qui traite des variations de forme que les mots subissent dans la phrase : les accords en nombre et en genre ainsi que les temps des verbes ;
- la morphologie lexicale, qui a trait aux mécanismes permettant de former des mots et qui est essentiellement dérivationnelle : elle se construit à partir du radical (qui est la base), par l'ajout d'affixes (préfixes et suffixes) (Riegel, Pellat, Rioul, 2009).

Selon Jaffré (2008), le principe sémiographique est principalement présent dans la morphographie notamment parce les unités qui jouent un rôle grammatical primordial n'ont pas toujours de correspondance phonique : « *Certains morphogrammes comme le « s » du pluriel ou le « e » du féminin sont devenus des unités à part entière alors qu'ils ne jouissent d'aucune autonomie* » (Jaffré, 2008 : 116). Historiquement, les lettres finales utilisées ne correspondent plus aujourd'hui à des phonèmes mais elles peuvent apparaître à l'oral dans certains contextes comme lors des liaisons (LES ENFANTS JOUENT_ > LES ENFANTS JOUENT_ILS ?) ou dans les dérivés (LAIT_ > LAITEUX) même si cette oralisation peut être partielle (TEMPS_ > TEMPOREL dans lequel le « s » n'est pas récupéré dans les dérivés). Selon la conception de Catach (1980), les unités qui composent la morphographie peuvent être distinguées en **morphèmes grammaticaux** et en **morphèmes lexicaux**. À l'image des morphèmes qu'ils représentent, les morphogrammes se divisent en morphogrammes grammaticaux et en morphogrammes lexicaux.

Les **morphogrammes grammaticaux** (morphologie flexionnelle) concernent les accords en genre et en nombre ainsi que toutes les flexions verbales qui permettent de différencier les personnes, les temps et les modes (par exemple à /dɔ̃n/ correspondent trois formes graphiques différentes : DONNE_, DONNES_, DONNENT_). Les morphogrammes grammaticaux fournissent ainsi une information grammaticale et signalent les relations de ce mot avec les autres éléments de la phrase. À l'écrit, la redondance des marques à l'intérieur de la proposition permet de maintenir l'information. Fayol et Jaffré (2014) montrent que la morphographie du genre et du nombre est un domaine qui comprend de nombreuses irrégularités. Ils soulignent ainsi la polysémie du morphogramme grammatical « -s » qui complexifie la morphographie du français. En français, le « -s » utilisé comme désinence verbale marque très souvent la deuxième personne du singulier (TU MANGES_) mais aussi la 1^{ère} personne du singulier dans de nombreux verbes (JE PRENDS_, JE NOURRIS_). Il est également la marque la plus fréquente du pluriel nominal (UN CHIEN > DES CHIENS_) et entre alors en concurrence avec le « -x » dans certains mots (UN CHEVAL > DES CHEVAUX). Par ailleurs, dans certains mots, l'utilisation du « -s » ou du « -x » n'a plus de fonction distinctive (UNE - DES FOIS, UNE - DES CROIX).

Les **morphogrammes lexicaux** (morphologie dérivationnelle) sont représentés par les affixes (préfixes et suffixes) ainsi que par les finales que Catach et al. (1980 : 249) divisent entre marques directes (comme le « -t » final de COMPARTIMENT_ que l'on retrouve dans COMPARTIMENTER_) et marques indirectes (« -x »/ « -s » de HEUREUX_/HEUREUSE_).

Jaffré (2008) classe les morphogrammes lexicaux en deux catégories : ceux qui servent à distinguer le féminin et le masculin (PETIT > PETITE) et dont la consonne finale muette au masculin va être rendue sonore par le féminin (comme le « t » de PETIT > PETITE). Il s'agit ici des marques directes décrites par Catach et al. (1980). Mais la construction du féminin n'est pas toujours aussi efficace : parfois, la consonne est transformée au féminin comme dans (HEUREUX > HEUREUSE). Ce sont pour Catach les marques indirectes. Jaffré (2008) souligne que les marques lexicales ne sont pas toujours régulières au sein des séries (MUSIQUE > MUSICIEN), voire complètement en décalage entre la base et ses dérivés ; les lettres muettes ne correspondant plus : PRINTEMPS > PRINTANIER.

Enfin, au sein des morphogrammes, Jaffré (2008 : 25) distingue les morphogrammes qu'il présente comme « *un ensemble mixte, à l'intersection de la phonographie et de la morphographie, qui est relativement stable et régulier.* ». Il s'agit ainsi des morphogrammes qui en représentant à la fois du sens et du son, ont des fonctions morphographique et phonographique qui sont liées. Lorsque le morphogramme est pris en charge par la phonographie, la difficulté pour le scripteur est moins importante (NOUS MANGEONS) ; dans le cas contraire, ce dernier doit se demander s'il y a une finale muette et choisir parmi les graphèmes envisageables.

L'analyse des graphies de la dictée d'IVQ nous permettra d'analyser ces difficultés dans les cas où un morphogramme est attendu (morphogramme lexical dans FORT, morphogrammes grammaticaux dans FROMAGES et SENTENT). Nous nous appuierons également sur la configuration de cette dictée pour analyser les cas où un morphogramme grammatical était souhaitable mais non attendu (cas où les items ont été dictés sans déterminant) avec TOMATE(S), CERISE(S), ceux où un morphogramme n'était pas souhaitable avec SEL, CONFITURE, ANTI-RHUME et ceux où l'ajout d'un morphogramme grammatical est erroné : PHARMACIE, ÉPICERIE, LIBRAIRIE, QUI NE SENTENT PAS FORT, LE FIANCÉ SOLENNEL.

Les pseudo-mots (pour lesquels il s'agissait de proposer une graphie plausible, conforme au principe phonographique du français) sont aussi l'occasion de tester la manière dont les scripteurs intériorisent des « moules graphiques » relatifs aux morphogrammes en position finale (par exemple /gobar/ avec la possibilité d'ajouter un -e, un -t, un -d ou de ne pas ajouter de finale muette).

Nous allons nous centrer maintenant sur ces cas de lettres finales muettes et plus largement les lettres muettes liées au principe de dérivation.

4.2. LES LETTRES Muettes ET LE PRINCIPE DE DÉRIVATION

L'hétérographie en français, comme pour la morphologie, est à la source de la volonté de lever à l'écrit des ambiguïtés et de garder trace à l'écrit d'informations qui ne sont pas nécessairement inscrites dans la chaîne sonore. Elle se traduit par une distinction graphique entre homophones, souvent sur la finale des mots (SAIN vs SAINT, COURS vs COURT). Mais c'est parfois la forme complète qui est impactée comme dans SANG et CENT, et pourtant ces deux mots ne sont pas plus différents d'un point de vue phonographique que le sont SANG et SANS que seul le dernier graphème distingue. Très souvent, le radical graphique qui est commun au mot simple et à ses dérivés trouve son origine dans l'étymologie, notamment à des formes empruntées au latin. Blanche-Benveniste et Chervel (1969) soulignent que les mots dérivés qui ont recours au latin, et qui sont donc plus « savants », ont une écriture plutôt simple, qui se conforme aux valeurs de base du code car ils ne sont « [...] pas tenus, comme les mots de base moins savants, d'inclure dans leur graphie l'amorce de dérivations qui découleraient d'eux. » (Blanche-Benveniste et Chervel, 1969 : 166). Mais le système de dérivation en français n'a pas uniquement recours à l'étymologie. Pour faire référence aux formes de dérivation non latines (SIROP, SIROTER), Blanche-Benveniste et Chervel (1969) parlent de dérivation « populaire ». Les différentes éditions du Dictionnaire de l'Académie au XVII^{ème} et XVIII^{ème} siècle témoignent des nombreux ajustements effectués à cette époque. Par souci de simplification, certaines lettres muettes disparaissent complètement (ex. ENCORES), d'autres partiellement car elles subsistent dans leurs dérivés (FRUICT devient FRUIT mais on conserve le « c » dans FRUCTUEUX). Certaines sont conservées (CERTES) alors que d'autres sont ajoutées (SOURIS) (Fayol et Jaffré, 2014).

Pour Blanche-Benveniste et Chervel (1969 : 164), le code a parfois su tirer parti des ressources qui sont offertes par les transformations, la plus importante étant « celle qui fait passer les consonnes de leur valeur muette en fin de mot à leur valeur de base dans le dérivé : petit, petite » et la plupart des consonnes muettes en fin de mot se justifient par leur rôle dans la dérivation. Même si certaines transformations font apparaître, entre les graphèmes, des relations qui ne sont pas inscrites dans le code, le remplacement de certains graphèmes trouve néanmoins sa justification dans le fonctionnement des dérivés. Par exemple le remplacement du « v » par un « p » dans certains dérivés (comme LOUP > LOUVETEAU et dans certaines séries comme RECEVOIR > RÉCEPTION) s'explique par une régularité dans le code qui veut que « v » ne peut pas se trouver en fin de mot car il n'a jamais la valeur zéro et ne peut non plus être suivi d'une consonne. Son remplacement par un « p » n'est pas non plus le fruit du hasard : dans l'évolution phonétique qui mène du latin au français, /p/ étymologique a souvent été remplacé par /v/ en position intervocalique (exemple RIPA > RIVE).

Certaines séquences sont impossibles à noter en français et c'est le recours à la sémiographie qui permet de relier les mots à leurs dérivés et de résoudre certains problèmes phonographiques :

- la cédille de balançoire permet ainsi de maintenir le « c » du radical BALANCER tout en conservant le /s/ ;
- le maintien de la valeur de base du « t » dans la séquence « -tion » /tjõ/, qui permet de distinguer la forme verbale PORTIONS (/pɔrtjõ/) du nom PORTION(S) (/pɔrsjõ/) s'explique également par le recours à la sémiographie : la valeur /t/ du graphème « t » de la forme verbale trouve sa justification par son rapport de dérivation à l'infinitif PORTER.

Les alternatives proposées par la dérivation forment également des réseaux associatifs qui indiquent les familles de mots ainsi que les formes canoniques de la dérivation :

« Le code s'est moulé au plus juste selon les besoins de la dérivation, et n'a pas cherché à généraliser des solutions graphiques en dehors des problèmes précis qui lui étaient soumis. Il y a des séquences de phonèmes qui sont impossibles à écrire en français si l'on s'en tient aux lois du code phonographique. Comme elles n'apparaissent que dans des séries morphologiques précises, la difficulté a été résolue par des artifices qui ne valent que pour des cas particuliers de dérivation. » (Blanche-Benveniste et Chervel, 1969 : 165).

Ainsi par exemple, pour le même phonème /ɛ/, la graphie « -ain » se transforme en « -aine » et « an » (GRAIN, GRAINE, GRANULÉS) alors que la graphie « -ein » se transforme en « -eine » et « én » (PLEIN, PLEINE, PLÉNITUDE⁴¹) : *« L'idéographie se manifeste ici comme une tendance à multiplier sur le mot graphique les signes paradigmatiques. Les paradigmes indiqués sont aussi bien d'ordre lexical que grammatical. » (Blanche-Benveniste et Chervel, 1969 : 166).*

Un même radical phonique est alors souvent associé à deux radicaux graphiques différents. Nous illustrons ce phénomène par deux exemples de dérivation : l'un portant sur le suffixe et l'autre sur le préfixe.

4.2.1. DÉRIVATION SUFFIXALE

Pour le radical DON, il existe deux séries qui se distinguent par le redoublement ou non de la consonne « -n » : DON, DONATEUR, DONATION (associés à des suffixes savants qui proviennent du latin : « -ateur », « -ation ») vs DONNER, DONNEUR, MALDONNE (dont le

⁴¹ Ces exemples sont cités dans Blanche-Benveniste et Chervel, 1969 : 165.

radical -DONN intervient dans des formations d'origine française avec les formes verbales en « -er », « -e » et le suffixe « -eur »). Dans cet exemple, il y a deux niveaux de dérivation que l'orthographe manifeste par des formes graphiques distinctes que l'oral ne distingue pas. De nombreux mots suivent cette répartition⁴², ce qui introduit un facteur de cohérence et une régularité dans l'écriture de ces finales que nous devons à la sémiographie :

« Beaucoup d'alternances entre consonne simple et consonne double, qui peuvent apparaître comme des irrégularités capricieuses de l'orthographe, s'expliquent par ce type de répartition [...] la répartition se fait toujours selon le même schéma : consonne géminée pour le mot de base et consonne simple pour le dérivé. » (Blanche-Benveniste et Chervel, 1969 : 169).

Blanche-Benveniste et Chervel soulignent également que le choix de la finale semble davantage s'adapter aux paradigmes grammaticaux qu'aux paradigmes lexicaux. Par exemple, COURT prend un « t » final pour permettre de marquer le féminin COURTE alors qu'il devrait prendre le « c » de RACCOURCIR. Inversement, CRU et NU ont perdu la finale en « -d » d'origine (qui venait de CRUDITÉ et NUDITÉ) pour disposer d'une forme directement disponible pour le féminin : CRUE et NUE.

Il apparaît enfin que « [...] dans un conflit entre dérivé savant et dérivé populaire, c'est presque toujours le dérivé savant qui impose sa marque : sirop porte le p de sirupeux et non le t de siroter. » (Blanche-Benveniste et Chervel, 1969 : 171).

4.2.2. DÉRIVATION SUR LE PRÉFIXE

La dérivation sur le préfixe se construit à partir de plusieurs règles. Lorsque la préfixation s'organise sur la base d'un mot qui a une existence autonome (ex. PORTER), il y a gémination de la consonne initiale (ex. PORTER > APPORTER ou FAIM > AFFAMER). Inversement, quand il n'y a pas de base autonome, il n'y a pas de gémination (ex. APAISER : A-PAISER). Cet exemple issu de Blanche-Benveniste et Chervel (1969) montre que cette règle n'est pas systématique : si elle l'était, la préfixation de APAISER se serait organisée sur la base du mot PAIX. La construction de la préfixation dépend donc également d'autres facteurs.

Dans l'exemple VRAISEMBLABLE, le « s » a valeur de /s/ alors qu'il ne devrait pas de par sa position. Selon Blanche-Benveniste et Chervel (1969), le /s/ semble conserver le statut d'initial de SEMBLABLE alors que VRAI et SEMBLABLE ne sont pas traités comme deux mots isolés : ils

⁴² « Les finales en -onation, -oniser, -onifier, -onor [etc.] qui comportent un élément radical et un suffixe savant, s'écrivent presque toujours avec un n simple, tandis que -onnaïlle, -onnaire, -onnaïe, -onnaire sont toujours en n double. » (Blanche-Benveniste et Chervel, 1969 : 168).

auraient un statut intermédiaire entre l'autonomie et la fusion. Enfin, SALIR et RESALIR qui fonctionnent comme SENTIR et RESENTIR n'appliquent pourtant pas la règle prévue. Blanche-Benveniste et Chervel (1969) soulignent ainsi que ce classement est plutôt arbitraire.

4.2.3. BILAN SUR LES LETTRES MUETTES ET LA DÉRIVATION

En somme, par rapport à ce problème des lettres muettes et de la dérivation, si la connaissance de la dérivation peut permettre de retrouver la graphie de base comme le « t » de CHANT > CHANTER, ce type d'opération n'est pas toujours possible (CHEMINOT, employé de chemin de fer vs CHEMINEAU, qui parcourt les chemins). Les travaux en psycholinguistique⁴³ (notamment ceux de Sénéchal et al., 2000, 2006 ; Pacton et al., 2002, 2006 ; Béguin et al. 2010) montrent que si cette stratégie est efficace, elle n'est pas la plus utilisée chez les élèves qui semblent privilégier la récupération de la forme lexicale enregistrée grâce à l'effet de fréquence. Par ailleurs, d'autres travaux en linguistique comme ceux de Gak, (1976) et de Blanche-Benveniste et Chervel (1969) ont montré que la discrimination graphique des homonymes ne relevait pas d'un procédé cohérent et homogène mais a eu recours à plusieurs facteurs comme l'étymologie (avec parfois de fausses parentés), la morphologie (IL MARCHE vs ILS MARCHENT), la tradition (comme le fait de conserver une graphie qui témoigne d'une prononciation révolue) ou même l'arbitraire. Gak (1976) cite différents exemples :

- le « d » de POIDS qui le distingue de POIS résulte d'un rapprochement erroné avec le latin PONDUS alors qu'il est issu de la forme PENSUM ;
- le « c » de SCEAU qui le distingue de SEAU est arbitraire puisqu'il résulte de la forme latine SIGILLUM.

C'est alors le contexte qui est présent à l'écrit dans un texte qui doit permettre au scripteur de choisir la forme adéquate (SANG plutôt que CENT) puis les connaissances lexicales pour retrouver la forme graphique normée. Dans les dictées de notre enquête, un des items peut poser ce problème, SEL, par rapport à l'homonymie avec SELLE bien que la graphie SEL soit la plus fréquente et la plus transparente d'un point de vue phonographique. En somme, les lettres muettes font partie d'un problème généralisé en français écrit, celui des formes homophones qui sont également hétérographes. Nous verrons ce qu'il en est dans la partie résultats (chapitre 5). Les confusions seraient davantage présentes sur ce que Gak (1976) nomme les « homonymies totales », qui ont la même représentation phonique avec une distinction graphique mais une parenté de sens comme MARTYR (supplice) et MARTYRE

⁴³ Se référer au chapitre 3, section 2.2.

(supplicié). Thimonnier (1967) ajoute à cette catégorie certains homophones dont le contenu est pauvre comme certains mots outils (CES vs SES, CE vs SE, etc.), dont on sait qu'ils représentent des entrées traditionnelles de l'enseignement scolaire et des manuels sur l'orthographe.

L'exemple de la dérivation n'est qu'une des difficultés de l'orthographe du français, mais elle est très révélatrice de la complexité du système auquel le scripteur est confronté lorsqu'il écrit.

4.3. LA PRONONCIATION DES FINALES ET LA QUESTION DU « -E » MUET FINAL

Selon Blanche-Benveniste et Chervel (1969), toutes les consonnes qui se trouvent en finale peuvent se prononcer sans le secours du « e », mais celles qui sont prononcées dans cette position ont souvent un caractère étranger et appartiennent donc à des séries de mots assez restreintes. La prononciation de la consonne (ex. CLUB, SUD) apparaîtrait comme un fait marginal à côté du type qui semble plus régulier, où elle n'est pas prononcée (ex. PIED, POING). Cette conclusion est basée sur une intuition. Blanche-Benveniste et Chervel (1969) soulignent en effet qu'il est impossible de déterminer avec précision si le type le plus régulier est effectivement celui où la consonne ne se prononce pas : NEF, SUBTIL, CAP vs CLEF, GENTIL ou DRAP. Catach (1980) estime la prononciation en hausse (FJORD, NET, VIS) sous l'effet des emprunts (voir section suivante).

À l'aide des outils actuels, des requêtes simples dans les dictionnaires peuvent permettre d'affiner ces analyses. On relève ainsi dans le *Petit Robert électronique* :

- 660 mots se terminant par la lettre « -f », dont beaucoup de mots en « -if », ce qui permet de penser que la finale se prononce le plus souvent mais 832 mots se prononcent /f/ en finale, parmi lesquels ceux en « -fe » et « phe » ;
- 180 mots se terminent par la lettre « p », là aussi a priori sonore (RAP, CAP).

Il faudrait également tenir compte de la variation dans la prononciation et de la dimension régionale : par exemple MOIS D'AOÛT se prononce selon les locuteurs /mwadut/ ou /mwadu/.

Nous disposons aujourd'hui de plusieurs outils permettant d'effectuer des calculs de fréquence applicables à la finale des mots. Cette analyse sera présentée dans la partie méthodologie, d'après les mots de la dictée et en lien avec les choix graphiques des scripteurs de notre enquête.

Arrêtons-nous brièvement à présent sur le « -e » muet en français. Selon Catach (1980), l'idée du « -e » caduc final qui permettait selon une conception traditionnelle de faire prononcer la

consonne qui précède, est mise en péril par l'arrivée dans le lexique du français de nombreux emprunts dont la consonne finale est prononcée (FJORD, HAREM) et par la prononciation de plus en plus fréquente des lettres finales dans les monosyllabes (SOC, NET, MAT, VIS). « *Dans la mesure où certaines distinctions phonologiques disparaissent, la fonction des graphèmes correspondants est remise en cause [...]* » (Catach, 1980 :59). Dans la dictée d'IVQ, plusieurs items entrent ainsi en concurrence avec d'autres mots dont l'absence de « -e » final n'empêche pas la prononciation de la dernière consonne : CONFITURE (item de la dictée) vs MUR ; TOMATE (item de la dictée) vs MAT /mat/. Pour d'autres en revanche, l'absence de « -e » final en revanche conditionne la prononciation correcte⁴⁴ de la consonne finale : FROMAGE (item de la dictée) vs AIRBAG, TAG /tag/ ; CERISE (item de la dictée) vs ANIS /anis/ pour qui l'absence de « -e » final en revanche conditionne la prononciation de la consonne finale. Dans la dictée, nous relevons également les items LIBRAIRIE et ÉPICERIE qui suivent la règle de formation des noms féminins en « -i » mais qui sont concurrencés par d'autres graphies comme FOURMI, SOURIS.

5. BILAN SUR LES DIFFICULTÉS ORTHOGRAPHIQUES DU FRANÇAIS

Nous l'avons vu, l'exemple de la dérivation n'est qu'une illustration de la complexité du système orthographique du français. Pour Blanche-Benveniste et Chervel (1969), malgré le fait qu'il apporte une cohérence incontestable, le fonctionnement graphique de la dérivation obéit à des règles contradictoires. En permettant parfois de regrouper des séries sémantiques grâce à un « *moule graphique commun* » (Blanche-Benveniste et Chervel, 1969 : 173), la dérivation dissocie également graphiquement, des mots qui ont un radical commun et de nombreuses fautes d'orthographe s'expliquent aussi par une fausse association sémantique, entraînant le choix d'un radical graphique erroné. Le recours au fonctionnement de la dérivation pour renforcer les connaissances lexicales est par ailleurs limité : « *Les principes qui régissent ces marques sont trop complexes pour qu'on puisse songer à une utilisation mécanique aussi élémentaire.* », tout comme le recours à l'étymologie : « *le modèle étymologique s'applique intégralement dans un cas sur deux seulement.* » (Blanche-Benveniste et Chervel, 1969 : 173). Pour Jaffré (2008), certaines lettres (les lettres étymologiques et historiques), témoins de l'histoire de la langue ou de sa filiation avec le grec ou le latin, vont complexifier l'orthographe du français car elles ne jouent plus aucun rôle dans le système graphique. Citons les consonnes doubles qui ne sont que graphiques (les consonnes doubles étymologiques : VILLE, BATTRE ; les consonnes doubles historiques :

⁴⁴ Par exemple dans CERISE, le « s » se prononce /z/ et non /s/.

HOMME, COTTE), certaines lettres latines (TEMPS issu de *tempus*) ou grecques (RHÉTORIQUE, SATYRE). À ces difficultés s'ajoutent les lettres muettes, qui ne correspondent à aucun phonème (comme le « p » de SCULPTURE ou le « s » de PARIS). Rappelons que pour Jaffré (2008, 33) c'est « [...] à l'aune de ces lettres non phonographiques que l'on peut mesurer le degré de complexité d'une orthographe alphabétique. Plus ces lettres sont nombreuses et plus l'orthographe est opaque. [...] ». Pour l'apprentissage du français, toutes ces spécificités représentent donc une source majeure de difficultés.

Pour Jaffré (2008 : 12) qui compare l'orthographe du français avec le chinois dont les lettrés ne maîtrisent jamais en totalité les signes, « on ne sait pas le chinois, on sait du chinois et il faudrait dire de même que l'on sait de l'orthographe du français ». Fayol (2013) souligne que nous manquons de données, notamment de suivis longitudinaux concernant les traitements syntaxiques et morphosyntaxiques et leur apprentissage mais qu'il apparaît clairement que celui-ci se fait lentement et ne peut se faire sans apprentissage. « Les données recueillies par les recherches conduites au cours des trois ou quatre dernières décennies montrent clairement que l'apprentissage de l'écrit est un apprentissage et non une simple acquisition comme pour l'oral. » (Fayol, 2013 :79). Lors de l'apprentissage, les enfants découvrent que l'écrit transcrit la parole et ils doivent ensuite apprendre les associations entre phonèmes et graphèmes, mais cette connaissance de base ne suffit pas. À cause de l'irrégularité du système orthographique du français, le recours à d'autres unités est nécessaire et d'autres connaissances et procédures doivent intervenir (comme la prise en compte des probabilités d'occurrence des lettres selon leur position dans le mot). Certains mots, doivent également faire l'objet d'un apprentissage explicite et systématique de leur orthographe. Enfin, une pratique fréquente des formes orthographiques et des règles syntaxiques est nécessaire ; pratique qui fait défaut aux adultes lorsqu'ils ont quitté le système scolaire. Cela étant, « L'adulte cultivé oubliera très vite la grammaire enseignée à l'école, mais il retiendra l'orthographe. » (Lucci et Nazé, 1979 : 71). Chervel (1969 : 275) précise que l'orthographe est finalement « ce qui reste de la grammaire lorsqu'on en a tout oublié. ».

Ce chapitre nous a permis de montrer la complexité de l'orthographe du français, même si tous les points qui en font à la fois sa spécificité et sa complexité n'ont pas été abordés de manière détaillée. Nous nous sommes en effet centrés sur les points d'orthographe qui sont présents dans la dictée d'IVQ pour décrire les difficultés orthographiques d'adultes. Cette description du système nous permet d'utiliser le graphème comme élément central de notre analyse des formes graphiques que nous articulerons en deux temps : une description des compétences phonographiques des enquêtés et la conformité de leurs réponses à la norme orthographique, en lien avec leurs connaissances lexicales et morphologiques.

En retour, notre étude permettra de montrer que la connaissance fine des productions des scripteurs est une manière de mieux comprendre le système : leurs erreurs et leurs réussites sont le reflet de ce qu'ils ont ou non acquis comme étant plus ou moins fréquent dans la langue et régulier.

CHAPITRE 3. APPORTS DE LA PSYCHOLINGUISTIQUE SUR L'APPRENTISSAGE DE L'ORTHOGRAPHE DU FRANÇAIS

1. RENDRE COMPTE DE LA DIFFICULTÉ ORTHOGRAPHIQUE DES MOTS : LES NOTIONS DE CONSISTANCE ET DE FRÉQUENCE.....	136
1.1 L'INDICE DE CONSISTANCE DANS LA BASE MANULEX-INFRA.....	136
1.2 LA FRÉQUENCE ORTHOGRAPHIQUE : UNE VARIABLE PRIMORDIALE DE LA COMPÉTENCE ORTHOGRAPHIQUE.....	140
1.3 L'ACQUISITION DU LEXIQUE.....	144
1.4 BILAN SUR L'APPRENTISSAGE STATISTIQUE DE L'ÉCRIT	150
2. L'ORTHOGRAPHE DU FRANÇAIS DU POINT DE VUE DE LA PSYCHOLINGUISTIQUE.....	152
2.1. LA MAITRISE DE LA MORPHOLOGIE FLEXIONNELLE, UNE COMPÉTENCE COMPLEXE	153
2.2. LA MORPHOLOGIE DÉRIVATIONNELLE ET L'ACQUISITION DES FINALES MUETTES	160
2.3. CONCLUSION SUR LA GESTION DE LA MORPHOLOGIE.....	163

1. RENDRE COMPTE DE LA DIFFICULTÉ ORTHOGRAPHIQUE DES MOTS : LES NOTIONS DE CONSISTANCE ET DE FRÉQUENCE

Les développements récents de la recherche en psycholinguistique permettent une approche plus fine du niveau de difficulté orthographique des mots du français, avec les notions de consistance et de fréquence.

Nous présentons ici deux bases de données : Manulex (Lété, Sprenger-Charolles et Colé, 2004) et Lexique, qui fournissent un indice de la complexité des mots d'après leur consistance (dans Manulex) et leur fréquence (dans Manulex et dans Lexique).

Nous présentons ensuite les travaux de Lété (Lété et al., 2008 ; Lété, 2018) concernant la notion de consistance et de fréquence ainsi que leur impact sur la production orthographique des enfants et des adultes.

1.1 L'INDICE DE CONSISTANCE DANS LA BASE MANULEX-INFRA

1.1.1 LA CONSISTANCE ORTHOGRAPHIQUE : UN INDICE DE LA COMPLEXITÉ ORTHOGRAPHIQUE

Peereman et al. (2004) définissent la consistance orthographique comme le « degré d'incertitude » lié, en production à l'écriture d'une association entre phonème et graphème (PO = phonologie vers orthographe) ou en réception, à la lecture d'une association entre graphème et phonème (OP = orthographe vers phonologie). Nous reprenons ici, pour cette définition, l'expression de Peereman et al. mais nous relevons ici un paradoxe qui consiste à associer un degré d' « incertitude » pour déterminer le degré de consistance (et non d'inconsistance) orthographique d'un mot.

Ainsi, lorsque la variété graphémique augmente, l'ambiguïté augmente et la consistance diminue. Le degré de consistance (ou d'inconsistance) met donc en relief le « degré de variabilité » des graphèmes pour un phonème en écriture. Plus précisément, le degré de consistance est lié au degré de fréquence des associations. En d'autres termes, plus une association est rencontrée fréquemment dans les mots du français, plus cette association est considérée comme consistante et donc accessible pour le scripteur comme pour le lecteur. En révélant le niveau de complexité orthographique, l'indice de consistance permet ainsi de prédire la difficulté à écrire ou à lire un mot. Pour calculer cet indice de consistance, Lété, Sprenger-Charolles et Colé (2004) ont développé un outil informatique, la base de données Manulex.

1.1.2 LA BASE DE DONNÉES MANULEX : PRÉSENTATION ET FONCTIONNALITÉS

Le développement de la base de données Manulex (Lété, Sprenger-Charolles et Colé, 2004) a été motivé par les limites des bases qui sont actuellement disponibles pour étudier l'acquisition de la lecture et de l'écriture. Tout d'abord, la plupart des bases existantes sont basées sur des corpus destinés à des adultes, ce qui rend leur utilisation inappropriée pour des enfants ou pour des adultes faiblement littéraciés (par exemple la base Frantext qui est issue de textes littéraires et contemporains). Par ailleurs, Manulex ajoute aux calculs de fréquence d'occurrence des mots (qui est déjà proposée dans d'autres outils), un calcul de la consistance des relations entre phonèmes et graphèmes. On sait en effet que la régularité des correspondances entre phonèmes et graphèmes est un facteur important de réussite orthographique ; facteur qu'il faut prendre en considération, tout comme la fréquence des mots, pour déterminer la complexité à orthographier un mot.

La base Manulex propose ainsi une description complète de la langue française écrite qui est adressée aux enfants à l'école primaire. Elle calcule en effet la fréquence d'occurrences des mots rencontrés par les enfants à partir d'un corpus de 54 manuels scolaires (ce qui représente 1,9 millions de mots et 48886 entrées lexicales). Nous utilisons l'outil Manulex-Infra (Ortégua et Lété, 2010) qui propose en plus des fonctionnalités de Manulex, une prise en compte des formes orthographiques (marques de flexion notamment), rendue possible grâce à une présentation des indices sur trois parties du mot (au maximum) : par exemple pour LIBRAIRIE : « LI- » (début de mot), « -BRAI- » (milieu) et « -RIE » (fin de mot).

1.1.2.1 CALCUL DE L'INDICE DE CONSISTANCE

Pour calculer les indices de consistance, la base de données Manulex-Infra⁴⁵ fournit donc une mesure statistique (en %) de la consistance orthographique des mots dans le sens PO : « phonologie-orthographe » (codage) et OP : « orthographe-phonologie » (décodage). La description des associations entre phonologie et orthographe (et inversement) a été effectuée sur la base de 37 phonèmes et 125 graphèmes avec 290 associations PO et OP. Pour calculer le degré de consistance des mots qui intéressent notre étude (lorsqu'il s'agit de les écrire, et donc d'associer un graphème au phonème), nous utiliserons les indices de Manulex-Infra dans le sens « PO », c'est-à-dire de la phonologie vers l'orthographe. Cette mesure est basée sur la fréquence d'association d'un phonème avec un graphème sur l'ensemble des mots de la base,

⁴⁵ Les bases Manulex et Manulex-Infra possèdent de nombreuses fonctionnalités dont nous ne détaillons pas l'ensemble, mais uniquement celles qui intéressent notre étude.

avec une distinction qui est faite sur la place de cette association dans le mot : début, milieu ou fin du mot. Cette distinction est intéressante dans la mesure où les valeurs de consistance peuvent varier fortement selon la position des unités graphémiques dans le mot, ce qui nous permettra d'identifier la partie du mot qui est la moins consistante et qui est susceptible de davantage poser problème aux enquêtés dans la dictée d'IVQ. Dans le calcul, la valeur de consistance est comprise entre 0 (consistance minimale) et 100 (consistance maximale). Nous présentons le détail du calcul qui permet d'interpréter l'indice de consistance dans le sens Phonologie-Orthographe du mot MAIN (/mɛ̃/ qui devient MAIN) :

- calcul de la fréquence d'association du phonème /m/ avec le graphème « m » en début de mot : le résultat est multiplié par 100 = ici, il était de 1, donc **l'indice est 100**. Cet indice de 100 est interprété comme suit : « *sur 100 apparitions du phonème /m/ en début de mot, il s'écrit toujours « m »*. Cette association est donc très consistante ;
- calcul de la fréquence d'association du phonème /ɛ̃/ avec le graphème « ain » en début de mot : ici, il est de 0,18, donc **l'indice est 18** : « *sur 100 apparitions du phonème /é/ en fin de mot, il s'écrit « ain » dans 18% des cas* ». Cette association est donc peu consistante ;
- l'indice de consistance total est ensuite obtenu en faisant la moyenne des indices (début, milieu et fin de mot). Pour MAIN, **l'indice de consistance est de 59**.

Pour estimer la fréquence et la consistance des relations entre les phonèmes et les graphèmes, les entrées lexicales de Manulex-Infra ont été segmentées en unités graphémiques selon le principe général qu'à un phonème correspond un graphème « /m/ = m) ou, beaucoup plus rarement, qu'à deux phonèmes correspond un graphème (par exemple dans le cas des digrammes : /wa/ = « oi »). Ortéga et Lété (2010) soulignent que ce découpage n'est généralement pas problématique pour le français mais que des choix ont dû être faits dans plusieurs cas : « *Les choix réalisés étaient gouvernés par un second principe selon lequel les segmentations devaient maximiser la mise en évidence des inconsistances de prononciations des chaînes orthographiques.* » (Ortéga et Lété, 2010 : 3). La base propose 125 graphèmes. Nous rappelons que le nombre de graphèmes varie selon le mode de calcul ; la fourchette étant comprise entre 130 (Catach, 1980) et 175 unités (Fayol et Jaffré, 2008) (voir section 3 du chapitre 2).

1.1.2.2 PRÉSENTATION DE DEUX VARIABLES : LE NIVEAU D'EXPERTISE ET LE TYPE DE CALCUL

La base Manulex-Infra permet également de calculer la consistance orthographique sur la syllabe. Lété (2008) montre que la consistance « Phonologie-Orthographe » (en production) est plus basse lorsqu'elle est calculée sur la syllabe (plutôt que sur le phonème), parce qu'une

unité syllabique correspond à plusieurs associations phonèmes-graphèmes. La comparaison des indices de consistance entre les mots devient notamment problématique lorsque le calcul s'est effectué sur la syllabe et que les mots n'ont pas le même nombre de syllabes ni le même nombre de phonèmes à l'intérieur des syllabes. Aussi, les indices de consistance des mots qui concernent notre étude ont été calculés sur les phonèmes.

Enfin, il est possible d'appliquer différentes variables sur les calculs dont nous présentons les deux principales.

Trois niveaux d'expertise : la base permet de sélectionner un niveau d'expertise constitué des manuels adressés à l'enfant durant sa scolarité. Trois niveaux d'expertise ont été déterminés en fonction du niveau de lecture qui est attendu en classe :

- en CP, vers l'âge de 6 ans, lorsque les enfants commencent à apprendre du lexique ;
- en CE1, vers l'âge de 7 ans, lorsque les enfants commencent à enrichir leur lexique orthographique ;
- au cycle 3 (du CE2 au CM2⁴⁶), entre 8 et 11 ans, période durant laquelle les enfants enrichissent leur vocabulaire grâce à une exposition répétée à l'écrit (Ortéga et Lété, 2010).

Une fonction dans Manulex-Infra permet de choisir les trois niveaux et de disposer d'une moyenne permettant d'accéder à l'ensemble des manuels du CP au CM2 et donc à l'intégralité du lexique (ou du moins une bonne partie) auquel l'enfant est exposé à l'école primaire.

Un calcul possible par « type » et par « token » : les calculs peuvent être effectués par « type » ou par « token ». Dans notre étude, nous avons choisi de calculer les indices en fonction de la fréquence réelle des mots dans les textes (par « token ») car nous postulons que la fréquence de rencontre des mots dans les manuels peut avoir un impact sur leur restitution en situation de scriptio, même chez des adultes qui ne sont plus confrontés à des manuels de lecture.

Plus récemment Peereman, Sprenger-Charolles et Massaoud-Galusi (2013) ont développé une nouvelle base de données (Manulex-Morpho) à partir des entrées lexicales les plus fréquentes de la base Manulex. La nouveauté de cette base est de prendre en compte l'apport de l'information morphologique sur la consistance des mots à partir de quatre catégories

⁴⁶ Dans ses travaux, Lété et ses collaborateurs incluent la classe de CE2 dans le cycle 3. Suite à une modification récente des cycles, le CE2 est maintenant intégré au cycle 2 et le cycle 3 comprend les classes de CM1, de CM2 (pour le primaire) et de 6^{ème} (pour le collège).

d'indices morphologiques : les flexions de genre et de nombre ; les flexions verbales ; la dérivation en « -ent » des adverbes ; les autres marques flexionnelles et dérivationnelles présentes en fin de mots. La consistance varie en effet en fonction de la place des graphèmes dans les mots et en français elle baisse surtout à la fin, à cause des lettres muettes et des marqueurs morphologiques. La prise en compte des indices morphologiques, qui fait diminuer la consistance des mots, montre pourquoi en français, écrire est plus difficile que lire : « [...] *si écrire est beaucoup plus difficile que lire en français, ceci est dû à des aspects liés à la morphologie flexionnelle des mots.* » (Lété, 2008). Nous n'avons pas utilisé cet outil dans notre étude, car nous avons trouvé dans Manulex-Infra, tous les items de la dictée d'IVQ y compris sous leur forme fléchie (FROMAGES, FIANCÉE).

La figure 2 est une capture d'écran qui montre le résultat obtenu lorsque l'on demande l'indice de consistance PO, dans tous les manuels du CP au CM2, avec un calcul par token pour l'item FROMAGES :

The screenshot shows the 'FROMAGES' software interface. At the top, there is a title bar 'Résultats de la requête sur Manulex-Infra'. Below it are several buttons: 'Retour', 'Frais-securs OFF', 'Réinitialiser le Tri', 'Export fichier' (with a sub-option 'Inclure l'en-tête'), and 'Code ASCII'. The main area displays search results for 'CP au CM2'. It shows 'SYNTHE : HC' and 'ORTHO : FROMAGES'. A table below the search results provides consistance indices for the word 'FROMAGES' across different positions: GRAM (78), ORTHO (78), COPGT (78), COPGI (91), COPGM (83), and COPGF (37). Navigation buttons like 'Précédente | 1 | Suivante' are visible.

GRAM	ORTHO	COPGT	COPGI	COPGM	COPGF
78	78	78	91	83	37

Figure 2. Capture d'écran de FROMAGES, Consistance OP (Manulex-Infra).

L'indice de consistance moyen de FROMAGES (COPGT pour Total) est de 78. Le mot, découpé en syllabes graphiques, présente un indice de consistance sur le début du mot (COPGI) de 91 points, sur le milieu du mot (COPGM) de 83 points et sur la fin du mot (COPGF) de 37 points. Le degré de complexité orthographique augmente sur la finale du mot, probablement à cause de la marque de flexion.

1.2 LA FRÉQUENCE ORTHOGRAPHIQUE : UNE VARIABLE PRIMORDIALE DE LA COMPÉTENCE ORTHOGRAPHIQUE

Pour déterminer la complexité orthographique des mots de la dictée, nous utilisons en plus de leur indice de consistance, un indicateur de leur fréquence. Pour Lété (2004 : 241) les valeurs des fréquences « renvoient en quelque sorte à la qualité – ou à la force – des représentations cognitives des mots en mémoire lexicale ». De nombreuses études ont montré que les performances orthographiques étaient meilleures pour les mots très fréquents que

pour les mots peu fréquents et que la fréquence joue même un rôle fondamental dans la plupart des tâches psycholinguistiques (Monsell, 1991, cité dans Lété, 2004).

Le calcul de la fréquence des mots est disponible dans plusieurs outils. Nous présentons le calcul qui est effectué dans la base de données Manulex-Infra puis nous présentons un autre outil que nous utilisons pour analyser les mots de la dictée d'IVQ : la base de données Lexique.

1.2.1 LE CALCUL DE L'INDICE DE FRÉQUENCE DANS MANULEX-INFRA

Dans la base Manulex-Infra, deux calculs sont proposés pour estimer la fréquence des mots : l'indice de fréquence brut et l'indice de fréquence estimée d'usage pour 1 million.

La fréquence brute (F) est le nombre d'occurrences d'un mot dans le niveau considéré. Mais cet indice ne reflète pas la rareté des mots. En effet, si un mot est très présent dans un manuel (ex. POINT qui apparaît 242 fois dans un seul manuel), il aura un indice de fréquence aussi élevé qu'un mot qui est moins présent dans chaque manuel mais qui apparaît dans tous (ex. PAPA qui apparaît 21 fois dans chaque manuel). Dans cet exemple, l'indice de fréquence brut ne permet pas de rendre compte de la rareté du mot POINT. La fréquence estimée d'usage (U) en revanche, est un dérivé de F qui a été ajusté avec un indice de dispersion (D), ce qui permet de disposer d'un indice qui prenne en compte la dispersion des mots dans les manuels ainsi que leur fréquence réelle afin de disposer d'un indicateur plus fiable de la fréquence des mots. La valeur de U est en effet celle que l'on pourrait trouver dans un corpus hypothétique, ce qui permet ainsi de l'utiliser pour faire des comparaisons avec des valeurs de fréquence issues de corpus de taille différente. Dans l'exemple cité ci-dessus, POINT a une valeur U = 326 et PAPA a une valeur U = 532, ce qui permet de conclure que PAPA est presque deux fois plus fréquent que POINT.

La figure 3 est une capture d'écran qui montre le résultat obtenu lorsque l'on demande les indices de consistance PO et l'indice U dans tous les manuels du CP au CM2, avec un calcul par token pour POINT (nom commun) :

Résultats de la requête sur Manulex-Infra

Retour Plein-écran OFF Réinitialiser le Tri Export fichier Inclure l'an tête Code ASCII

CP au CM2
SYNTH 18 HC
ORTHO 18 point

Résultats 1 – 1 sur 1
Nombre de Résultats -1- Changer

Précédente | 1 | Suivante

GRAM	ORTHO	U	COPGT	COPGI	COPGM	COPGF
sc	point	326,34	75,43	100,00	89,83	26,66

Précédente | 1 | Suivante

Figure 3. Capture d'écran de POINT, Consistance OP (Manulex-Infra)

La figure 4 est une capture d'écran qui montre le résultat obtenu lorsque l'on demande les indices de consistance PO et l'indice U dans tous les manuels du CP au CM2, avec un calcul par token pour PHARMACIE :

Résultats de la requête sur Manulex-Infra

Retour Plein-écran OFF Réinitialiser le Tri Export fichier Inclure l'an tête Code ASCII

CP au CM2
ORTHO 18 pharmacie

Résultats 1 – 1 sur 1
Nombre de Résultats -1- Changer

Précédente | 1 | Suivante

GRAM	ORTHO	U	COPGT	COPGI	COPGM	COPGF
sc	pharmacie	26,84	87,17	8,08	83,28	23,68

Précédente | 1 | Suivante

Figure 4. Capture d'écran de PHARMACIE, Consistance OP (Manulex-Infra)

Nous utiliserons l'indice U pour prendre en compte la fréquence des mots de la dictée d'IVQ dans les résultats des productions des scripteurs adultes.

Pour appuyer nos résultats et disposer d'indices de consistance et de fréquence des mots de la dictée d'IVQ, nous utilisons l'outil Manulex-Infra, bien que ce dernier ne reflète pas complètement l'exposition à l'écrit des adultes au quotidien puisque l'indice de fréquence est calculé en fonction de la présence⁴⁷ des mots dans les manuels scolaires. Pour compléter les propositions de Manulex, nous utilisons la base de données Lexique 3.

1.2.2 LA BASE DE DONNÉES LEXIQUE 3 : BRÈVE PRÉSENTATION

La création de la base de données Lexique provient d'un besoin identifié par la recherche en psycholinguistique. Ce besoin peut être brièvement résumé de la manière suivante : une base de données accessible librement aux chercheurs, dont le corpus corresponde à l'écrit qui est proposé aux adultes (à la différence de Manulex qui est basé sur l'exposition des enfants aux

⁴⁷ Il n'existe pas à notre connaissance de base de données qui fournisse des indices de fréquence (ou de familiarité) issus de corpus d'écrits ordinaires adressés aux adultes (revues, affichages présents dans la rue, dans les magasins, etc.).

manuels scolaires), qui comprennent les formes fléchies des mots et qui proposent des estimations de fréquences plus actuelles que celles des bases qui sont mises à disposition. Ainsi, New, Pallier et ses collaborateurs (2001) ont créé la base de données Lexique (actuellement dans sa troisième version).

Lexique propose de calculer la fréquence des mots d'après deux corpus. Le premier corpus est composé de textes issus de Frantext mais dont la limite est qu'il présente un style soutenu (textes littéraires), qui ne reflète pas nécessairement l'usage ordinaire de la langue française. Le deuxième corpus est issu de l'oral et il a pour objectif de combler les limites du premier corpus en utilisant des données qui reflètent davantage l'usage commun de la langue. Ce second corpus est composé de sous-titres de films et de séries. La base de données Lexique permet ainsi de disposer d'informations plus actuelles et plus complètes que les outils proposés comme Brulex⁴⁸ (Content, Mousty et Radeau, 1990) qui ne propose pas de formes fléchies et dont le corpus n'est pas récent (textes littéraires qui datent de 1919 à 1964).

La base de données Lexique propose donc un calcul de la fréquence (U) de mots sélectionnés d'après deux corpus :

- un corpus des 218 textes littéraires publiés entre 1950 et 2000⁴⁹, qui sont présents dans la base Frantext. Ces textes sont essentiellement des romans, mais la base comprend également des recueils de poésie, des essais. Ce corpus permet de calculer la fréquence de mots basés sur la langue écrite ;
- un corpus de sous-titres d'environ 10 000 films ou séries, qui permettent de calculer la fréquence de mots issus de la langue orale.

La base propose également un calcul de fréquence sur les formes fléchies des mots (désinences verbales, flexions de genre et de nombre sur les noms et les adjectifs). Deux estimations de fréquence sont ainsi fournies : une sur le lemme et une sur la forme orthographique.

Les résultats de Lété et al. (2008) montrent qu'une exposition répétée aux mots apprend implicitement aux enfants que certaines graphies sont plus fréquentes que d'autres dans certains contextes, ce qui impacte aussi leur acquisition du lexique. Cet effet de fréquence

⁴⁸ BRULEX est une base de données informatisée qui regroupe 35 746 entrées lexicales du *Petit Robert* et leurs fréquences à partir du TLF. Les fréquences ont été estimées à partir d'un corpus de textes littéraires datant de 1919 à 1964 (26 millions de mots).

⁴⁹ Cette sélection correspond à la partie la plus récente de Frantext.

dans les compétences orthographiques des enfants est observé de manière très précoce, dès le CP, ce qui confirme que le lexique orthographique se constitue très rapidement avec l'exposition à l'écrit. Nous postulons que cet effet de fréquence continue de jouer un rôle dans les compétences orthographiques des adultes. Nous manquons de données concernant les mots auxquels les adultes sont confrontés dans le quotidien, aussi, pour calculer les indices de fréquence des mots de la dictée d'IVQ, nous utiliserons les données issues de la base Lexique, davantage appropriée aux adultes que la base Manulex.

1.3 L'ACQUISITION DU LEXIQUE

Dans cette partie, nous présentons les procédures qui permettent aux enfants et aux adultes d'orthographier les mots, d'après les récentes recherches effectuées dans le domaine de la psycholinguistique, en particulier ceux de Fayol et Jaffré (2014), de Lété et al (2008) et Lété (2018). Nous présentons en particulier les travaux qui concernent l'apprentissage statistique du lexique, en lien avec la notion de fréquence orthographique.

1.3.1 LES PROCÉDURES MISES EN ŒUVRE DANS L'APPRENTISSAGE DU LEXIQUE

Pour orthographier les mots, les recherches en neuropsychologie depuis les années 1970 ont avancé deux types de procédures (ou deux voies), dont l'existence a plus récemment été confortée par l'imagerie cérébrale (Dehaene, 2007).

Il existerait alors deux types de processus d'acquisition de l'écrit :

- la procédure infra-lexicale qui traite les mots par transcodage des relations entre phonèmes et graphèmes et qui est mesurée par l'indice de consistance = **effet de consistance** ;
- la procédure lexicale qui permet de restituer les mots fréquents par récupération directe en mémoire = **effet de fréquence**.

Nous détaillons ci-dessous ces deux procédures.

1.3.1.1 LA PROCÉDURE INFRA-LEXICALE

Pour orthographier un mot inconnu, le scripteur utilise une procédure qui lui permet de traiter les mots par transcodage des relations entre phonèmes et graphèmes. Chez Besse et al. (2009), cette procédure est appelée **voie phonographique**, dans le sens où elle consiste à associer une ou plusieurs lettres pour transcrire un phonème. Concrètement, le scripteur stocke temporairement en mémoire la forme orale entendue le temps de décomposer le mot, puis il procède à un traitement phonographique : c'est-à-dire qu'il va faire correspondre des

graphèmes aux phonèmes entendus, tout en gardant en mémoire la forme d'ensemble du mot et l'état d'avancement de la transcription qu'il est en train d'effectuer. Cette procédure est particulièrement couteuse d'un point de vue cognitif, d'autant plus si le mot est long.

Pour décrire cette procédure qui consiste littéralement à construire une forme graphique, Fayol et Jaffré (2008), utilisent plutôt les termes de **voie d'assemblage** ou de **voie indirecte**, que nous retrouvons aussi chez Lété et al. (2008). Si le principe reste le même, il nous semble néanmoins que le terme « voie phonographique » ne comporte pas la dimension de fréquence qui est associée à la voie d'assemblage ou voie indirecte décrite par Fayol et Jaffré (2008) ainsi que par Lété et al. (2008). Selon ces auteurs, les scripteurs avec l'expertise, vont davantage procéder à une transcription de segments phoniques plus large en formes orthographiques lorsque celles-ci sont fréquentes : ils vont par exemple repérer que le segment phonique /mã/ s'écrit plus fréquemment « -ment » lorsqu'il se trouve en finale des mots.

Selon Lété et al. (2008), la probabilité de réussir le mot selon cette procédure est mesurée par l'indice de consistance : c'est-à-dire le degré de probabilité pour qu'un graphème (ou une association de graphèmes) soit utilisé dans une configuration en particulier (par exemple en finale des mots). Selon Lété et al. (2008), les enfants (de manière relativement précoce) et les adultes utilisent les associations les plus fréquentes lorsqu'ils écrivent un mot dont ils ne connaissent pas la graphie exacte. Pour cela, ils récupèrent en mémoire des instances qu'ils auront identifiées comme fréquentes dans certaines configurations.

L'utilisation de cette procédure entraîne bien souvent des erreurs qui sont phonologiquement plausibles et qui sont liées à la bi-univocité du système graphique du français, qui comporte plus de graphèmes que de phonèmes en français : par exemple écrire *taumate*⁵⁰ pour TOMATE.

1.3.1.2 LA PROCÉDURE LEXICALE

La procédure lexicale est utilisée lorsque les scripteurs sont confrontés à des mots qu'ils connaissent parce qu'ils les ont fréquemment rencontrés et qu'ils les ont enregistrés dans leur lexique orthographique. Cette procédure est appelée **procédure lexicale**, **récupération directe** ou encore **voie d'adressage** selon les auteurs. Elle consiste à récupérer directement en mémoire la forme complète du mot pour la restituer. Il s'agit d'une procédure efficace pour

⁵⁰ Cet exemple est issu du corpus de dictée de l'enquête IVQ.

le scripteur, qui sollicite une attention minimale et qui s'effectue sans erreur si le mot a été enregistré correctement.

Selon Fayol (2013), les adultes disposeraient ainsi d'un stock lexical important enregistré en mémoire, comme un dictionnaire mental. Les recherches dans ce domaine tendent vers un lexique commun à la lecture et à l'écriture et son existence est indispensable pour orthographier les mots qui utilisent des correspondances phonographiques peu fréquentes et qui sont peu consistants comme CHAOS, PAYS, etc.

1.3.2 L'IMPACT DE LA CONSISTANCE SUR LES COMPÉTENCES EN ORTHOGRAPHE LEXICALE

Dans notre étude sur les difficultés orthographiques d'adultes faiblement littéraciés, nous mettons en lien l'indice de consistance des mots proposés dans la dictée d'IVQ avec les connaissances orthographiques des enquêtés. Il existe peu d'études qui se sont intéressées aux connaissances lexicales des adultes et la plupart des études se sont axées sur l'état des connaissances lexicales chez des profils particuliers, comme les dyslexiques ou les dysgraphiques (par exemple les travaux de Holmes, cités par Fayol et Jaffré, 2014). En revanche ; plusieurs études ont été menées chez les enfants (Fayol 2013 ; Fayol et Jaffré, 2014), dont les résultats ont été confortés par les travaux récents de Lété (2008, 2018). Ces études montrent que pour orthographier les phonèmes qui sont peu consistants, les enfants prennent en compte de manière progressive, au fur et à mesure de leurs apprentissages et de leur expérience avec l'écrit, le contexte environnant ces phonèmes.

Certains chercheurs (Fritz, 1985 cité dans Lété et al., 2008) postulent l'existence de stades dans l'acquisition, qui apparaîtraient successivement (les enfants utiliseraient une procédure puis une autre) mais les recherches plus récentes de Lété et al. (2008) montrent une conception plus interactive des processus cognitifs utilisés. Les chercheurs ont étudié l'évolution des connaissances orthographiques des enfants du CP au CM2 à partir des mesures de consistances de la base Manulex-Infra et des résultats présentés dans la base de données Éole⁵¹ (Pothier et Pothier, 2003). Leurs résultats confirment l'hypothèse selon laquelle les enfants utilisent les deux procédures (lexicale et infra-lexicale) dans chaque niveau d'apprentissage. Ce serait la longueur du mot qui déterminerait l'utilisation d'une procédure plutôt qu'une autre. Ainsi, au début de leur apprentissage de l'écrit, les enfants semblent

⁵¹ La base de données Éole présente pour près de 12000 mots, le pourcentage de graphies correctes qui ont été produites par des enfants du CP au CM2. Chaque mot présent dans Éole a été orthographié par 40 enfants dans chaque niveau.

utiliser préférentiellement la procédure lexicale⁵² pour les mots monosyllabiques, et la procédure infra-lexicale⁵³ pour les autres mots qu'ils n'ont pas encore enregistrés en mémoire. Ces procédures seraient mobilisées de plus en plus rapidement et de manière plus efficace avec le niveau d'expertise. Ces travaux confirment également que l'effet de fréquence est présent dès le CP : « *the present work showed that children also make use of orthographic information stored in lexical memory because a word-frequency effect on spelling-accuracy scores was found as early as first grade* » (Lété et al., 2008 : 970). Ils montrent également que leur lexique orthographique est d'abord constitué de mots courts : « *Par exposition répétée aux mots, les enfants apprennent graduellement que certaines graphies sont plus fréquentes que d'autres dans certains contextes.* » (Lété, 2008 : 95).

Lété (2008) montre que la complexité de l'orthographe du français est adressée aux enfants dès leurs premières années d'apprentissage : « *les 290 associations trouvées dans les 45080 formes orthographiques de Manulex-Infra (niveau CP-CM2) sont toutes présentes au CP, dans seulement 10861 formes orthographiques.* » (Lété, 2008 : 91). Dès le CP, les enfants sont donc exposés à toutes les inconsistances de leur langue, ce qui représente une situation idéale pour l'acquisition de l'écrit. Cependant, même s'ils sont exposés à ces inconsistances de manière précoce, la consistance des relations entre phonèmes et graphèmes n'en reste pas moins la principale difficulté dans l'acquisition de l'orthographe chez les enfants. Et cette difficulté liée à la complexité orthographique et à l'effet de consistance ne disparaîtrait pas avec le niveau d'expertise : « *Le fait que les associations phonème-graphème continuent à jouer un rôle important chez les enfants plus âgés est un résultat attendu puisque l'information contextuelle ne permet pas de réduire toutes les ambiguïtés orthographiques.* » (Lété, 2008 : 95).

Par ailleurs, on trouve souvent dans la littérature, l'idée selon laquelle la consistance orthographique (c'est-à-dire le degré de difficulté d'un mot) serait liée à sa fréquence. Ainsi, plus un mot serait rencontré fréquemment et plus son degré de consistance augmenterait. L'outil Manulex montre que cette corrélation est incorrecte et que la fréquence et la consistance sont deux mesures différentes qui n'agissent pas l'une sur l'autre. Elles peuvent montrer en revanche l'utilisation par le scripteur de deux procédures différentes : la récupération directe en mémoire (voie directe) avec la mobilisation du lexique orthographique ou le transcodage des relations entre phonèmes et graphèmes (voie indirecte). Ainsi, l'utilisation de la voie indirecte, par transcodage, permettra à un scripteur

⁵² Appelée aussi voie directe (Fayol et Jaffré, 2008).

⁵³ Appelée aussi voie indirecte (Fayol et Jaffré, 2008).

d'orthographier un mot rare mais consistant ou de fournir une graphie plausible d'un pseudo-mot (c'est-à-dire, dont la graphie restitue la forme la forme orale du pseudo-mot tout en tenant compte à l'écrit des contraintes phonographiques du français).

1.3.3 L'ACQUISITION DE L'ORTHOGRAPHE LEXICALE CHEZ LES ENFANTS

De récents travaux en linguistique (Pothier et Pothier, 2014) et en psycholinguistique (Lété, 2018) ont estimé les connaissances en orthographe lexicale des enfants jusqu'au cycle 3. Nous présentons dans ce paragraphe l'Échelle d'acquisition en orthographe lexicale (EOLE) conçue par Pothier et Pothier (2014), qui rend compte du niveau d'acquisition de l'orthographe lexicale des enfants à l'école primaire. Nous présentons ensuite les travaux récents de Lété (2018), qui se sont appuyés sur les résultats présentés dans ÉOLE et sur les mesures de fréquence de Manulex pour estimer le vocabulaire dont disposent les enfants.

1.3.3.1 L'ÉCHELLE D'ACQUISITION EN ORTHOGRAPHE LEXICALE : L'OUTIL EOLE

ÉOLE est un outil qui donne des repères sur le niveau d'acquisition de l'orthographe lexicale d'élèves du CP au CM2, à qui ont été dictés des mots en contexte (ex. déterminant + nom + adjectif). Pour construire la dictée, 11694 termes ont été sélectionnés comme étant les plus courants dans des corpus représentatifs du français écrit des années 2000. 40 élèves dans chaque niveau de classe ont écrit 50 mots, ce qui fournit environ 200 graphies pour chacun des mots testés. À chaque terme est associé son pourcentage de réussite orthographique, ce qui permet d'identifier les formes orthographiques qui sont les plus précocement acquises et celles qui posent problème. L'enjeu de cet outil est d'améliorer l'enseignement de l'orthographe et l'évaluation des élèves en donnant des repères fiables sur ce que l'on peut attendre des élèves en fonction de leur niveau scolaire.

Cet outil s'inscrit dans la lignée de l'échelle Dubois-Buyse (1^{ère} édition 1940 ; réédition de Ters et al., 1995) et du Vocabulaire Orthographique de Base (Ters et al., 1977). Il améliore ces premiers outils par la nature du corpus de référence et les tests menés auprès des enfants, mais il reste limité pour une analyse plus fine des réussites et des difficultés orthographiques. En effet, il n'intègre pas les mots-outils et il ne tient pas compte de manière systématique des classes grammaticales et des formes fléchies. Ainsi, à la manière des entrées d'un dictionnaire, les mots apparaissent sous leur forme lemmatisée. Le parti-pris étant l'orthographe lexicale, la dictée ne proposait vraisemblablement pas de formes fléchies : les formes verbales ont été données à l'infinitif, les noms et adjectifs au singulier. En revanche, pour quelques mots, une distinction est faite sur le genre, dans les cas où les mots peuvent être des adjectifs et des noms : par exemple FINAL – nom ou adjectif, masculin singulier et FINALE – nom ou adjectif,

féminin singulier. Néanmoins, il faut noter qu'il manque une désambiguïsation du lemme : les scores de réussite ne précisent pas si l'item a été dicté comme un nom ou comme un adjectif.

En outre, un outil d'aide à l'évaluation comme EOLE, se devrait aujourd'hui de compléter les données par une analyse de la consistance orthographique des formes concernées et de la fréquence d'exposition comme c'est le cas dans Manulex (Lété, Sprenger-Charolles, Colé, 2004). Manulex, qui permet de prédire la réussite ou la difficulté ne fournit quant à lui de données sur les productions effectives des élèves. Le travail de Lété (2018) a justement permis d'utiliser les données d'EOLE en les associant avec celles de Manulex, pour les entrées lexicales communes, qui fournit des résultats intéressants. Néanmoins, l'absence de scores de réussite sur les formes fléchies limite les analyses proposées.

1.3.3.2 LE VOCABULAIRE DES ENFANTS : UNE ESTIMATION DES COMPÉTENCES LEXICALES

Pour Lété (2018), le vocabulaire d'un enfant peut être évalué à partir de la quantité de mots à laquelle il est exposé lors de ses lectures. Lété (2018 : 72) distingue le vocabulaire du lexique et le définit comme : « *la mise en pratique du lexique d'une langue par un individu, par la lecture et la production écrite* ». À partir du lexique disponible dans Manulex (6 800 lemmes⁵⁴ au CP, 10 500 au CE1 et 22 500 au cycle 3), il a quantifié le vocabulaire dont les enfants disposeraient en étant exposés à un seul manuel de lecture pour chaque année d'apprentissage. Un manuel de CP contient environ 1 900 lemmes, un manuel de CE1 environ 1000 lemmes supplémentaires et entre le CE2 et le CM2, 2000 de plus. Lété (2018) estime ainsi qu'en fin de cycle 3, un enfant dispose d'un vocabulaire de l'ordre de 5000 lemmes, en réception de l'écrit, qui pour la plupart d'entre eux sont également utilisés à l'oral et à l'écrit. Ce résultat calculé d'après la fréquence des mots dans Manulex va dans le sens des résultats obtenus par Pothier et Pothier (2003) concernant l'acquisition de l'orthographe lexicale chez les enfants. Pothier et Pothier (2003) considèrent qu'un mot est acquis par une classe d'âge lorsqu'il a été réussi par 75% des enfants interrogés dans leur enquête. Selon leurs résultats, les enfants disposeraient d'un stock lexical orthographique de 300 mots au CP, 1 000 mots au CE1, 2 000 mots au CE2, 3 500 au CM1 et 5 000 au CM2. Ainsi, selon Pothier et Pothier (2003) et Lété (2018), les enfants en fin de cycle 3, disposeraient en moyenne d'un stock lexical orthographique constitué de 5000 mots, qu'ils mobilisent en réception de l'écrit mais aussi en production écrite. Ce nombre de mots disponible en fin de cycle 3, soit vers l'âge de 10 ans

⁵⁴ Selon Lété (2018), un lemme (qui est la forme graphique de base d'un mot, sans marque de flexion) représente en moyenne 1,3 forme orthographique dans Manulex.

serait néanmoins 15 fois moins important que le nombre de mots que plus tard, un adulte est susceptible de rencontrer et d'acquérir dans des écrits qui lui sont destinés.

Lété (2018) décrit également les caractéristiques du vocabulaire acquis chez les enfants. Selon ses données, environ 3000 mots nouveaux apparaissent entre le CE1 et le CM2. Ces mots constituent les $\frac{3}{4}$ du stock lexical d'un enfant de cycle 3 et ils se distinguent des mots acquis en CP par leur fréquence lexicale et leur longueur. Ainsi, les mots qui sont acquis tardivement sont plus rares et un peu plus longs que ceux acquis en CP. Pour Lété (2018), le lexique acquis aux débuts des apprentissages de manière précoce constitue le « lexique de base », constitué de mots fréquents et essentiellement bi-syllabiques. Ce lexique est très fréquent parce qu'on le retrouve dans tous les manuels jusqu'en cycle 3 et parce qu'il est également présent dans la plupart des écrits qui sont destinés aux adultes. En revanche, Lété (2018) a observé que les mots acquis plus tardivement (et qui sont moins fréquents), ne sont pas plus complexes (moins consistants) que ceux acquis en CP ou en CE1.

1.4 BILAN SUR L'APPRENTISSAGE STATISTIQUE DE L'ÉCRIT

Les travaux présentés dans cette partie, qui sont issus de la recherche en psycholinguistique, portent sur un apprentissage statistique de l'écrit. Les bases de données disponibles fournissent de précieuses informations sur la consistance orthographique (Manulex), la fréquence d'exposition (Manulex, Lexique) et sur les productions effectives des élèves en fonction de leur niveau scolaire (EOLE). Dans ce domaine, il manque néanmoins d'outils permettant de poursuivre les analyses. Il serait par exemple utile de disposer de données sur la fréquence de morphèmes (ex. préfixe ANTI-) et sur les finales muettes (ex. : les mots qui se terminent par /ar/ se terminent-ils plus fréquemment par « -ar », « -are », « -arre », « -art », « -ard » ou « -ars » ?). Cela permettrait de mieux comprendre ce qui est acquis aisément par les élèves et ce qui fait obstacle, par rapport au système orthographique du français. Concernant la fréquence des finales dans le lexique français, nous avons utilisé le *Petit Robert électronique* : sa fonction « recherche » fournit des listes de mots issues du dictionnaire (par exemple tous les mots qui se terminent par /ar/). Cet outil fournit un indicateur de la fréquence des finales (indiqué en nombre de mots). Il effectue pour cela une recherche sur l'intégralité du *Petit Robert*, dont le corpus ne correspond pas non plus à l'écrit ordinaire adressé régulièrement à un adulte. Concernant la fréquence d'exposition à l'écrit des adultes, il manque d'outils représentatifs d'un corpus d'écrits ordinaires adressé à des adultes. Car même si dans la base Lexique le corpus de sous-titres est censé combler le manque de données concernant le langage courant, il reste axé sur un corpus oral qui ne permet donc pas à l'adulte, même s'il entend fréquemment un mot, de mémoriser une forme orthographique.

Concernant les enfants, les travaux réalisés à partir de ces outils sont nombreux et apportent une connaissance de plus en plus fine du type d'acquisition dans le domaine de l'écrit, des variables qui sont en jeu et de l'organisation de ces acquisitions selon le niveau d'apprentissage. Les récents travaux de Lété (2018) montrent que lorsqu'ils orthographient, les enfants réinvestissent les connaissances que l'exposition à l'écrit (par l'intermédiaire de leurs lectures) leur a permis d'acquérir. Ainsi, les enfants exploitent, en production, les régularités statistiques de leur langue qui sont traitées lorsqu'ils lisent : « *Cette sensibilité de l'enfant aux propriétés statistiques de son environnement langagier pourrait dès lors être invoquée comme mécanisme essentiel à la construction de sa base de connaissances linguistiques dans laquelle des unités probables sont renforcées en mémoire, les autres plus ambiguës ou plus rares étant progressivement oubliées* » (Lété, 2018 : 70). Ce traitement s'effectue progressivement, au fil de leurs apprentissages et de leur exposition à l'écrit. Il apparaît également que les enfants utilisent différentes stratégies pour orthographier, et que le choix de leur utilisation est lié à de nombreux paramètres (comme le niveau d'apprentissage, le stock lexical, la fréquence et la consistance des mots à orthographier, leur longueur) qui montre une exploitation complexe du traitement de l'écrit. D'autres travaux, comme ceux réalisés par Pacton et al. (2002), montrent que les enfants de CP ont déjà des connaissances implicites concernant les régularités orthographiques de leur langue (par exemple que certaines consonnes sont doublées mais pas d'autres, que les voyelles ne le sont pas, etc.) et cela avant même qu'un enseignement explicite ne leur soit dispensé. Lorsque leurs compétences augmentent, ils appliquent des règles qu'ils ont acquises implicitement. Les données issues des productions effectives des enfants (dans EOLE) confortent celles révélées dans d'autres travaux, comme ceux de Fayol et Jaffré (2014) qui montrent que l'acquisition du français ne s'effectue pas de manière linéaire. Les nouveaux apprentissages orthographiques introduisent des doutes et des erreurs qui n'apparaissent pas auparavant : par exemple CHARLATAN dans EOLE qui est réussi à 75% en CP, à 29% en CE1, à 17% en CE2, à 30% en CM1 puis à 23% en CM2.

Selon Lété (2018 : 71), cet apprentissage représente un « *type particulier d'apprentissage implicite, non enseigné, pour lequel on considère que l'apprenant s'ajuste aux régularités statistiques des événements linguistiques auxquels il est confronté grâce à des processus associatifs élémentaires* ». Ces « *événements linguistiques* » renvoient entre autre à la fréquence d'exposition et au degré d'incertitude des relations entre phonèmes et graphèmes en lien avec la consistance orthographique. Lété (2008) souligne que si l'effet de consistance des mots sur les performances en lecture chez les enfants et les adultes a été beaucoup étudié, il l'a peu été en ce qui concerne la production orthographique, notamment des adultes. « *Un*

effet avéré de ces variables statistiques (comme ici la consistance) permet de considérer que l'enfant (et l'adulte) exploite ces régularités dans la compréhension et la production du langage » (Lété, 2008 : 95). Aussi, nous postulons que l'effet de consistance continue de jouer un rôle dans les connaissances orthographiques de certains adultes et notamment chez ceux qui ont un stock lexical plus faible.

2. L'ORTHOGRAPHE DU FRANÇAIS DU POINT DE VUE DE LA PSYCHOLINGUISTIQUE

Pour comprendre l'origine des difficultés orthographiques des adultes, nous présentons dans ce chapitre les apports de la psycholinguistique sur les processus qui sont à l'œuvre lorsque les enfants apprennent l'orthographe, puis lorsqu'ils ont affaire à la morphologie flexionnelle et dérivationnelle.

Les études menées en psycholinguistiques montrent qu'écrire et orthographier sont des opérations coûteuses en attention et mémoire, ce qui affecte les performances orthographiques. La complexité de notre orthographe, une sensibilité interindividuelle différente (qui découle aussi de la question de la représentation de l'orthographe et du rapport à l'écrit propre à notre société et propre à chaque personne), les déterminants sociaux sont autant de facteurs qui peuvent expliquer que même à l'âge adulte, écrire un message simple n'est pas une formalité pour tous.

Dans cette partie, nous nous centrons sur quelques difficultés orthographiques du français : la morphologie flexionnelle avec la gestion de l'accord en nombre, la morphologie dérivationnelle avec le problème des finales muettes et le lexique en général, dont la complexité est mesurée par la consistance et dont la restitution par les scripteurs dépendent également d'un deuxième paramètre basé sur la fréquence d'exposition. Notre choix a été guidé par les difficultés auxquelles ont été confrontés les enquêtés dans la dictée d'IVQ (que nous présentons dans le chapitre suivant consacré à la méthodologie). Certaines acquisitions complexes, comme l'accord du participe passé ou les désinences verbales en /e/, constituent un point central des difficultés orthographiques chez les élèves et font partie des erreurs qui persistent chez les adultes (Brissaud, 1999 ; Brissaud, Chevrot, Lefrançois, 2006 ; David, Guyon, Brissaud, 2006). Ces compétences n'ont pas été testées dans la dictée et nous considérons en effet que leur réalisation correspond à un degré supérieur de maîtrise orthographique qu'il n'est pas opportun de tester lorsque l'on souhaite évaluer les compétences orthographiques utiles pour un usage littéracie de l'écriture.

2.1.LA MAITRISE DE LA MORPHOLOGIE FLEXIONNELLE, UNE COMPÉTENCE COMPLEXE

Nous avons vu dans le chapitre 2 (section 4) dans la partie consacrée aux morphogrammes grammaticaux qu'une des principales difficultés de l'orthographe du français concerne l'écriture de la morphologie flexionnelle. En effet, la spécificité de la morphologie flexionnelle du français est d'être écrite mais très souvent absente de l'oral. Ainsi dans l'exemple « ILS MANGENT » vs « IL MANGE » les deux syntagmes ne se distinguent pas à l'oral. À l'écrit, en revanche, deux informations grammaticales sont présentes dans le pronom « ILS » ainsi que dans la flexion verbale avec « -NT ». Ce phénomène est parfois qualifié de redondance à l'écrit et serait une sorte de compensation à la perte d'éléments qui ne sont présents qu'à l'oral (gestuelle, intonation, etc.). Il apparaît surtout que cette spécificité du français a deux conséquences décrites par Fayol et Jaffré (2014). Tout d'abord, le fait que les enfants (ou les apprenants étrangers) ne peuvent découvrir ces marques que lors de l'apprentissage de l'écrit. En effet, à l'oral, ils n'ont que peu été confrontés à l'information grammaticale systématiquement présente à l'écrit : IL VIENT vs ILS VIENNENT mais IL MANGE vs ILS MANGENT. Ensuite, une fois l'apprentissage effectué, les enfants et les adultes doivent gérer des marques sans pouvoir se référer (sauf dans quelques cas) à des indices qu'ils peuvent entendre. Cela a pour conséquence de mobiliser une concentration importante lorsqu'ils écrivent, qu'ils doivent combiner avec leur mémoire de travail car selon la longueur des phrases, le temps de maintien en mémoire d'informations concernant le genre et le nombre peut être important.

2.1.1. LA MISE EN PLACE DES PLACE DES COMPÉTENCES AVANT LE PASSAGE À L'ÉCRIT

Avant d'entrer dans l'écrit, les enfants sont donc confrontés à un oral qui ne leur présente que peu de marques grammaticales, notamment pour l'accord en nombre des noms et des adjectifs, ce qui les fait entrer tard dans l'acquisition de la morphologie flexionnelle et rend particulièrement difficile l'apprentissage des marques de nombre (Fayol et Jaffré, 2014). Ce domaine de l'écrit représente alors une source d'erreurs bien connue des enseignants. Le genre grammatical apparaît en revanche plus facile à acquérir car il se manifeste par plusieurs catégories d'indices. La référence au genre naturel (LE CHAT vs LA CHATTE) va représenter la première confrontation au genre chez les enfants mais elle est limitée car la plupart des êtres animés n'ont pas de distinction graphique permettant de leur attribuer un genre (LE DAUPHIN, qu'il soit féminin ou masculin, ne change pas d'orthographe et le déterminant reste au masculin).

Pour déterminer le genre d'un nom, l'indice le plus fiable est la dimension morphosyntaxique, grâce aux marques qui sont présentes sur les mots qui entourent le nom : sur le déterminant (LA POMME vs LE MELON) et sur l'adjectif lorsque la marque de genre est audible : LA POMME VERTE vs LE MUR VERT. La dernière catégorie d'indices est d'ordre infra-lexicale, en lien avec les terminaisons des noms qui sont plus fréquemment associées au masculin ou au féminin : par exemple les noms en « -on » sont plus souvent masculins (MELON, FRELON, JETON, etc.) tandis que ceux en « -ette » sont plutôt féminins (MURETTE, MIETTE, BALAYETTE, etc.). Fayol et Jaffré (2014) rapportent que dès l'âge de 4 ou 5 ans, les enfants sont en mesure de déterminer le genre (probable) d'un mot nouveau qu'on leur présente. Cette habileté précoce montre que très tôt, les enfants ont mémorisé des formes orales associées à leur déterminant et qu'ils sont parvenus avant même d'entrer dans l'apprentissage de l'orthographe, à en extraire des régularités.

Lorsqu'ils entrent à l'école primaire, les enfants disposent donc d'une connaissance des notions de genre et de nombre sur les noms ainsi sur certains verbes qui voient leur forme varier (IL BOIT vs ILS BOIVENT). Il est cependant difficile de savoir si les connaissances en morphologie des jeunes enfants proviennent d'un effet de fréquence d'exposition à des formes qui sont davantage rencontrées au féminin ou au pluriel ou s'il s'agit déjà d'une compétence leur permettant de combiner le radical des mots avec des affixes. Certaines généralisations produites par de jeunes enfants comme « JE BOIVAIS » pour « BUVAIS » semblent plutôt attester qu'ils sont capables de produire des combinaisons issues de régularités, et les recherches actuelles sur cette question penchent alors pour une mémorisation de formes fléchies issue de la fréquence⁵⁵ d'exposition (Fayol et Jaffré, 2014).

Enfin, le problème de l'accord peut entrer en conflit avec d'autres principes et en particulier la distinction graphique entre homophones lexicaux (ex. VOIE vs VOIX). Le « x » de VOIX comme le « s » de PAYS (item de la dictée proposée dans l'enquête IVQ), ne sont pas le fait d'un accord en nombre et valent aussi bien au singulier qu'au pluriel.

2.1.2. LA MISE EN PLACE DES COMPÉTENCES AVEC LE PASSAGE À L'ÉCRIT

Lorsque les enfants entrent dans l'écrit, ils sont confrontés à différentes formes orthographiques autour d'un lemme, qui présentent des variations morphographiques visibles comme CHIEN vs CHIENS vs CHIENNE et dont certaines se distinguent également à

⁵⁵ Nous renvoyons à la section 1. du chapitre 3 pour une présentation de la notion de fréquence et des outils qui permettent de calculer la fréquence d'exposition aux mots et à leurs formes fléchies.

l'oral comme CHEVAL vs CHEVAUX. Par ailleurs, certaines formes orthographiques vont être plus fréquemment rencontrées au pluriel plutôt qu'au singulier (LES CHAUSSURES plus fréquent que LA CHAUSSURE) ou inversement (LA MER plus fréquent que LES MERS). Fayol et Jaffré (2014) montrent que la fréquence d'exposition à une forme plutôt qu'à une autre (par exemple CHAUSSURES plus fréquemment rencontré au pluriel) joue un rôle important dans la restitution orthographique et que les enfants auraient tendance à reproduire la forme fréquemment rencontrée. Cousin et al. (2006) montrent notamment que cette sensibilité à la fréquence de la forme orthographique proviendrait d'une exposition précoce à ces formes avant que les enfants aient compris que celles-ci peuvent comporter des flexions qui marquent le nombre ou le féminin (notamment jusqu'au cycle 2). À ces apprentissages réalisés de manière implicite lors d'activités de déchiffrage s'ajoute à l'école, un enseignement explicite des règles d'accord du nom, de l'adjectif et du verbe.

2.1.2.1 L'ACCORD EN NOMBRE : APPLICATION DE RÈGLES ET DE PROCÉDURES

Différentes études (Pacton et al., 2002 ; Fayol et Miret, 2005) se sont intéressées aux procédures qui sont utilisées dans la réalisation des accords en genre et en nombre. Deux procédures ont été identifiées :

La première procédure consiste à appliquer une règle « générale » lorsqu'il s'agit de réaliser un accord sur un mot inconnu. Pour l'accord en nombre, cette règle consisterait à ajouter un « -s » au radical du mot. Cette règle peut s'appliquer pour former le pluriel des noms et des adjectifs et peut s'appliquer à tous les items. Cependant Fayol et Jaffré (2014) soulignent que cette procédure est coûteuse en attention⁵⁶ et qu'elle est lente puisque sa mise en œuvre n'est pas automatisée mais demande de recourir à des ressources cognitives du fait qu'elle requiert de l'attention. En ce sens, elle est donc source d'erreurs alors même que les scripteurs qui l'utilisent connaissent les règles d'accord et qu'ils savent également voir leurs erreurs et les corriger grâce à la révision orthographique (Largy, 2001).

La seconde procédure consiste à récupérer directement en mémoire des mots ou des associations de graphèmes qui forment des groupes stables. À la différence du premier, ce mécanisme n'est pas généralisable ; il ne s'applique qu'à un seul item et implique donc que ceux-ci aient été suffisamment fréquemment rencontrés pour que leur forme ait été enregistrée et stabilisée en mémoire. Ainsi certains items fréquemment enregistrés sous leur

⁵⁶ Nous présentons plus précisément les notions de coût en attention et en mémoire dans la section 2.1.2.2 de ce chapitre.

forme fléchie (ex. LES PARENTS) auront tendance à être utilisés sous cette forme dans tous les contextes. Cette procédure utilise des activités automatiques et par contraste par rapport à la première procédure, elle est rapide et peu coûteuse d'un point de vue cognitif (Fayol et Got, 1991).

Pour Fayol et Got (1991), l'utilisation d'une procédure ou d'une autre ne s'effectue pas de manière tranchée : bien souvent les deux procédures fonctionnent simultanément et peuvent apporter deux réponses différentes. Par exemple, le scripteur sait qu'un déterminant au singulier placé devant un nom indique que celui-ci sera également au singulier (LE PARENT). En revanche, l'utilisation de la procédure de récupération directe en mémoire lui fera choisir préférentiellement la graphie PARENTS (plus fréquente).

Enfin, certaines configurations sont particulièrement propices à la survenue d'erreurs, lorsque d'autres difficultés que celle d'accorder en nombre se greffent, et cela même chez des adultes experts qui connaissent parfaitement les règles d'accord. Par exemple en présence d'un sujet inversé (PENDANT DES HEURES TOURNE LE DISQUE), d'un sujet complexe parce qu'il comporte un ou plusieurs compléments (LE CHIEN DES VOISINS ABOIE) ou de la configuration pronom - pronom - verbe lorsque l'on est en présence de pronoms clitiques (IL LES VOIT) En effet, pour accorder un verbe avec son sujet, le français procède généralement par effet d'attraction : le sujet qui précède le verbe permet de l'accorder en fonction de ses caractéristiques. Dans LE DISQUE TOURNE PENDANT DES HEURES, il n'y a pas d'ambiguïté : le sujet précède le verbe qui s'accordera à la 3^e personne du singulier. La procédure qui fonctionne par effet d'attraction permet alors d'accorder le verbe efficacement et à moindre coût. En revanche dans les configurations complexes décrites plus haut, le risque est d'accorder à tort le verbe avec le nom ou le pronom qui le précède.

Concernant les adultes, Fayol et Got (1991) soulignent que cette automatisation s'observe surtout chez ceux qui ont une pratique régulière de l'écrit. Et même s'ils utilisent des procédures de correction, les erreurs liées à l'automatisation des procédures peuvent facilement passer inaperçues ou alors seront perçues et corrigées mais avec un important effort de concentration et de mémoire. Ainsi, même s'ils connaissent les règles d'accord en nombre du verbe, du nom et de l'adjectif, les adultes experts ne les utilisent pas en situation de rédaction car l'effort d'attention et de mémoire serait trop important et entraverait la tâche de rédaction qu'ils effectuent au même moment (Fayol et Jaffré, 2008).

2.1.2.2 LE COUT EN ATTENTION ET EN MÉMOIRE : EFFETS SUR LES COMPÉTENCES ORTHOGAPHIQUES

Aux difficultés liées à l'organisation structurelle du français s'ajoute une difficulté supplémentaire que les travaux en psycholinguistique appellent le cout en attention et en mémoire. Fayol (2013) et Jaffré et Fayol (2014) ont montré que les erreurs surviennent plus particulièrement lorsque l'attention est focalisée sur une autre tâche, comme la recherche d'idées (lors de la rédaction) et cela chez les enfants comme chez les adultes. Un exercice de rédaction écrite demande en effet au scripteur de mener en même temps deux activités coûteuses en terme d'attention et de mémoire (trouver des idées et les mettre en mots). Or, nous l'avons vu, la capacité d'attention et la mémoire sont nécessaires au scripteur pour orthographier et réaliser des accords ; notamment dans les configurations complexes décrites ci-dessus, qui demandent une réflexion de la part du scripteur.

Dans une expérience menée par Largy et al. (1996), on a demandé à des adultes d'orthographier la phrase « IL LES TIMBRE » qui comporte deux difficultés pouvant induire en erreur (un pronom clitique) et un homophone verbal au pluriel dont le pluriel, justement, est plus fréquemment rencontré sous sa forme nominale (DES TIMBRES). Les résultats montrent que lorsque l'attention des scripteurs est focalisée sur l'accord du verbe (ils doivent compléter la désinence verbale), ils ne commettent pas d'erreurs. En revanche lorsqu'ils devaient transcrire la phrase tout en ayant à accomplir une tâche qui sollicite leur mémoire (retenir 3 mots), les adultes confondaient fréquemment l'homophone verbal avec sa forme nominale (IL LES TIMBRES).

2.1.2.3 LE TYPE D'ERREURS EN LIEN AVEC LE NIVEAU D'APPRENTISSAGE

Lors de l'apprentissage de l'accord en nombre et en genre, les élèves doivent acquérir en parallèle deux compétences : savoir reconnaître la catégorie grammaticale des mots à accorder et connaître les marques flexionnelles pour chacune d'elles. Différentes études menées auprès d'élèves ont montré que la catégorie grammaticale joue un rôle dans l'acquisition de la compétence en morphographie.

Une enquête réalisée par Thévenin, Totereau, Jarousse et Fayol (1999) auprès d'élèves de cycle 2, montre que l'accord en nombre sur les noms est mieux réalisé que celui sur les adjectifs et enfin que celui sur les verbes, le niveau de compétence augmentant avec le niveau scolaire (47% de réussite au CP sur l'accord en nombre du nom contre 91% en CE2). Dans une récente étude, Geoffre, Totereau et Brissaud (2019) ont exploité le corpus de dictées issu de la recherche « Lire et Écrire au CP » (Goigoux, 2016) pour rendre compte de l'évolution des

performances orthographiques des élèves, du début du CP à la fin du CE1. En fin de CP, il apparaît que 75% des élèves n'utilisent aucune marque de pluriel dans le syntagme (LES LAPINS COURENT) et seuls 17% marquent un des deux accords. Ainsi, la transition vers le traitement grammatical dans l'écriture de phrases ne se fait pas avant la fin du CP, et pour une minorité d'élèves. En fin de CE1, le critère de position du déterminant (qui déclenche l'accord), est primordial, ce qui montre la mise en place d'une procédure automatisée des accords en nombre. Dans LES SALADES VERTES, l'accord en nombre de SALADES est mieux réalisé que celui de VERTES. Ce phénomène est aussi observé chez des élèves plus âgés (Geoffre, Brissaud, 2012), sur lesquels on observe également une surgénéralisation des pluriels, en lien avec l'augmentation des compétences (Totereau et al., 1997). Enfin, Geoffre, Totereau et Brissaud (2019) montrent qu'en CE1, l'accord en nombre du verbe est encore faible, d'autant plus lorsque le marqueur de pluriel n'est pas placé immédiatement à côté du verbe (LES LAPINS COURENT).

Les travaux de Pacton et al. (2002) montrent ensuite que dès le cycle 3, la plupart des enfants connaissent les règles d'accord sur le nom, l'adjectif et le verbe et savent utiliser les procédures qui permettent de réaliser les accords. En particulier si la configuration des accords à réaliser dans le syntagme n'est pas complexe (homophonie, pronoms clitiques, sujets inversés, etc.) et que la mémoire et l'attention des scripteurs n'est pas accaparée par d'autres tâches. Et pourtant, au collège, d'autres travaux comme ceux réalisés par Totereau, Brissaud, Reilhac et Bosse (2013) montrent que les accords en nombre sur le nom, l'adjectif et le verbe restent une compétence complexe qui n'est toujours pas acquise en fin de 6^e. Les résultats obtenus sur un exercice de dictée montrent que seulement un élève sur trois (dans le collège le moins favorisé) et deux élèves sur trois (dans les deux autres collèges) marquent l'accord nominal. Pour Totereau et al. (2013), ces performances ne signifient pas que la règle n'est pas connue des élèves, mais que les erreurs sont davantage liées à la mise en place de la procédure d'accord, trop coûteuse en situation de dictée. L'accord en nombre de l'adjectif apparaît comme la tâche la moins bien réussie, avec un élève sur trois seulement (dans le collège le plus favorisé) qui réalise l'accord dans le syntagme « LES YEUX VERTS ». Ainsi, au collège, l'accord en nombre du nom, du verbe et surtout de l'adjectif continue de poser problème aux élèves. Certains travaux (David, Guyon et Brissaud, 2006 ; Thévenin, Totereau, Jarousse et Fayol, 1999) suggèrent que les élèves montreraient des progrès significatifs dans la gestion des accords en nombre au terme d'un entraînement. Cela explique peut-être pourquoi, dans l'étude de Pacton et al. (2002), les résultats observés chez des élèves de début de cycle 3 semblent meilleurs que ceux d'élèves de collège observés dans d'autres études.

Les erreurs relevées sur les marques de nombre dans l'enquête de Totereau et al. (2013) révèlent qu'il s'agit principalement d'omissions de marques. Très fréquentes au cycle 2 (Pacton et al., 2002), elles tendent à diminuer au cycle 3 et sont peu fréquentes chez les adultes. Dans notre analyse de la dictée issue de l'enquête IVQ et celle de l'enquête complémentaire, nous observerons la restitution des marques de nombre sur le nom, le verbe et l'adjectif. Par ailleurs, le type d'erreurs semble évoluer en fonction du niveau de compétences et notamment lorsque les élèves manipulent facilement l'accord en nombre (en lien avec le niveau d'automatisme). En effet, de nouvelles erreurs surviennent lorsque l'accord en nombre est automatisé (à partir du cycle 3). Apparaissent alors des cas de remplacement d'une flexion verbale par une autre marque : sur les verbes (ILS JOUE-S pour JOUENT), sur les noms (LES FEUILLENT pour FEUILLES) et sur les adjectifs (LES POMMES ROUGENT pour ROUGES). Ces erreurs disparaissent généralement au cycle 3 mais elles peuvent se maintenir et apparaître chez les adultes dans certains cas, notamment lorsque les formes homophones sont en concurrence sur le plan grammatical : par exemple le nom « FOUILLE » et le verbe « FOUILLER », ce qui amène à des erreurs de type « LES FOUILLENT ARCHÉOLOGIQUES » pour « FOUILLES ». Dans cet exemple, c'est l'effet de fréquence qui va conduire à des configurations erronées ; les scripteurs ayant tendance à utiliser les formes fléchies qu'ils vont rencontrer plus fréquemment, ici le « -nt » du verbe FOUILLER. Dans ce cas, pour Largy et al. (1996), on ne peut pas considérer que le scripteur s'est trompé de catégorie grammaticale en confondant le verbe FOUILLER avec le nom FOUILLE. Dans cet exemple, le scripteur aurait activé la valence la plus fréquente, ici la valence verbale « -nt » récupérée directement en mémoire (car plus économique) sans appliquer de règle d'accord, qu'il maîtrise pourtant. Jaffré et Fayol (2014) soulignent que ces erreurs surviennent surtout chez les adultes lorsque l'attention et la mémoire sont sollicitées.

Ainsi, la réalisation d'un accord en nombre est bel et bien une tâche complexe dans laquelle de nombreux éléments entrent en considération pour expliquer la survenue d'erreurs. Nous les résumons ci-dessous :

- la complexité de la tâche à réaliser en elle-même, en lien avec la difficulté structurelle du français ;
- un contexte plus ou moins propice à la survenue d'erreurs avec par exemple la présence de pronoms clitiques, d'un sujet inversé, etc. ;
- l'utilisation de deux procédures (récupération directe en mémoire et utilisation de règles) qui peuvent entrer en concurrence lorsqu'ont été mémorisées des formes fléchies fréquentes qui ne correspondent pas au contexte linguistique de la tâche ;

- l'attention et la mémoire qui sont sollicitées à la fois pour la réalisation de l'accord, avec l'utilisation de règles et de procédures de correction qui sont coûteuses, auxquelles s'ajoutent éventuellement le coût de la situation de scripture (le fait de rédiger et de chercher des idées, le fait de se concentrer sur de nombreuses tâches dans une dictée, etc.) ;
- d'autres paramètres que Jaffré et Fayol (2014 : 100) présentent comme des différences interindividuelles, parmi lesquelles : « *Les connaissances préalables, la capacité de mémoire de travail, la disponibilité de l'attention, la fatigue, etc.* » et qui interviennent également pour expliquer l'origine des erreurs dans la gestion des accords en nombre.

Tous ces éléments permettent de comprendre en quoi la réalisation des accords en nombre en français est une tâche complexe et la maîtrise de la morphologie flexionnelle, une compétence qui peut, selon les circonstances, poser problème même aux scripteurs adultes les plus experts.

2.2.LA MORPHOLOGIE DÉRIVATIONNELLE ET L'ACQUISITION DES FINALES MUETTES

Lors de leur apprentissage du lexique, les enfants découvrent que certains mots relevant de la morphologie dérivationnelle sont construits à partir d'un radical puis de suffixes (par exemple PETIT qui devient PETITE au féminin). Selon Béguin, Saint-Laurent et Giasson (2010), les recherches effectuées dans ce domaine montrent que pour écrire, le recours à la morphologie dérivationnelle serait une stratégie orthographique intervenant tardivement chez les enfants. Cette stratégie serait la même que celle utilisée par les adultes et elle consisterait à stocker en mémoire à long terme, des informations spécifiques sur les caractéristiques orthographiques du français. Cette procédure dite « lexicale » est en lien avec la fréquence des mots : plus un mot est rencontré fréquemment par le scripteur et plus il sera fixé en mémoire donc disponible par récupération directe.

Fayol et Jaffré (2014) se sont demandé si les enfants utilisent la morphologie dérivationnelle lorsqu'ils doivent transcrire les finales muettes des mots. Ils soulignent que l'observation de leurs productions écrites ne permet pas de déterminer s'ils ont utilisé ou non la morphologie dérivationnelle pour aboutir à une transcription exacte. En effet, la transcription correcte de la finale de PETIT peut provenir de l'utilisation de la morphologie dérivationnelle (PETIT > PETITE) mais les enfants comme les adultes peuvent y avoir recours sans en avoir conscience. La restitution de la graphie peut également être issue du lexique orthographique disponible

en mémoire, auquel le scripteur a recours par récupération directe. Certaines graphies montrent en revanche une généralisation de certaines finales comme dans ABRI-7. Une expérience a été menée au Québec auprès d'enfants francophones (Sénéchal et al., 2000 et 2006) qui ont dû restituer des mots possédant une finale muette pouvant être expliquée par la morphologie dérivationnelle (ex. BAVARD) ou ceux qui n'en dépendent pas mais qui ont une bonne probabilité sur la consonne attendue (ex. FOULARD). L'analyse des résultats montre qu'au début du primaire, les mots les mieux restitués (75%) sont les mots réguliers qui ne comportent pas de finale muette, puis ceux qui dépendent de la morphologie dérivationnelle (59%) et enfin les autres (39%). Les erreurs ont majoritairement portées sur des omissions de finale (75%).

Ces résultats suggèrent que le recours à la morphologie (pour écrire les mots dits morphologiques) est quand même une stratégie efficiente (puisque ce sont les mots qui n'en dépendent pas comme FOULARD qui sont les moins bien restitués) et que les élèves savent dès la fin du primaire tirer parti de leurs connaissances en morphologie pour orthographier (Béguin et al., 2010). En revanche, lorsque les enfants ont expliqué la procédure utilisée, c'est la récupération directe en mémoire qui semble avoir été privilégiée. Plusieurs études ont analysé les raisons de cette faible mobilisation des connaissances en morphologie dérivationnelle chez les enfants (Pacton et al., 2002 ; Pacton et al., 2006 ; Béguin et al., 2010) Il ressort de ces études plusieurs explications qui sont généralisables à tous les scripteurs.

La **première explication** concerne la formation même de la morphologie dérivationnelle en français, avec des dérivés dont le lien avec le radical n'est pas toujours facilement identifiable (voir chapitre 2, section 4.1.1).

La **deuxième explication** est en lien avec l'apprentissage des suffixes qui selon Pacton et al. (2002) est incomplet, car il n'est pas étendu à l'ensemble des mots concernés par la dérivation à partir d'un radical. Certains phénomènes ne sont pas ramenés à des règles alors qu'ils le pourraient. Par exemple, en français, lorsque les animaux possèdent un diminutif, le suffixe utilisé « -eau » (RENARD > RENARDEAU) fonctionne dans tous les cas sauf pour un seul animal (CHIOT).

Les expériences menées par Pacton et al. (qui sont confirmées par d'autres travaux comme ceux de Lété, 2008, 2018 ; Fayol et Jaffré, 2014) montrent que les enfants prennent conscience très tôt (dès le CE1) des régularités statistiques de leur langue avec des configurations de graphèmes qu'ils repèrent comme plus fréquentes que d'autres dans certaines positions (par exemple « -eau » est la configuration la plus fréquente pour transcrire /o/ en position finale) même lorsque ces régularités ne leur sont pas enseignées de manière

explicites. Dans l'étude menée par Pacton et al. (2002), les enfants dès le CE1 généralisent cette régularité de la langue et l'appliquent aux pseudo-mots lorsqu'on leur demande d'en transcrire. En revanche, ils ne parviennent pas à formuler des règles (par exemple : un diminutif en /o/ placé en fin de mot prend la forme « -eau » comme dans BATEAU) : « *Les élèves pourtant sensibles aux régularités statistiques n'extraient pas d'eux-mêmes des règles pourtant simples qui leur permettraient de réguler leur productions orthographique.* » (Fayol et Jaffré, 2014 : 110).

Les travaux de Pacton et al. (2002) rendent également compte que l'utilisation des stratégies (récupération directe en mémoire ou recours à la morphologie) pour orthographier des mots qui comportent une marque de dérivation était liée au niveau d'expertise des enfants en langue écrite (lecture et écriture). Il apparaît également que les finales muettes des mots qui possèdent un dérivé féminin ont été mieux orthographiés que les autres, avec des taux de réussite qui augmentent avec le niveau d'habileté en lecture-écriture. Cependant, quel que soit le niveau d'expertise, le meilleur taux de réussite est toujours observé pour les mots qui possèdent un dérivé féminin, ce qui montre que le recours à la morphologie est une compétence qui selon les circonstances, peut intervenir très tôt.

Enfin en 2006, Pacton et Casalis ont vérifié si la capacité des enfants à utiliser une procédure morphologique pouvait influencer la graphie de mots qui ne comportent pas de lettre finale, comme par exemple NUMÉRO qui pourrait être écrit de manière erronée avec un « -t » final, à cause de son dérivé NUMÉROTÉ. L'expérience a montré que les enfants qui sont meilleurs en orthographe réussissaient mieux à restituer les finales muettes des mots qui comportent une marque de dérivation (moins d'erreurs d'omission mais aussi de substitution de lettres muettes). En revanche, chez ces mêmes enfants, on retrouve davantage d'erreurs concernant l'ajout de finales muettes sur les mots dits « pièges » qui n'en possèdent pas (NUMÉRO-T, FAVORI-S). Ces résultats laissent penser qu'ils surgénéralisent l'utilisation de la morphologie pour certains mots.

Béguin et al. (2010) a mesuré l'effet de la fréquence des mots sur la restitution des finales muettes de mots morphologiques chez des élèves considérés comme bons ou faibles en orthographe. Pour le choix des mots, la fréquence orthographique a été contrôlée avec Manulex (mots rares opposés à des mots fréquents), ainsi que sur trois catégories de mots : des mots morphologiques qui ont un dérivé féminin, des mots morphologiques qui n'ont pas de dérivé féminin mais qui sont dérivables par un mot de la même famille et des mots lexicaux. Les résultats montrent que chez tous les élèves, quel que soit le type d'item, la fréquence joue un rôle important : les finales muettes sur les mots rares sont moins bien restituées que celles

présentes sur les mots fréquents. En revanche, les élèves bons en orthographe sont beaucoup moins sensibles à l'effet de fréquence sur les mots morphologiques qui ont un dérivé féminin, ce qui confirme qu'ils utilisent (davantage que les élèves faibles en orthographe) une procédure morphologique pour les orthographier et que la mise au féminin semble être une stratégie morphologique plus accessible (car plus enseignée ?). En revanche, contrairement aux autres élèves, les faibles en orthographe réussissent mieux les mots lexicaux que les mots morphologiques, ce qui indique qu'ils utilisent peu la stratégie morphologique et qu'ils ont plutôt recours à leur lexique orthographique (et sont donc plus tributaires de la fréquence des mots).

2.3. CONCLUSION SUR LA GESTION DE LA MORPHOLOGIE

La gestion de la morphologie flexionnelle en français est une compétence complexe qui semble poser problème à tous les scripteurs, même les plus experts. Différents processus entrent en compte et nous avons vu qu'il ne s'agit pas simplement d'un problème de connaissances mais aussi d'une difficulté à mettre en œuvre les connaissances à cause de procédures parfois concurrentes. Ainsi, même lorsque les règles sont connues, c'est souvent la procédure la moins coûteuse en attention et la plus rapide qui est choisie ; et donc la récupération directe d'instances. En somme, si l'on considère que dans la grande majorité des cas les règles sont connues, il faudrait que les scripteurs soient régulièrement entraînés à réaliser des accords et qu'ils mettent en place des procédures de contrôle, notamment lorsqu'ils se trouvent dans une situation davantage propice à la survenue d'erreurs. Il est rare en effet, dans une tâche d'écriture, que la réalisation des accords soit la seule difficulté dont le scripteur ait à se soucier : il est généralement en train d'effectuer une autre tâche coûteuse en mémoire et en attention comme rédiger, réfléchir, etc. qui affectent sa mémoire temporaire de travail et son attention au moment où la difficulté apparaît.

Par ailleurs, l'apprentissage de la morphologie ne s'effectue pas de manière linéaire puisqu'au cours de l'apprentissage, l'automatisation des procédures fait de nouveau apparaître des erreurs, là où la récupération directe en mémoire d'instances fréquentes entre en concurrence avec l'application connue de règles. Ainsi, les erreurs en morphologie ne sont pas nécessairement liées à une méconnaissance des règles ou à leur mauvaise utilisation : la réalisation des accords est en effet extrêmement coûteuse à la fois en mémoire mais aussi en attention. Ainsi, même les scripteurs experts qui possèdent des mécanismes de contrôle efficaces peuvent se trouver en difficulté dans certaines configurations particulièrement propices à la survenue d'erreurs. Il serait alors bien audacieux pour un scripteur d'affirmer qu'en toute circonstance, sa gestion des accords se réalise sans obstacle. Selon Fayol et Jaffré

(2014), chez les adultes, la gestion des accords se ferait justement sans réelle « gestion » notamment chez les scripteurs experts qui procèdent par automatisation et qui récupèrent directement en mémoire des instances fréquentes car leur mobilisation est rapide et souvent efficace. Ce sont alors les procédures de contrôle, efficaces chez le scripteur expert, qui permettraient de rectifier les erreurs dans la plupart des cas. Mais chez les adultes dont le niveau d'expertise est plus faible, les erreurs surviendront alors dans des contextes particuliers (sujet inversé, présence de pronoms clitiques, homophones, etc.).

Pour les lettres finales muettes, il existe de nombreuses petites « règles », généralement apprises à l'école qui comportent des exceptions (par exemple les noms féminins en /i/ se terminent en « ie » sauf BREBIS, SOURIS, FOURMI PERDRIX et NUIT). Mais leur apprentissage ne garantit ni que les scripteurs les connaissent toutes, ni qu'ils les appliquent en situation de scripture lorsqu'ils les connaissent (difficulté du transfert entre la connaissance d'une règle et son application, notamment en terme de coût cognitif). Selon Fayol et Jaffré (2014), il semblerait même que les adultes régulièrement au contact de l'écrit ne les utilisent pas, préférant récupérer directement en mémoire les formes orthographiques qu'ils ont enregistrées (que celles-ci soient correctes ou non). Concernant l'apprentissage du vocabulaire, la littérature rapporte que l'apport de la lecture sur l'apprentissage implicite des régularités de la langue écrite serait modeste en production. Cette conception « naturelle » de l'apprentissage de l'orthographe serait en effet gênée par les irrégularités du système orthographique du français, avec des régularités qui ne sont jamais étendues à tous les mots. Comme le suggèrent les travaux de Lété (2008, 2018), l'apport de la lecture se ferait davantage par l'enregistrement du vocabulaire dans le lexique orthographique des enfants, au contact régulier de l'écrit et de la fréquence de rencontre des mots.

Comme le soulignent Fayol et Jaffré (2014), si l'apprentissage de règles et de procédures permettant de les appliquer devait réussir, la morphologie (notamment flexionnelle) ne devrait pas poser tant de difficultés. En France, l'instruction passe beaucoup par l'apprentissage de règles et d'exercices d'application permettant justement d'automatiser des procédures : « *En d'autres termes, alors que l'enseignement pense privilégier la mémorisation des règles et leur application par calcul, il favorise aussi (surtout ?) la mémorisation d'instances.* » (Fayol et Jaffré, 2014 : 121). L'efficacité de l'enseignement explicite de la morphologie flexionnelle semble en effet relative, à en croire les résultats des dernières évaluations disponibles. Dans l'enquête « Lire, écrire, compter » conduite en 1987, 2007 et

2015 par la DEPP⁵⁷, un échantillon d'élèves de fin de CM2 a été évalué en lecture, en calcul et en orthographe. Ils ont passé une dictée destinée à vérifier leur maîtrise des règles grammaticales en situation de production écrite et leurs connaissances lexicales. Le résultat de cette enquête montre qu'en vingt ans, les compétences orthographiques des élèves d'une même classe d'âge ont diminué, avec un peu plus de 50% des erreurs liées à la morphologie flexionnelle (Andreu, Steinmetz, 2016).

⁵⁷ Direction de l'évaluation, de la prospective et de la performance.

OBJECTIFS DE LA RECHERCHE

PROBLÉMATIQUE

L'apprentissage de l'écrit, comme le rapporte Fayol (2013), est systématique depuis environ 150 ans en France. Ainsi l'analphabétisme a quasiment disparu, mais l'illettrisme subsiste encore comme le montre la dernière enquête de l'INSEE (IVQ 2011). Comme cette appellation d'illettrisme implique une évaluation et que la frontière entre ce qui relève de l'illettrisme et ce qui n'en relève pas ne va pas de soi, nous ne reprendrons pas à notre compte cette étiquette mais parlerons de personnes en difficulté avec l'écrit ou de personnes faiblement littéraciées.

Dans notre étude, nous nous intéressons aux compétences orthographiques de scripteurs âgés de 16 à 65 ans et résidant en France métropolitaine, qui ont été considérés en difficulté dans le domaine de l'écrit dans l'enquête IVQ (INSEE, 2011). Le matériau de notre recherche se compose de dictées qui reflètent une situation de scription ordinaire (rédiger une liste de courses). Elles sont issues d'une première enquête d'IVQ et d'une enquête complémentaire conduite en 2014 et combinent différents items, ce qui permet de tester les compétences orthographiques de plusieurs manières (mots et pseudo-mots, mots fréquents et rares, pseudo-mots à priori simples ou à priori compliqués, syntagmes induisant des marques du pluriel). Le parti-pris dans cette thèse est d'envisager les compétences orthographiques de ce public d'adultes en difficulté avec l'écrit dans le cadre de la littéracie : leurs dictées sont analysées non seulement en termes de lacunes mais également en termes de savoirs et savoir-faire. En outre, nous défendons l'idée que la connaissance du système orthographique du français peut être mise à profit pour mieux cerner les compétences des scripteurs.

Notre problématique est la suivante : comment peut-on caractériser linguistiquement les compétences orthographiques d'adultes faiblement littéraciés pour établir des types de compétences et des profils de scripteurs ?

Notre hypothèse est qu'on peut décrire les compétences orthographiques de scripteurs même en grande difficulté autrement que sous l'angle d'une opposition entre forme correcte et forme erronée, et qu'une analyse plus fine est possible en mobilisant les apports de la linguistique. Plus précisément, nous proposons d'isoler la composante phonographique comme une première composante clé de la compétence orthographique et de montrer en

quoi cette composante est décisive pour cerner des degrés de compétence. Cependant, ce premier niveau ne suffit pas pour préciser les compétences plus avancées : c'est alors la morphographie qui doit retenir l'attention.

Par ailleurs, nous faisons l'hypothèse qu'un test ne portant même que sur quelques items, à condition qu'ils soient soigneusement sélectionnés, doit permettre de préciser les degrés de compétence. D'abord au niveau phonographique avec des mots présentant des correspondances stables entre phonèmes et graphèmes et des mots présentant des difficultés liées à l'existence de plusieurs graphèmes possibles pour un même phonème, puis au niveau morphographique avec la présence de marques de pluriel dans le groupe nominal et entre groupe sujet et groupe verbal.

L'enjeu de notre étude est d'affiner l'analyse de ce que savent et ne savent pas faire les publics en difficulté avec l'écrit. Un second enjeu est de fournir des recommandations pour l'évaluation de l'orthographe des publics en difficulté avec l'écrit, notamment par rapport au choix des mots ou plus largement des items dans une évaluation de type dictée. Il est important de s'interroger sur le choix du type de test ainsi que sur son contenu lorsqu'il s'agit d'évaluer l'écriture et l'orthographe de publics dont on sait qu'ils ne maîtrisent pas la norme orthographique. Que cherche-t-on à évaluer exactement ? La capacité à communiquer par écrit pour un usage courant ou la maîtrise parfaite du système orthographique d'une langue ? De quelles compétences a-t-on besoin pour communiquer par écrit et quels critères prendre en considération pour évaluer ces compétences ? Enfin, où se situe la frontière entre la maîtrise du système ou d'une partie du système et sa non maîtrise ?

Notre recherche se veut ainsi un préalable servant à mieux outiller les actions de formation, par la prise en compte des difficultés des scripteurs de manière plus ciblée, et par la mise en place de diagnostics permettant d'orienter le public vers des modules adaptés.

Comme nous l'avons dit, notre étude s'intéresse, aux difficultés orthographiques rencontrées par les Français lorsqu'ils sont dans une situation de la vie quotidienne, en caractérisant leurs connaissances en orthographe : au-delà de la distinction usuelle entre orthographe lexicale et grammaticale, nous nous penchons sur les compétences en orthographe lexicale, en morphographie mais également en phonographie.

Nous le ferons concrètement, en nous appuyant d'abord sur 441 dictées produites par des scripteurs considérés en difficulté avec l'écrit dans l'enquête d'utilité publique conduite à domicile IVQ (INSEE 2011) puis en comparant ces productions avec les écrits de 87 scripteurs

qui n'ont pas été considérés en difficulté avec l'écrit dans IVQ 2011 et que nous avons rencontrés, à domicile également, en 2014, lors d'une extension de l'enquête IVQ.

Nous décrivons dans un premier temps, les graphies choisies par les scripteurs pour transcrire la dictée, en utilisant le critère orthographique : le mot possède-t-il la forme orthographique attendue pour pouvoir affirmer qu'il respecte la norme ? Possède-t-il une marque flexionnelle ? Si oui, cette marque est-elle celle attendue (accord en nombre ou en genre) ? Dans un deuxième temps, après avoir écarté les formes qui respectent la norme, nous décrivons finement les choix graphiques des scripteurs en nous positionnant au niveau du phonème et de sa transcription (le graphème). Cette analyse nous permettra de voir si les enquêtés qui n'ont pas su restituer la forme normée, ont pu en revanche produire une graphie de type phonographique qui rende le message lisible. L'écriture sera ici analysée comme un système de représentation de la parole au-delà (ou en deçà) de la norme, ce qui nous amènera à repenser le concept de maîtrise de l'écriture (et surtout de l'orthographe) comme critère de compétence à l'écrit pour participer à l'élaboration du taux d'illettrisme dans la population.

Ainsi, nous verrons qu'un scripteur qui a mis en place un système d'écriture suffisamment efficace pour transmettre un message lisible, ne devrait pas être considéré comme illettré. Ce serait nier que ce dernier possède des connaissances phonographiques non négligeables, les connaissances orthographiques relevant d'un niveau supérieur de maîtrise qu'il n'aurait pas acquis dans sa totalité.

CHAPITRE 4. PRÉSENTATION DES ENQUÊTES ET DES CHOIX MÉTHODOLOGIQUES

1. PRÉSENTATION DE LA DICTÉE D'IVQ : CHOIX DES ITEMS ET PERSPECTIVES D'ANALYSES	173
1.1. LE CHOIX DES ITEMS	173
1.2. NOTRE APPROCHE DE LA DICTÉE D'IVQ	176
1.3. LES PSEUDO-MOTS DE LA DICTÉE D'IVQ	185
2. TRAITEMENT DES DICTÉES.....	186
2.1. CRITÈRES ET ATTENTES DE L'INSEE SUR LA DICTÉE : LE CALCUL DES SCORES	186
2.2. UN NÉCESSAIRE RECODAGE DU CORPUS POUR UNE ANALYSE DIFFÉRENTE	187
3. ÉVALUER LES CAPACITÉS D'ÉCRITURE D'ADULTES EN DIFFICULTÉ À PARTIR D'UNE DICTÉE : LA NÉCESSITÉ D'UNE ANALYSE PARTICULIÈRE.....	195
3.1. LE CRITÈRE ORTHOGRAPHIQUE APPLIQUÉ AUX MOTS	195
3.2. LE CRITÈRE PHONOGRAPHIQUE APPLIQUÉ AUX MOTS.....	198
3.3. LE CRITÈRE PHONOGRAPHIQUE APPLIQUÉ AUX PSEUDO-MOTS	203
4. LE PROJET « COMPÉTENCES ORTHOGRAPHIQUES DES ADULTES ET CONTEXTES SOCIAUX D'USAGE »	204
4.1. OBJECTIFS ET INTÉRÊT DU PROJET POUR L'ANALYSE DE LA DICTÉE D'IVQ	204
4.2. UNE ENQUÊTE DE TERRAIN.....	206
4.3. LE TRAITEMENT DES DONNÉES DE L'ENQUÊTE COMPLÉMENTAIRE	219
5. SÉLECTION DU CORPUS POUR NOS ANALYSES	229

Évaluer le niveau de littéracie chez les adultes n'est pas une préoccupation récente (voir chapitre 1, section 2) et les tests existants (IALS, ALL, etc.) présentent une carence en matière d'écriture puisqu'ils sont essentiellement basés sur l'évaluation des compétences en lecture et en compréhension. Comme nous l'avons vu, l'enquête « Information et Vie Quotidienne » (IVQ) a été conçue pour disposer d'une mesure des compétences en littéracie des adultes et de leur usage de l'information au quotidien. Cette enquête a la particularité de prendre en compte tous les niveaux de compétences (notamment les faibles niveaux pour estimer l'importance de l'illettrisme) et elle évalue ainsi toutes les composantes de l'écrit nécessaires à la communication : **compréhension écrite** et **orale** (comprendre ce qui est lu et entendu), **lecture** (savoir déchiffrer), **numératie** (savoir compter), **écriture** (savoir écrire).

Notre étude s'intéresse au versant écriture de l'enquête IVQ 2011 et propose d'analyser 447 dictées produites par les enquêtés des régions⁵⁸ Auvergne, Bourgogne, Provence-Alpes-Côte d'Azur et Rhône-Alpes considérés en difficulté dans cette enquête⁵⁹. Les enquêtés qui ont passé la dictée (16% au niveau national), nous le rappelons, sont ceux qui ont été considérés en possible difficulté à l'écrit suite au test d'orientation (composé d'exercices de compréhension orale, de lecture et de numératie).

Pour mieux évaluer les compétences spécifiques dans le domaine de l'écrit des personnes jugées en difficulté avec l'écrit, nous disposons d'un second corpus de dictées produit par des personnes qui n'ont pas été considérées en difficulté avec l'écrit dans l'enquête IVQ⁶⁰, en 2011. Le recueil de données a été réalisé dans le cadre du projet « Compétences orthographiques des adultes et contextes sociaux d'usage. ». Nous avons ainsi fait passer en 2014, la dictée d'IVQ augmentée de quelques items, assortie d'un questionnaire et d'un entretien, à 87 enquêtés interrogés en 2011 dans le cadre d'IVQ mais qui ne présentaient alors pas, selon l'INSEE, de difficulté particulière avec l'écrit (et n'avaient donc pas été évalués en production écrite).

Dans cette partie dédiée à la méthodologie, nous présentons dans un premier temps les items de la production écrite choisis par l'équipe de Besse (2004), les objectifs et les critères de correction qui ont été appliqués par l'INSEE. Nous verrons que l'analyse de la partie production écrite de l'enquête IVQ a consisté largement à évaluer la maîtrise de l'orthographe

⁵⁸ Ce découpage des régions est celui qui existait avant le redécoupage de 2015.

⁵⁹ Voir annexe 1, un exemple de dictée produite par un enquêté de cet échantillon de scripteurs en difficulté avec l'écrit (enquêté « bas »).

⁶⁰ Voir annexe 2, un exemple de dictée produite par un enquêté de cet échantillon de scripteurs qui n'ont pas été considérés en difficulté avec l'écrit dans IVQ 2011 (enquêté « haut »).

chez les scripteurs. Pour évaluer le niveau de difficulté de la dictée et les compétences mobilisées par les scripteurs, nous observons les items de la dictée à la lumière de différents travaux réalisés dans le domaine de la psycholinguistique. Pour effectuer ce travail, nous présentons les items de la dictée selon leur indice de consistance et de fréquence.

Nous proposons ensuite d'analyser ces dictées d'adultes faiblement littéraciés en terme de compétences et pas seulement en terme de manques en matière d'orthographe. Nous décrivons alors le travail de vérification des données et le recodage que nous avons effectué autour de la problématique de l'informatisation d'un corpus manuscrit. Puis nous détaillons les critères de correction et les différents barèmes que nous avons appliqués.

Dans une seconde partie, nous présentons enfin le second corpus (et les conditions de son recueil) qui nous permet de bénéficier d'écrits d'adultes qui n'ont pas été considérés en difficulté avec l'écrit dans l'enquête IVQ de 2011 et qui vont servir d'écrits de référence par rapport à l'usage et non par rapport à la norme.

1. PRÉSENTATION DE LA DICTÉE D'IVQ : CHOIX DES ITEMS ET PERSPECTIVES D'ANALYSES

1.1. LE CHOIX DES ITEMS

C'est l'équipe PsyEF dirigée par J-M. Besse qui a été chargée d'élaborer le module « bas » destiné aux enquêtés repérés en difficulté avec l'écrit suite au module d'orientation, dont la production écrite. Si les concepteurs du test présentent l'exercice (Besse et al., 2009) comme une *production écrite*⁶¹, celle-ci, qui se présente sous la forme d'une liste de mots à reproduire *sous la dictée*, n'en reste pas moins une dictée ; aucun ajout personnel n'étant autorisé dans ce type d'exercice. Aussi, nous utiliserons désormais le terme de dictée. Les enquêtés disposaient d'une feuille leur permettant de fournir leurs réponses de façon manuscrite, sans qu'aucune indication ne leur soit donnée quant à la manière d'organiser la dictée (ponctuation, signe typographiques, format liste ou texte, etc.). Cette liberté au sein d'un exercice contraint n'a pas été exploitée par l'INSEE mais pourrait faire l'objet d'une analyse⁶².

⁶¹ Dans l'objectif de rendre l'épreuve « rassurante » et de pas faire référence à un exercice connoté scolairement, Besse et al (2009) ont choisi de ne pas faire mention d'une « dictée ». Ce choix va de pair avec le format de cet exercice, présenté comme une liste de courses. Le terme « dictée » n'apparaît ni dans le protocole, ni dans les consignes à destination des enquêtés.

⁶² Les aspects liés à la matérialité de l'écriture (utilisation des majuscules, organisation spatiale, etc.) ont été codées en ce sens pour les 106 dictées du corpus rhônalpin (voir section 2.2.2 qui présente brièvement ce travail).

La version proposée dans l'enquête de 2011, se présente sous la forme suivante (les pseudo-mots sont indiqués en italiques) :

(RAYON) PHARMACIE :

ANTI-RHUME *MICATOL* OU *SIRAP*

ALCOOL A 90° *DUX*

(RAYON) ÉPICERIE :

TOMATES DE PAYS

CONFITURE

CERISES

SEL GOBAR

TROIS FROMAGES QUI NE SENTENT PAS FORT

(RAYON) LIBRAIRIE :

LE FIANCÉ SOLENNEL DE *JOC PÉVANOR*

L'objectif de cet exercice de production écrite était pour Besse et son équipe (2009), de :

- rendre compte des connaissances lexicales ;
- définir les compétences en morphographie ;
- vérifier les compétences en phonographie ;
- vérifier l'utilisation des procédures utilisées pour orthographier (voie d'assemblage et/ou d'adressage).

Nous détaillons chaque compétence testée :

Rendre compte des connaissances lexicales. Besse et al. ont choisi 12 mots en fonction de trois critères : la régularité des correspondances phonie-graphie (leur consistance orthographique), la longueur phonologique et la fréquence :

- La régularité des correspondances phonie-graphie :
 - o 6 mots considérés comme réguliers (consistants) : SEL, ANTI, TOMATES, CERISES, ÉPICERIE, CONFITURE.
 - o 6 mots considérés comme irréguliers (peu consistants) : PAYS, RHUME, ALCOOL, FIANCÉ, SOLENNEL, PHARMACIE.
- La longueur phonologique :
 - o 3 phonèmes : SEL, ANTI, PAYS, RHUME ;
 - o 5 phonèmes : TOMATES, CERISES, ALCOOL ;
 - o 6 phonèmes : FIANCÉ ;

- 7 phonèmes : ÉPICERIE⁶³, CONFITURE, SOLENNEL, PHARMACIE.

Leur fréquence a été estimée dans l'échelle orthographique Dubois-Buyse avec des mots dont l'orthographe est considérée comme connue à l'âge de 14 ans⁶⁴. Pour Pothier et Pothier (2003) il s'agit des mots sur lesquels il est observé plus de 75% de réussite chez les élèves en fin de cycle 3.

Définir les compétences en morphographie des enquêtés lorsqu'il s'agit de produire des marques de nombre muettes selon des règles syntaxiques simples (accord déterminant-nom et sujet-verbe) :

- accord en nombre d'un nom : FROMAGES ;
- accord en nombre d'un verbe : SENTENT.

Vérifier les compétences en phonographie et la capacité à utiliser la voie d'assemblage pour transcrire les 6 pseudo-mots :

- MICATOL : /mikatɔl/ ;
- SIRAP : /sirap/ ;
- DUX : /dyks/ ;
- GOBAR : /gobaR/ ;
- JOC : /ʒɔk/ ;
- PÉVANOR : /pevanɔR/.

Vérifier l'utilisation de la voie d'assemblage et/ou d'adressage :

- voie d'assemblage et/ou d'adressage pour orthographier les mots réguliers ;
- voie d'adressage pour orthographier les mots irréguliers et les pseudo-mots.

Selon Besse et al. (2004), la voie d'assemblage consiste à identifier les sons puis à les transcrire avec les graphèmes qui correspondent, en utilisant une procédure phonographique pour orthographier. Cette procédure est utilisée chez le scripteur qui cherche à écrire un mot qui lui est inconnu ou une partie de mot. La voie d'adressage consiste à rappeler un mot ou une partie de mot stocké en mémoire. Besse et al. (2009) rappellent que l'utilisation de la voie

⁶³ Nous notons ici que la prise en compte de la chute du schwa amène à considérer que cet item est composé de 6 phonèmes : /epiSri/.

⁶⁴ Dans l'échelle Dubois-Buyse, les mots classés dans les échelons inférieurs à 35 correspondent aux mots considérés comme connus à l'âge de 14 ans parce qu'ils ont été réussis par au moins 75% des élèves (Ters et al., 1995).

d'assemblage (uniquement) en français conduit souvent à une écriture orthographiquement incorrecte et à tendance à perturber l'écriture.

1.2. NOTRE APPROCHE DE LA DICTÉE D'IVQ

Nous ne discuterons pas ici de la pertinence de cette dictée pour déterminer les difficultés en orthographe des scripteurs. Cet aspect sera traité dans la partie consacrée aux perspectives didactiques, une fois analysées les graphies produites.

Nous retenons que cet exercice, même s'il possède des limites en termes d'analyses (items peu nombreux, phénomènes linguistiques restreints notamment pour évaluer les compétences en morphographie), nous permet de disposer d'un corpus inédit d'écrits de personnes considérées en possible difficulté avec l'écriture. Par ailleurs, sa forme (une liste composée d'items et de syntagmes) nous permet d'établir des mesures quantitatives et de décrire finement les réponses des enquêtés lorsqu'ils doivent produire des mots écrits et transcrire des mots qui n'existent pas (pseudo-mots). Cette opportunité rare d'observer des comportements de scripteurs en difficulté nous permettra de repérer différents types de compétences et d'établir des profils de scripteurs.

En première approche, parmi les phénomènes linguistiques les plus marquants de cette dictée, nous pouvons citer :

- lettres internes muettes : le « h » de RHUME ;
- lettres finales muettes : le « -s » de PAYS, le « -t » de FORT, le « -e » de PHARMACIE, ÉPICERIE ET LIBRAIRIE ;
- graphies particulières, qui possèdent une structure rare : PAYS, SOLENNEL ;
- choix des digrammes « en » ou « an » pour le phonème /ã/ : SENTENT, ANTI ;
- accord nominal sur FROMAGES et accord du verbe SENTENT, dans le contexte particulier d'une expansion par proposition relative (TROIS FROMAGES QUI NE SENTENT PAS FORT) ;
- transcription des 6 pseudo-mots présentés comme des noms de marque et le nom d'un auteur, qui demande au scripteur de recourir spécifiquement à un traitement phonographique des 18 phonèmes proposés :
 - /i/, /a/, /o/, /ɔ/, /e/, /y/ ;
 - /b/, /d/, /g/, /ʒ/, /k/, /l/, /m/, /n/, /p/, /R/, /s/, /t/, /v/.

Ainsi, la dictée ne présente pas de difficultés linguistiques majeures : le lexique semble accessible et la morphographie qui représente une source majeure de difficultés en français est peu présente. Aussi, nous proposons d'enrichir les analyses orthographiques proposées

par l'INSEE par une analyse des graphies erronées pour observer la compétence phonographique élargie aux mots. Nous profiterons également de l'opportunité de disposer de graphies de pseudo-mots pour décrire finement les transcriptions proposées en observant particulièrement le choix des finales et en le comparant avec les choix opérés par des scripteurs plus experts.

1.2.1. LES MOTS DE LA DICTÉE D'IVQ PRÉSENTÉS SELON LEUR INDICE DE CONSISTANCE ET DE FRÉQUENCE

Pour déterminer le niveau de difficulté de la dictée et les compétences mobilisées par les adultes, nous choisissons de regarder les items à la lumière des travaux de Lété (2008) sur la notion de fréquence et de consistance décrites dans Manulex⁶⁵ (Lété et al., 2004 ; Peereman et al., 2007 ; Peereman et al., 2013). Nous utilisons également la base de données Lexique 3 qui fournit des indices de fréquence⁶⁶.

Nous présentons les indices de consistance et/ou de fréquence fournis par chaque base dans deux tableaux distincts.

1.2.1.1 CLASSEMENT DES MOTS DE LA DICTÉE D'IVQ SELON LEUR INDICE DE CONSISTANCE ET DE FRÉQUENCE CALCULÉS DANS LA BASE MANULEX-INFRA

Pour rendre compte du degré de difficulté de la dictée, nous présentons dans le tableau 3, le classement des mots de la dictée d'IVQ (sous la forme de lemme) selon leur indice de consistance (compris entre 0 et 100) et de fréquence U ⁶⁷. Pour chaque mot, nous indiquons son rang de classement par rapport aux autres mots, selon l'indice de consistance et de fréquence calculé dans la base Manulex-Infra (1 étant le mot qui apparaît comme le plus consistant et le plus fréquent dans Manulex). Notons que l'item ANTI est absent de Manulex-Infra.

Cet outil permet également de choisir entre le lemme ou la forme fléchie. Lorsque nous les avons trouvés dans la base, nous avons relevé les indices de consistance et de fréquence des formes fléchies telles qu'elle étaient attendues dans l'exercice (FROMAGES, SENTENT). Nous

⁶⁵ Pour une présentation plus détaillée de l'outil MANULEX et de ses exploitations possibles, se référer à la partie théorique (chapitre 3, section 1).

⁶⁶ Pour une présentation de l'outil LEXIQUE, voir la partie théorique (chapitre 3, section 1.2.1.2)

⁶⁷ Pour rappel, U représente la fréquence estimée d'usage. Cet indice prend en compte la dispersion des mots dans le corpus ainsi que leur fréquence réelle, ce qui permet de disposer d'un indicateur plus fiable de la fréquence des mots (pour davantage d'explications, se référer au chap.3, section 1.2).

avons également relevé les indices de consistance et de fréquence des graphies observées dans les réponses des enquêtés (indiquées en italique dans le tableau), par exemple :

- le « -s » de TOMATE(S) ou de CERISE(S), qui n'était pas attendu puisque les items étaient dictés sans déterminant ;
- les marques de genre ou de nombre erronées : FIANCÉE pour FIANCÉ, SOLENNELLE pour SOLENNEL, LIBRAIRIES pour LIBRAIRIE, etc. ;
- les homophones SELLE vs SEL, RHUM vs RHUME qui pourraient expliquer certaines confusions ;
- l'item FORT selon qu'il est utilisé en tant qu'adjectif ou adverbe (comme dans la dictée) ;
- etc.

Les calculs sont effectués « par Token »⁶⁸, sur le mot entier (moyenne des différentes associations dans le mot) et à partir de tous les manuels de la base (moyenne du CP au CM2).

⁶⁸ Les calculs "par token" prennent en compte le nombre d'occurrences des mots dans les textes (c'est-à-dire leur fréquence réelle), ce qui permet de prendre en considération la rareté d'un mot. Sans cela, un mot qui apparaît fréquemment dans un texte aurait le même poids qu'un mot qui apparaît plus rarement.

mots	consistance	rang « consistance »	fréquence U	rang « fréquence »
SOLENNEL	86,64	1	2,53	14
<i>SOLENNELLE</i>	<i>67,22</i>		<i>0,94</i>	
TOMATE	82,99	2	23,87	6
<i>TOMATES</i>	<i>79,37</i>		<i>23,77</i>	
FROMAGE	81,72	3	118,16	3
FROMAGES	78,72		16,04	
SENTIR	81,04	4	22,68	7
SENTENT	48,76		5,92	
<i>SENTES</i>	<i>8,75</i>		<i>0,02</i>	
SEL	80,52	5	78,43	4
<i>SELLE</i>	<i>48,07</i>		<i>23,26</i>	
PAYS	79,25	6	309,72	1
FORT	78,31	7	302,95	2
<i>FORTS (ADJ)</i>	<i>67,37</i>		<i>18,55</i>	
LIBRAIRIE	77,63	8	11,13	11
<i>LIBRAIRIES</i>	<i>75,41</i>		<i>0,02</i>	
CONFITURE	76,01	9	42,33	5
<i>CONFITURES</i>	<i>73,97</i>		<i>8,10</i>	
CERISE	72,03	10	11,67	10
<i>CERISES</i>	<i>84,55</i>		<i>25,41</i>	
PHARMACIE	67,17	11	16,84	9
<i>PHARMACIES</i>	<i>66,12</i>		<i>0,46</i>	
ÉPICERIE	64,86	12	7,14	13
<i>ÉPICERIES</i>	<i>64,06</i>		<i>0,01</i>	
RHUME	61,74	13	20,21	8
<i>RHUM</i>	<i>58,96</i>		<i>7,84</i>	
ALCOOL	61,47	14	7,35	12
FIANCÉ	59,19	15	0,46	15
<i>FIANCÉE</i>	<i>62,41</i>		<i>2,10</i>	

Tableau 3. Classement des mots de la dictée d'IVQ et de certaines graphies produites par les enquêtés selon leur degré de consistance (consistance PO totale, du CP au CM2, calcul par token) et leur indice de fréquence, calculés dans Manulex-Infra.

L'observation des indices de consistance montre que dans cet exercice, les enquêtés ont été confrontés à :

- des mots plutôt réguliers, dont le degré de consistance est élevé (avec un indice supérieur à 80 sur 100) : TOMATE, FROMAGE, SENTIR (sans la forme fléchie), SEL ;
- les mots CONFITURE, LIBRAIRIE, FORT, PAYS sont également plutôt consistants, avec des indices proches de 80, compris entre 76 et 79 sur 100 ;
- des mots moins réguliers dont le degré de consistance est plus faible comme RHUME, ALCOOL, ÉPICERIE et PHARMACIE qui ont un indice de consistance compris entre 61 et 67 sur 100. Le mot le moins consistant de la dictée d'IVQ est FIANCÉ, avec un indice de 59 sur 100.

Nous relevons que le mot SOLENNEL est considéré comme très consistant par Manulex. Il semblerait que ce soit la rareté de la correspondance du phonème /a/ avec la réalisation graphique « en » qui explique ce résultat. Nous observons le même phénomène sur le mot PAYS, qui est considéré comme plutôt consistant avec un indice de 79 sur 100.

Nous verrons avec ces deux exemples que les calculs proposés par cet outil ne prédisent pas le taux de restitution de cet item SOLENNEL. Puisque cet item est considéré comme étant le mot le plus consistant de la dictée, il devrait également être le mieux restitué ; or nous observerons le contraire.

La fréquence des mots peut également constituer une explication aux difficultés rencontrées par les scripteurs⁶⁹, notamment lorsque la graphie est plutôt irrégulière. Dans ce corpus, s'opposent ainsi :

- des mots fréquents comme PAYS, FORT qui ont un indice élevé (U = 302 et 309), ainsi que FROMAGE (U = 118), qui est le troisième mot le plus fréquent de la dictée ;
- des mots dont l'exposition semble bien moindre : ÉPICERIE, SOLENNEL, FIANCÉ.
- parmi les formes qui ont un ou plusieurs homophones comme SEL vs SELLE : c'est la forme la plus fréquente (SEL) qui a été utilisée dans la dictée ;
- la graphie RHUM (qui désigne la boisson et que nous retrouvons dans certaines dictées est également bien moins fréquente que la forme attendue RHUME (U = 7 vs 20).

⁶⁹ Se reporter à la partie théorique (chapitre 3, section 1.2) pour les études qui montrent l'impact de la fréquence dans les compétences orthographiques.

COMPARAISON DES INDICES DE FRÉQUENCE ENTRE LEMMES ET FORMES FLÉCHIES

Selon Manulex, les items TOMATE et CERISE ne sont pas plus fréquents, qu'ils soient présentés sous forme de lemme ou sous leur forme fléchie (TOMATES, CERISES).

En revanche FROMAGE apparaît plus fréquemment au singulier dans les manuels (indice U = 118 vs 16), tout comme CONFITURE (indice U = 42) vs CONFITURES (indice U = 8,1).

Les items ANTI, ALCOOLS, PHARMACIES, ÉPICERIES, LIBRAIRIES ne se trouvent pas dans la base ou ont un indice de fréquence très faible (indice U < 1).

Par ailleurs, nous observons parmi les mots qui possèdent un ou plusieurs homophones, que les enquêtés n'ont en grande majorité pas fait de confusion : ils ont écrit SEL plutôt que SELLE ou CELLE. En revanche pour RHUME, la proximité avec la graphie RHUM, qui désigne la boisson, a perturbé les enquêtés « bas » chez qui l'on retrouve de nombreuses formes RHUM.

1.2.1.2 CLASSEMENT DES MOTS DE LA DICTÉE SELON LEUR INDICE DE FRÉQUENCE CALCULÉ DANS LA BASE LEXIQUE 3

Nous présentons ensuite dans le tableau 4, l'indice de fréquence U des mots de la dictée d'IVQ calculé dans la base Lexique 3. Les mots sont présentés sous forme de lemme ou avec la forme fléchie pour les mots TOMATES, FROMAGES, SENTENT, CERISES.

Le rang indique le classement du mot par rapport aux autres mots de la dictée, selon l'indice de fréquence calculé dans la base Lexique (le mot classé en 1 représente celui qui apparaît comme le plus fréquent, par rapport aux autres mots de la dictée). Pour faciliter la lecture, les mots de la dictée sont classés par ordre décroissant, du plus fréquent au moins fréquent, dans la base « livres ».

mots	U Lexique « livres »	rang	U Lexique « films »	rang
SENTIR	718,78	1	535,40	1
PAYS	242,55	2	202,57	2
FORT	212,77	3	171,91	3
ALCOOL	42,84	4	41,30	5
SEL	33,24	5	21,83	7
FROMAGE	26,96	6	27,22	6
FROMAGES	6,15		1,54	
FIANCÉ	23,11	7	51,51	4
SOLENNEL	19,66	8	4,58	15
CONFITURE	15,41	9	7,41	11
TOMATE	13,31	10	20,77	8
TOMATES	7,57		12,89	
SENTENT ⁷⁰	11,08	11	11,21	
PHARMACIE	10,20	12	10,46	9
CERISE	10,07	13	5,26	14
CERISES	6,76		2,52	
LIBRAIRIE	9,59	14	5,97	13
ÉPICERIE	9,46	15	6,79	12
RHUME	5,61	16	8,17	10

Tableau 4. Classement des mots de la dictée d'IVQ selon leur indice de fréquence calculé dans la base LEXIQUE 3.

Ce classement des items selon leur indice de fréquence U calculé dans la base Lexique 3 montre une différence parfois importante selon que cet indice est calculé dans les manuels scolaires (Manulex-Infra), dans la littérature (Lexique « livres ») et dans les films (Lexique « films »).

Dans Manulex-Infra, les mots les plus fréquents⁷¹ sont PAYS, FORT, FROMAGE et les moins fréquents : ÉPICERIE, SOLENNEL, FIANCÉ.

En revanche, dans Lexique 3, les mots les plus fréquents sont SENTIR, PAYS, FORT (dans les bases « livres » et « films », les 3 mots les plus fréquents sont identiques même si les indices varient) et les 3 mots les moins fréquents sont LIBRAIRIE, ÉPICERIE, RHUME dans la base « livres » et LIBRAIRIE, CERISE, SOLENNEL dans la base « films ».

⁷⁰ Pour la forme fléchie SENTENT, la base n'indique pas s'il s'agit du présent de de l'indicatif ou du subjonctif.

⁷¹ Nous ne faisons pas référence au mot SOLENNEL comme étant le plus fréquent car nous estimons qu'il s'agit ici d'un biais.

Cela questionne sur le corpus d'exposition qui serait le plus approprié à des adultes. Nous reviendrons sur cette question lors de l'analyse des dictées, lorsque nous observerons les mots les mieux réussis et les moins bien réussis dans les deux enquêtes.

1.2.2. LES MOTS DE LA DICTÉE D'IVQ PRÉSENTÉS SELON LEUR SCORE DANS EOLE

Nous présentons dans le tableau 5, les scores donnés dans EOLE pour les mots de la dictée d'IVQ. Certains mots de nos dictées n'y apparaissent pas car ils ne font pas partie des 12000 termes qui ont été retenus dans EOLE pour être les plus courants du français écrit des années 2000.

Nous avons ajouté des remarques sur la forme attendue dans la dictée, qui ne testait pas toujours uniquement la forme lexicale (par ex. dans FROMAGES QUI NE SENTENT PAS FORT).

mots d'IVQ	CP	CE1	CE2	CM1	CM2
ALCOOL	0	5	25	43	59
CERISE	19	58	76	85	96
CONFITURE	50	86	96	96	96
ÉPICERIE	4	19	22	40	40
FIANCÉ	12	14	28	58	55
FORT ⁷²	25	71	70	96	100
FROMAGE	27	77	80	85	93
LIBRAIRIE	3	33	68	63	63
PAYS	5	52	69	78	87
PHARMACIE	10	38	57	78	87
RHUM	0	0	9	16	15
RHUME	0	8	3	32	42
SEL	54	76	91	98	100
SENTIR	30	40	68	92	90
SOLENNEL	0	0	0	2	7
SOLENNELLE	0	3	6	5	19
TOMATE	63	67	88	92	90

Tableau 5. Pourcentages de réussite de mots de la dictée d'IVQ d'après EOLE.

Les données issues d'EOLE complètent les données sur la fréquence et la consistance orthographique qui sont fournies par les bases Manulex-Infra (Lété, Sprenger-Charolles, Colé, 2004) et Lexique. Il améliore notamment ces outils qui permettent de prédire la réussite ou la

⁷² Sur ce mot, la catégorie grammaticale du mot utilisé dans la dictée n'est pas précisée : nom, adjectif ou adverbe.

difficulté mais qui ne fournissent pas en revanche de données sur les productions effectives des élèves. Il reste néanmoins limité pour une analyse plus fine des réussites et des difficultés orthographiques parce qu'il n'intègre pas les mots-outils et ne tient pas compte de manière systématique des classes grammaticales et des formes fléchies.

Le fait de disposer de données sur les productions reste un atout majeur car les indices de réussite montrent que l'acquisition est en règle générale linéaire. Le nombre de termes atteignant un score de 75 % de réussite est de 347 au CP, 1013 au CE1 ; 2188 au CE2, 3606 au CM1, 5054 au CM2). En revanche, les nouveaux apprentissages orthographiques introduisent aussi des doutes et des erreurs qui n'apparaissent pas auparavant, que nous observons dans les mots des dictées, avec ARRACHER, PAQUET et VIS.

Pothier et Pothier (2003) estiment que les mots qui atteignent 75% de réussite peuvent être considérés comme acquis par l'ensemble d'une classe d'âge (ou comme devant être acquis pour ceux qui n'en maîtrisent l'orthographe).

Nous observons dans la figure 5, les taux de réussite en CM2 sur les items de la dictée d'IVQ. Les mots qui sont réussis à plus de 75% sont indiqués en gris clair, ceux qui sont réussis à moins de 75% sont en gris foncé.

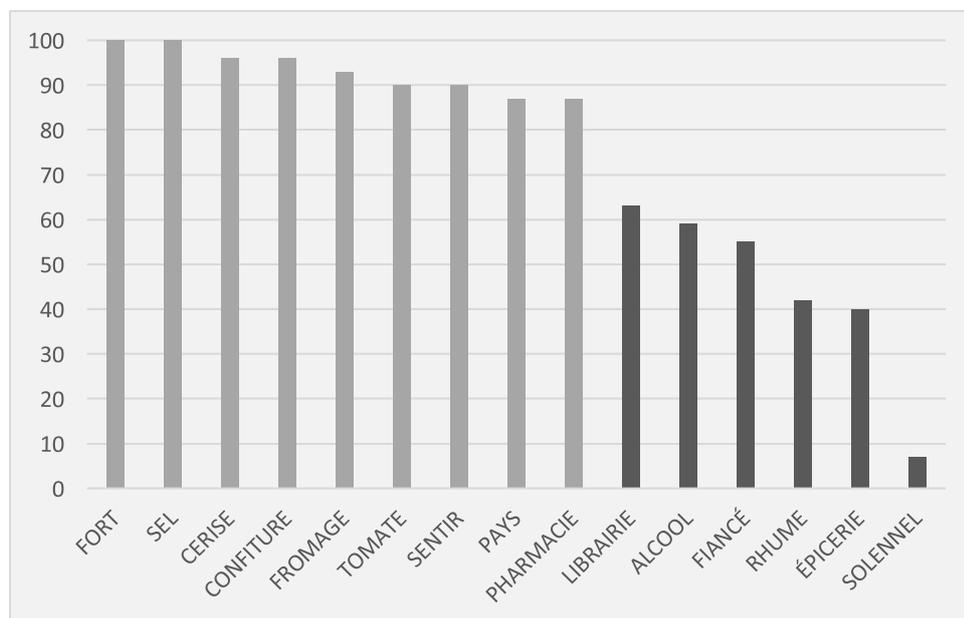


Figure 5. Classement des mots d'IVQ selon leur pourcentage de réussite dans EOLE, chez des élèves de CM2.

Dans EOLE, certains mots de la dictée d'IVQ obtiennent de très bons taux de réussite chez des élèves de CM2 : FORT, SEL, CERISE, CONFITURE, FROMAGE. En revanche, 6 mots obtiennent moins de 75% de restitution orthographique dont en particulier RHUME (42%), ÉPICERIE (40%) et SOLENNEL (7%) qui sont réussis par moins d'un élève sur deux. Nous regarderons dans les

résultats si ces 6 mots (que Pothier et Pothier, 2003 considèrent comme non acquis en CM2) ont été réussis par les enquêtés « bas ».

1.2.3. LE CHOIX DES MOTS : BILAN

Les mots de la dictée d'IVQ présentent des différences en termes de consistance, avec des mots très consistants qui ont un indice supérieur à 80 sur 100 et des mots moins consistants avec pour indice le plus faible 59 sur 100 (FIANCÉ). Nous émettons une réserve sur le calcul de l'indice de consistance des mots SOLENNEL et PAYS, qui selon Manulex-Infra apparaissent comme très consistants.

En nous basant sur le calcul de l'indice de fréquence proposé dans Lexique 3, il apparaît que les mots les plus fréquents (SENTIR, PAYS, FORT) font partie des mots les plus consistants, d'après Manulex-Infra. Les moins fréquents dans Lexique 3⁷³ comme ÉPICERIE, RHUME, SOLENNEL font partie des mots qui sont les moins consistants du corpus.

L'observation des items de la dictée d'IVQ d'après les travaux réalisés en psycholinguistique que nous mettrons en lien avec les résultats obtenus par les enquêtés pour chaque item, permettra de nous interroger sur la pertinence de certains mots pour tester des compétences de base. Par exemple SOLENNEL, mot peu fréquent et dont la correspondance /a/ avec la graphie « en » pourrait se révéler difficile à transcrire. La difficulté de cet item transparait par ailleurs dans EOLE, avec seulement 2% d'élèves de CM1 et 7% d'élèves de CM2 qui produisent une graphie orthographiquement conforme.

1.3. LES PSEUDO-MOTS DE LA DICTÉE D'IVQ

La dictée d'IVQ comporte des pseudo-mots qui ont été choisis par l'équipe de Besse et al. pour vérifier la capacité des scripteurs à recourir à un traitement phonographique de formes entendues lorsque celles-ci n'existent pas.

Les six pseudo-mots : MICATOL, SIRAPE, DUX, GOBAR, JOC, PÉVANOR rassemblent 18 phonèmes :

- 6 phonèmes vocaliques : /a/, /e/, /i/, /o/, /ɔ/, /y/ ;
- 13 phonèmes consonantiques : /b/, /d/, /g/, /ʒ/, /k/, /l/, /m/, /p/, /R/, /s/, /t/, /v/, /n/.

Outre la vérification du niveau de conscience phonographique des enquêtés, nous avons choisi d'analyser les choix graphiques des enquêtés au vu de la diversité des graphies possibles

⁷³ Nous utilisons ici les indices de fréquence les plus élevés dans Lexique « livres » et « films ».

pour une seule forme orale : par exemple *micatol*, *mycatol*, *micathole*, etc. pour MICATOL (/micatɔl/). Nous verrons alors si certaines graphies sont utilisées préférentiellement par les scripteurs.

Nous observerons également le choix des scripteurs concernant la possibilité d'ajouter une finale muette sur certains pseudo-mots ainsi que le choix de la finale :

- MICATOL : avec ou sans « -e » final ;
- SIRAP : avec ou sans « -e » final ;
- GOBAR : avec ou sans « -e », « -d », « -t » final ;
- PÉVANOR : avec ou sans « -e », « -d », « -t » final.

2. TRAITEMENT DES DICTÉES

Dans cette section, nous présentons dans un premier temps les critères et les attentes de l'INSEE pour la correction des dictées, qui ont conditionné le calcul des scores de réussite. Dans un deuxième temps, nous présentons nos critères et les différences de choix que nous avons fait. Ces différences ont conduit à un nécessaire recodage de ce corpus de dictées, qui est fonction de nos attentes en matière d'analyses.

2.1. CRITÈRES ET ATTENTES DE L'INSEE SUR LA DICTÉE : LE CALCUL DES SCORES

Pour évaluer les performances à l'écrit et déterminer le niveau de compétence en écriture des enquêtés, l'INSEE a réalisé des scores de réussite des items de la dictée à partir d'une analyse quantitative des données. Cette évaluation classique consiste à repérer les graphies incorrectes et à établir des scores de réussite.

L'INSEE a fait le choix de ne corriger que 18 items sur les 39 dictés car certains ont été jugés redondants ou non pertinents. L'évaluation des dictées a consisté à vérifier :

- le respect de la norme orthographique sur 12 mots avec 1 point attribué par mot : PHARMACIE, ANTI⁷⁴, RHUME, ALCOOL, ÉPICERIE, TOMATE(S), PAYS, CONFITURE, CERISE(S), SEL, FIANCÉ, SOLENNEL ;
- le respect de la morphographie avec 2 points attribués pour le « -s » de FROMAGES et 2 points pour le « -nt » de SENTENT ;

⁷⁴ Le mot composé ANTI-RHUME a été analysé comme deux mots distincts ANTI et RHUME.

- le respect du principe phonographique avec 1 point attribué pour chacun des 6 pseudo-mots qui présentent une graphie plausible.

Pour établir des statistiques, l'INSEE a saisi les réponses des enquêtés en attribuant pour chaque item, un code « Faux », « Absent » ou « Juste » permettant l'attribution de points. Un total de 22 points détermine ainsi le niveau de l'enquêté et cinq profils de scripteurs ont été définis :

- 0 : personnes n'ayant pas pu faire les exercices du fait de trop grandes difficultés ;
- 1 : difficultés graves (moins de 40 % de réussite = score inférieur à 9/22) ;
- 2 : difficultés fortes (entre 40 et 60 % de réussite = score compris entre 9 et 13/22) ;
- 3 : difficultés partielles (entre 60 et 80 % de réussite = score compris entre 14 et 17/22) ;
- 4 : pas de difficulté (plus de 80 % de réussite = score supérieur à 17/22) ;

Les profils 0, 1 et 2 qui regroupent les enquêtés qui n'ont produit aucun item et ceux dont le profil montre des difficultés graves ou fortes relèvent selon l'INSEE de situations dites « préoccupantes » à l'écrit, parmi lesquelles seront identifiées les personnes qui relèvent de l'illettrisme.

2.2. UN NÉCESSAIRE RECODAGE DU CORPUS POUR UNE ANALYSE DIFFÉRENTE

Pour mener notre étude, l'ANLCl a mis à notre disposition les 2047 dictées en version papier recueillies par l'INSEE en 2011 ainsi que le fichier informatique (réalisé par l'INSEE), qui a permis de mesurer de manière quantitative les données.

Après avoir comparé les dictées en version papier et le fichier informatique, nous avons fait un premier constat pour notre étude, compte tenu de nos objectifs : nous ne pouvions utiliser en l'état le fichier fourni par l'INSEE.

2.2.1. LE FICHIER DES RÉPONSES : UNE PERTE D'INFORMATION POUR OBSERVER LA VARIÉTÉ DES GRAPHIES

La saisie des réponses des scripteurs et le codage réalisé par l'INSEE avait pour objectif de permettre une analyse quantitative des données qui mette en relief la capacité des scripteurs à respecter l'orthographe de mots dictés (et les compétences en morphographie sur deux marques de nombre) et à respecter le principe phonographique du français en proposant une transcription acceptable pour les pseudo-mots. L'objectif n'était pas de décrire la variété des graphies.

Aussi, tous les items n'ont pas été saisis et étaient donc absents du fichier : QUI, NE, PAS, FORT, LIBRAIRIE. Certaines graphies sur les mots ont été uniformisées, notamment l'accentuation et les diacritiques (*fiancé* prend la forme de « *fiance* » dans le fichier) et la variété des graphies considérées comme conformes au critère phonographique sur les pseudo-mots n'a pas été relevée (toutes les graphies conformes comme *mikatol*, *mycathol*, etc. ont pris la forme « *micatol* » dans le fichier).

2.2.1.1 LE TRAVAIL DE SAISIE ET DE CODAGE RÉALISÉ PAR L'ANLCI

Pour faire apparaître la variété des graphies produites par les enquêtés, l'ANLCI a procédé dans un deuxième temps à un travail de saisie des graphies des pseudo-mots lorsque ceux-ci étaient conformes aux attentes des concepteurs du test. Apparaissent donc désormais les formes exactes produites par les scripteurs (*mikatol* saisi « micatol » par l'INSEE devient « mikatol » dans le fichier). En revanche, l'ANLCI n'a saisi ni l'accentuation, ni les diacritiques.

Les codeurs de l'ANLCI ont également ajouté dans le fichier informatique, les mots QUI, FORT et LIBRAIRIE qui n'avaient pas été retenus par l'INSEE pour calculer le score à l'épreuve de dictée (et n'avaient donc pas été relevés).

Ce travail a été réalisé après le calcul des scores proposés par l'INSEE et n'a pas été pris en compte dans les résultats publiés par l'INSEE, ni dans le calcul du taux d'illettrisme.

Nous détaillons ci-dessous les points qui nous empêchent de procéder à une analyse fine (à la fois quantitative et qualitative) des graphies des enquêtés, que celles-ci soient erronées ou correctes et que nous avons corrigées ou ajoutées dans le fichier.

2.2.1.2 LES SAISIES QUI S'ÉCARTENT DES RÉPONSES DES ENQUÊTÉS

Nous avons vu que pour pouvoir décrire les choix graphiques des enquêtés, le fichier informatique des données tel que nous l'avons reçu ne permet pas toujours d'observer la diversité des réponses lorsque celles-ci ont été considérées comme correctes. Nous présentons les choix qui ont été faits par l'INSEE.

LA MORPHOGRAPHIE : MARQUES NON ATTENDUES ET MARQUES ATTENDUES

Les mots TOMATES et CERISES ont été dictés sans déterminant. Néanmoins, si une marque de nombre n'était pas attendue, elle est en revanche souhaitable (on n'achète pas une seule cerise). L'observation de sa présence (ou de son absence) sur les graphies peut permettre de rendre compte des comportements des scripteurs face à des pluriels « implicitement » suggérés par le contexte. Ces marques de nombre n'entrent pas dans le barème

orthographique prévu par l'INSEE et les finales sur ces deux mots ont été uniformisées. Les graphies ont systématiquement été enregistrées avec la marque du pluriel : la graphie *tomate* est devenue « *tomates* » dans le fichier, tout comme la graphie *cerise* a été transformée en « *cerises* ».

Sur l'item FORT, une erreur (semble-t-il) de codage a conduit à uniformiser la marque de nombre sous la forme FORT, qui est erronée puisque l'item est ici un adverbe. C'est ainsi la graphie FORTS avec un « -s », qui a systématiquement été saisie sur cet adverbe (invariable)⁷⁵.

Enfin, sur les mots FIANCÉ et SOLENNEL, l'ajout erroné par les scripteurs, d'une marque de genre ou de nombre n'apparaît pas dans le fichier. Lorsqu'ils ont été orthographiés *fiancée-s* et *solennel-l-e-s* par les scripteurs, leurs réponses ont été uniformisées sous la forme qui était attendue par les concepteurs du test : « *fiance* » et « *solennel* ».

Les compétences en morphographie ont été évaluées sur deux marques de nombre : le « -s » de FROMAGES et le « -nt » de SENTENT. L'item SENTENT a donné lieu à des variantes telles SEMBLENT et SONT. L'INSEE a considéré ces dernières comme correctes, car le nombre a bien été marqué et les graphies ont été saisies dans le fichier informatique sous la forme attendue SENTENT.

LES IMPUTATIONS ET LE TRAITEMENT DES GRAPHIES ILLISIBLES

Les graphies qui ont été considérées comme "illisibles" ont été codées de la même manière que les graphies incorrectes (code 9), si bien que nous ne pouvons pas savoir, à partir du code, si la graphie est effectivement incorrecte ou s'il s'agit d'un problème de lecture (la graphie pouvant donc être correcte).

L'INSEE a également procédé à un travail d'imputation de mots orthographiés correctement pour les personnes qui n'ont pas répondu à tous les items. Ces travaux d'imputation se sont appuyés sur les données individuelles collectées lors de la passation du module biographique.

LE TRAITEMENT DES DIACRITIQUES ET DES AUXILIAIRES D'ÉCRITURE

L'accentuation n'a pas non plus été relevée dans le fichier informatique. L'INSEE a fait le choix de neutraliser tous les accents produits par les enquêtés et de considérer comme correctes toutes les formes non accentuées produites :

⁷⁵ Cette uniformisation des réponses FORTS n'a pas eu d'incidence dans le calcul des scores effectués par l'INSEE car cet item n'entre pas dans ce calcul.

- les graphies *fiancé* et *épicerie* apparaissent donc systématiquement sous la forme « *fiance* » et « *epicerie* » (sans accent), si bien qu'à partir du fichier, on ne peut pas savoir si l'enquêté a utilisé le graphème « é ».

À l'inverse, l'ajout erroné d'un accent, d'une cédille, d'une apostrophe, d'une segmentation à l'intérieur du mot par les enquêtés, n'a pas été relevé :

- *sél*, *tomâte* apparaissent sous la forme « *sel* », « *tomate* » dans le fichier ;
- *pharmaçie* apparaît sous la forme « *pharmacie* » dans le fichier ;
- *al'coole* apparaît sous la forme « *alcoole* » dans le fichier ;
- *p vanord* apparaît sous la forme « *pvanord* » dans le fichier ;

Les majuscules n'ont pas non plus été relevées : *Micatul* apparaît sous la forme « *micatul* ».

Ces constats montrent que le codage, s'il peut être utilisé pour établir des fréquences de réussite orthographique, ne peut pas être utilisé en l'état pour étudier la variation graphique des formes erronées pour les mots et ne peut donc pas être utilisé pour rendre compte précisément des capacités à l'écrit de ce public. C'est pour cette raison qu'un travail de saisie supplémentaire a été réalisé par l'ANLICI, qui a notamment corrigé les pseudo-mots (sur lesquels les graphies considérées comme phonographiquement correctes avaient été harmonisées).

Nous précisons que la reprise du codage et la correction des graphies ont eu très peu d'incidence dans le calcul des scores orthographiques. Par exemple « *fiance* » ou « *tomate* » saisi par l'INSEE et corrigées par les graphies réellement produites par les scripteurs, soit *fiancé* et *tômate* (que nous considérons en revanche comme orthographiquement non conforme). Si la reprise du codage a permis en effet de repérer des erreurs de saisie, inévitables sur un corpus aussi important et dont le traitement s'est effectué avec des moyens humains, selon Conseil (2017), les corrections apportées qui modifient le score des enquêtés représenteraient moins d'1% du corpus national. Pour qu'une erreur de saisie ou de codage ait un impact, il faudrait qu'une graphie correctement produite ait été saisie ou codée comme non conforme à la norme chez un enquêté dont le score serait proche du seuil de 14 (limite en dessous de laquelle les enquêtés ont été considérés en difficultés graves ou fortes avec l'écrit). Nous verrons en revanche qu'un calcul différent du barème orthographique peut avoir un impact sur le profil des scripteurs.

Néanmoins, un travail de vérification et d'ajout d'informations graphiques nous a semblé indispensable pour pouvoir observer avec précision la variété des graphies produites. Nous

présentons ensuite précisément le travail que nous avons effectué sur le fichier, d'après la version papier des dictées des enquêtes.

2.2.2. VÉRIFICATION DE LA SAISIE, RECODAGE ET CHOIX MÉTHODOLOGIQUES

Avant de réaliser nos analyses, nous avons effectué un travail de vérification du codage déjà réalisé et des graphies relevées pour vérifier la conformité entre la saisie réalisée et les dictées manuscrites⁷⁶. Nous avons recherché les erreurs de saisie, inévitables lorsque les traitements sont effectués manuellement et avons également souhaité discuter en équipe des choix concernant la transcription de certaines graphies. La principale difficulté réside en effet dans l'interprétation des graphies manuscrites puisque dans certains cas, plusieurs lectures étaient possibles. Lorsque nous étions confrontée à ce type de graphies, une concertation en équipe a été organisée avec un travail de comparaison des graphies produites par le scripteur sur le reste de la dictée. Lorsque nous avons considéré la reconnaissance des lettres trop aléatoire, nous avons attribué un code « illisible ».

En parallèle de ce travail de vérification, nous avons rectifié les graphies uniformisées par l'INSEE (et qui n'avaient pas été corrigées par l'ANLCI) :

- retrait du « -s » de TOMATE et CERISE lorsqu'il avait été ajouté ;
- rétablissement de l'accentuation telle que nous avons pu l'observer dans les dictées papiers : *fiancé, fiancè ; epicerie, épicerie, èpicerie, Epicerie, Épicerie* y compris sur les formes non conformes à la norme comme *péi* (qui avait été saisi *pei*) pour PAYS ;
- rétablissement des diacritiques et des auxiliaires d'écriture : *tômate, pharmacie, l'ibrairie, etc. ;*
- matérialisation des segmentations par un espace dans le fichier informatique : *é picerie ;*
- rétablissement des majuscules : *Tomate, toMate ;*
- correction des imputations de réponses en indiquant la présence de non-réponse dans le fichier ;
- attribution d'un code « illisible » pour les graphies dont nous avons considéré la reconnaissance des lettres trop aléatoire et que nous ne souhaitons pas non plus indiquer comme une réponse manquante ;

⁷⁶ Le travail de vérification du fichier, de recodage et d'ajout de données non saisies a été réalisé sur le corpus qui a été utilisé pour notre étude : les 441 dictées des régions Auvergne, Bourgogne, Provence-Alpes Côte d'Azur et Rhône-Alpes.

- codage des métadonnées (organisation de la copie, présentation des items : liste ou texte, etc.) que nous avons souhaité analyser⁷⁷.

LA PRISE EN COMPTE DE L'ACCENTUATION ET DES DIACRITIQUES DANS LE BARÈME

La prise en compte ou non de l'accentuation dans la réussite orthographique a nécessité un traitement différent selon les items, au vu des réponses recueillies.

Dans la dictée, le mot ÉPICERIE apparaît de manière isolée, présenté comme le nom d'un rayon. Il a très souvent été transcrit avec une majuscule à l'initiale non accentuée (*Epicerie*). Par souci de cohérence (avec pour objectif de disposer d'un fichier qui soit le reflet le plus précis possible des dictées papier), nous ne souhaitons pas imputer de « e » accentué (*Epicerie* = *épicerie*) car nous ne pouvons pas savoir si l'enquêté aurait accentué le « e » initial d'ÉPICERIE s'il avait été écrit en minuscule. L'accent était parfois présent mais mal orienté (*èpicerie*) et la réalisation phonique de ce phonème serait donc davantage /ɛ/ que /e/. À ce sujet, Lucci et Millet (1994) indiquent que près de 2/3 des variations qui affectent l'ensemble des diacritiques correspondent à une omission ou à une neutralisation des accents graves et aigus. Les auteurs soulignent une évolution dans la manière de graphier les accents : le plus souvent ils ne sont pas transcrits et lorsqu'ils le sont, ils apparaissent souvent sous la forme d'un trait horizontal (accent plat) ou vertical ; la distinction accent aigu/accent grave tendant à être neutralisée. Nous avons en effet repéré ce phénomène dans les dictées (notamment sur FIANCÉ transcrit *fiance*) et sur certaines graphies, il était souvent difficile de déterminer si l'enquêté avait souhaité réaliser un accent grave ou aigu. Aussi, les accents mal orientés n'ont pas été considérés comme non conformes (*fiancè* accepté).

Sur le mot ÉPICERIE, nous n'avons donc pas considéré comme erronée d'un point de vue orthographique et phonographique, l'absence d'un accent. L'accentuation du « e » initial d'ÉPICERIE a donc été neutralisée dans le barème : pour transcrire le phonème /e/ à l'initiale, nous attendions donc le graphème « e », et avons considéré comme correctes (orthographiquement et phonographiquement) les graphèmes « e », « é », « E », « É » ou les accents plats.

En revanche, nous avons considéré comme erronées d'un point de vue orthographique, les adjonctions d'accents (*épicerie, cerise, rhûme*) ou de diacritiques (*çerises*). Nous avons

⁷⁷ Nous avons envisagé une analyse de tous les aspects liés à la matérialité de l'écriture (utilisation des majuscules, organisation spatiale, etc.) et avons codé en ce sens les 106 dictées du corpus rhônalpin. C'est Claude Ponton, chercheur en TAL au Lidilem et que nous remercions pour ce travail, qui a généré la base de données nécessaire à cette analyse, qui reste à faire.

également considéré comme orthographiquement incorrecte l'adjonction d'accents circonflexes dans des graphies telles que *tômates* ou *fromâges*, même si cela n'affecte pas la phonographie.

Enfin, le rétablissement dans le fichier informatique, de l'accentuation présente sur **FIANCÉ** nous permet désormais de distinguer les graphies *fiance*, des graphies *fiancé*. Nous avons décidé en premier lieu de considérer la graphie *fiance* = /fjãs/ vs /fjãse/ attendue, comme non conforme à la fois orthographiquement, mais aussi phonographiquement. Après analyse des réponses des enquêtés, nous avons finalement choisi de considérer la graphie *fiance* comme conforme orthographiquement. En effet, dans les dictées des enquêtés « bas », 34 graphies apparaissent sous la forme *fiance* et 14 sous la forme **FIANCE** (pour laquelle nous ne pouvons pas savoir si l'enquêté aurait accentué le graphème si le mot avait été écrit en minuscules). Considérer la graphie *fiance* comme orthographiquement non conforme aurait fait baisser le score de réussite orthographique sur cet item. Il nous a semblé que si les enquêtés ont réussi les deux difficultés principales de ce mot : le choix du digramme « an » pour transcrire /ã/ et du graphème « c » pour /s/, la graphie *fiance* ne reflète pas une difficulté d'origine phonographique pour transcrire /s/.

2.2.3. RÉFLEXION MÉTHODOLOGIQUE SUR LA CONSTITUTION D'UN BARÈME

L'intérêt de notre travail de recodage et de saisie réside donc dans cette possibilité nouvelle d'élargir les critères d'évaluation (avec davantage de mots et de phénomènes à étudier) et de pouvoir désormais observer la variation issue de la transcription non conforme des mots et des pseudo-mots.

L'exploitation de ce corpus pose les bases d'une réflexion méthodologique sur la manière de traiter des données produites par des scripteurs dont on sait qu'ils ont du mal à atteindre la norme orthographique. Ainsi, une évaluation classique qui ne consisterait qu'à repérer les graphies incorrectes et à établir des scores de réussite ne nous permettrait pas de rendre compte des capacités des enquêtés à communiquer par écrit ni de définir précisément leurs compétences.

En tenant compte de la spécificité de ce corpus, nous avons adopté un traitement différent de celui appliqué par l'INSEE en proposant d'établir différents scores afin de valoriser les différentes compétences de ce public.

Pour définir les différents barèmes, nous avons tenu compte de plusieurs questionnements :

En premier lieu, cherche-t-on à évaluer la capacité à communiquer par écrit pour un usage courant ou cherche-t-on à mesurer la maîtrise parfaite du système orthographique d'une langue ?

Lorsque l'on interroge des personnes dont on sait que la production d'écrit est une difficulté pour elles, il convient d'observer leurs compétences au-delà de la capacité à respecter l'orthographe. D'autant plus que dans l'enquête IVQ et dans le dépistage de l'illettrisme, nous sommes davantage sur le versant d'un usage littéraire de l'écriture que sur celui de l'utilisation d'une orthographe normée.

Ensuite, de quelles compétences a-t-on besoin pour communiquer par écrit et quels critères prendre en considération pour évaluer ces compétences ?

Si l'on considère que le critère de réussite dans la transmission d'un message relève principalement de sa compréhension pour celui qui le lit, alors la maîtrise du système phonographique du français nous semble être un critère important à prendre en compte.

Enfin, nous pouvons nous interroger sur la frontière entre la maîtrise du système ou d'une partie du système et sa non maîtrise.

Nous souhaitons donc, au-delà du calcul d'un score de réussite orthographique, observer la manière dont des scripteurs en difficulté réagissent lorsqu'ils ne connaissent pas l'orthographe d'un mot. Une analyse des graphies erronées nous permettra de rendre compte de leurs compétences en phonographie, en déterminant si la forme proposée restitue la forme phonologique de l'item. Notre analyse des formes erronées distinguera alors si cette forme est phonographiquement acceptable (et tient compte des règles de position et de distribution relatifs à l'écriture du français) ou si elle représente une transcription phonétique de la forme entendue, dont la lecture n'est pas toujours aisée.

Nous avons également souhaité observer les connaissances des scripteurs dans le domaine de la morphographie en élargissant les critères (au vu du peu de marques de nombre présents dans la dictée). Pour cela, nous évaluons également le comportement des scripteurs face à l'absence de marque sur l'adverbe FORT, sur les marques de nombre souhaitables mais non attendues (car non marquées par un déterminant) que nous nommons marques « implicites » ainsi que sur celles qui ne sont pas souhaitables (par exemple sur le nom des rayons) ou encore les marques ajoutées de manière erronée (exemple. *fiancée* dans LE FIANCÉ SOLENNEL).

Nous avons également souhaité observer les comportements des scripteurs vis-à-vis de la graphie des finales des pseudo-mots (choix d'ajouter ou non une finale muette et choix de la finale muette). Nous regarderons alors si le comportement des scripteurs face aux finales

semble constituer un marqueur de réussite orthographique et/ou l'enregistrement de régularités. Pour cela, nous comparerons les finales choisies avec la fréquence des finales sur les mots du français, pour voir si les choix des scripteurs seraient la marque d'une sensibilité à une régularité du système orthographique du français.

Dans la deuxième partie de ce chapitre, nous présentons nos choix méthodologiques pour décrire de façon quantitative et qualitative la capacité d'écriture des scripteurs à partir de la dictée.

3. ÉVALUER LES CAPACITÉS D'ÉCRITURE D'ADULTES EN DIFFICULTÉ À PARTIR D'UNE DICTÉE : LA NÉCESSITÉ D'UNE ANALYSE PARTICULIÈRE

Nous présentons les critères d'évaluation et les différents barèmes que nous avons appliqués sur la dictée pour valoriser les différentes compétences de scripteurs considérés en difficulté avec l'écrit : connaissance du système orthographique et de la morphographie, connaissance du système phonographique au-delà de la norme orthographique⁷⁸.

3.1. LE CRITÈRE ORTHOGRAPHIQUE APPLIQUÉ AUX MOTS

Dans notre étude, les performances orthographiques des enquêtés ont été observées sur 16 mots de la dictée. Ce critère orthographique, qui vise à vérifier les connaissances lexicales et grammaticales des enquêtés distingue deux scores : un score pour l'orthographe lexicale et un score pour la morphographie.

3.1.1. UN SCORE POUR L'ORTHOGRAPHE LEXICALE

Pour définir les compétences lexicales des enquêtés, nous avons vérifié la conformité de la graphie proposée avec la norme orthographique de 16 mots :

PHARMACIE – ANTI – RHUME – ALCOOL – ÉPICERIE – TOMATES – PAYS – CONFITURE – CERISES – SEL – FROMAGES – SENTENT – FORT – LIBRAIRIE – FIANCÉ – SOLENNEL.

Par rapport aux items corrigés par l'INSEE et pris en compte dans le barème, nous avons ajouté l'item LIBRAIRIE ainsi que FROMAGE(S) et SENTE(NT) dont seules les marques de nombre avaient été prises en compte. Dans ce barème, nous n'avons tenu compte que de l'orthographe lexicale, sans considérer les marques morphographiques (marques de genre ou

⁷⁸ Se référer à l'annexe 3 pour voir le relevé et le classement des graphies des enquêtés « bas ».

de nombre). Lorsque nous avons rencontré des substitutions de mots : par exemple *camembert* à la place de FROMAGES, nous avons codé la réponse comme une absence de réponse.

Dans le tableau 6, nous listons les graphies des mots que nous avons considérées comme conformes d'un point de vue orthographique.

mots	graphies considérées comme orthographiquement conformes
PHARMACIE	<i>pharmacie(s)</i>
ANTI	<i>anti</i>
RHUME	<i>rhume(s)</i>
ALCOOL	<i>alcool(s)</i>
ÉPICERIE	<i>épicerie(s), épicerie(s)</i>
TOMATES	<i>tomate(s)</i>
PAYS	<i>pays</i>
CONFITURE	<i>confiture(s)</i>
CERISES	<i>cerise(s)</i>
SEL	<i>sel(s)</i>
FROMAGES	<i>fromage(s)</i>
SENTENT	<i>sente(s)(nt), sont, semblent</i>
FORT	<i>fort(s)</i>
LIBRAIRIE	<i>librairie(s)</i>
FIANCÉ	<i>fiancé(e)(s), fiancé (s), fiancé(e)(s)</i>
SOLENNEL	<i>solennel(s)(lle)(lles)</i>

Tableau 6. Relevé des graphies considérées comme orthographiquement conformes (16 mots).

Nous présentons ensuite le calcul du score pour la morphographie.

3.1.2. LA PRISE EN COMPTE DE LA MORPHOGRAPHIE

Nous évaluons la morphographie d'après les deux accords en nombre présents dans le syntagme suivant : TROIS FROMAGES QUI NE SENTENT PAS FORT.

Nous attribuons 1 point pour la restitution de chaque marque de nombre attendue. Le score morphographique est donc de deux points.

L'INSEE avait fait le choix de noter sur deux points chacune des deux marques de nombre. Nous estimons que l'attribution de 4 points pour les deux marques de morphographie pèse lourd dans le calcul du score total. Dans notre étude, nous avons choisi de n'attribuer qu'un point par marque de nombre réussie. Les compétences en morphographie font par ailleurs également l'objet d'une analyse indépendante. En rectifiant le barème appliqué sur les deux

items porteurs d'une marque de nombre (1 point pour l'orthographe de l'item et 1 point pour l'accord en nombre, même si la marque proposée ne correspond pas à celle attendue), nous avons établi un barème qui pénalise moins les enquêtés qui ne maîtrisent pas la morphographie.

Nous avons considéré comme acceptables toutes les marques de nombre proposées, y compris celles qui diffèrent de la marque attendue sur le verbe SENTIR (*sentent, sentes, sont*⁷⁹), seule la volonté de marquer le nombre étant observée.

Pour élargir notre vision des compétences en morphographie des enquêtés, nous avons observé les autres marques de genre et de nombre présentes sur certains items (*solennelle, fiancée*, etc.), mais ces réponses analysées de manière qualitative ne rentrent pas dans le score morphographique.

Nous observons ainsi le comportement des enquêtés sur les marques de nombre présentes sur les mots qui ont été dictés sans déterminant. En effet, certains mots dictés pouvaient être écrits soit au pluriel soit au singulier mais ils appellent néanmoins implicitement une marque de nombre : (DES) TOMATES, (DES) CERISES. D'autres en revanche ne demandent pas de pluriel parce qu'ils font référence au nom d'un rayon : (RAYON) PHARMACIE, ÉPICERIE, LIBRAIRIE ou à une quantité : (DE L') ANTI-RHUME, (DE L') ALCOOL, (DE LA) CONFITURE⁸⁰, (DU) SEL.

Nous avons également relevé la présence d'une marque de nombre sur l'adverbe FORT, mais cet item ne rentre pas dans l'évaluation des compétences en morphographie des enquêtés car c'est l'absence de marque qui correspond à une réussite et nous ne pouvons déterminer si ceux qui ont réussi ont eu conscience de ne pas devoir marquer le nombre sur un adverbe. Nous considérons que l'invariabilité des adverbes est une connaissance complexe, tout comme la distinction des homophones grammaticaux (*a* vs *à*⁸¹ dans la dictée) qui requièrent une connaissance déjà avancée du système global et une capacité d'analyse catégorielle importante (Catach, 1980), compétences qui ne nous semblent pas opportunes de juger dans une dictée destinée à évaluer les compétences orthographiques d'un public en difficulté.

⁷⁹ Nous avons accepté les mots « sont » et « semblent » relevés dans les dictées car nous considérons qu'une telle réponse (correcte d'un point de vue syntaxique) pourrait provenir de la manière dont l'enquêteur a dicté le texte.

⁸⁰ On pourrait aussi envisager ici (DES) CONFITURES.

⁸¹ Une analyse de ces homophones que l'on trouve dans l'énoncé ALCOOL À 90° DUX n'a pas été proposée, car le taux de non réponse est particulièrement élevé ; la plupart des enquêtés ayant choisi de ne pas le produire.

Nous avons observé le comportement des scripteurs sur ces différents phénomènes dans une perspective descriptive, mais leurs réponses ne rentrent pas dans le score morphographique.

3.2. LE CRITÈRE PHONOGRAPHIQUE APPLIQUÉ AUX MOTS

Évaluer les compétences à l'écrit de publics en difficulté dans ce domaine soulève des questions qui ne se posent pas lorsque l'on dispose d'écrits produits par des adultes qui n'ont pas de difficulté particulière.

Nous avons vu que l'évaluation classique de l'INSEE qui consiste à repérer les graphies incorrectes et à établir des scores de réussite ne nous permet pas de rendre compte des capacités des enquêtés à communiquer par écrit : en effet, les erreurs orthographiques n'affectent pas nécessairement la qualité du message écrit et la capacité d'écriture ne s'arrête pas aux compétences orthographiques. Par ailleurs, seule une analyse qualitative des réponses erronées permet de définir précisément leur niveau de compétences.

Nous avons donc analysé les réponses erronées des enquêtés et appliqué un barème phonographique sur les mots, valorisant ainsi la capacité des scripteurs à coder les phonèmes en proposant une forme graphique des mots conforme au système phonographique du français lorsqu'ils n'en connaissent pas l'orthographe. Nous considérons, en effet, que des graphies telles que *farmacie* ou *épisserie* n'altèrent pas la capacité communicationnelle de leurs auteurs.

Cette analyse nous permet de distinguer la proportion d'enquêtés capables de restituer la forme phonologique des mots dictés, de ceux pour qui les connaissances phonographiques sont insuffisantes pour passer du canal oral au canal écrit (l'INSEE n'a fait ce travail que sur les pseudo-mots) et ainsi de mieux identifier le niveau de compétences de l'enquêté. Selon Fayol (2013), les compétences phonographiques représentent en effet une étape intermédiaire et essentielle avant l'acquisition des conventions d'écriture.

Nous détaillons ci-dessous les différents codages appliqués pour classer les réponses non conformes au principe orthographique⁸².

3.2.1. LES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES

Le critère phonographique appliqué sur les mots consiste à vérifier la conformité de la correspondance phonie-graphie de la graphie proposée par rapport à la forme phonologique

⁸² Voir l'annexe 3 pour un relevé complet des graphies produites par les enquêtés « bas », classées selon nos critères.

du mot. L'analyse des graphies non conformes montre plusieurs catégories d'erreurs, qui ne modifient cependant pas la forme sonore du mot :

- redoublement de consonnes : *tommatte ;*
- choix de digrammes : *taumat ;*
- accentuation : *frômage ;* diacritiques : *pharmaçi ;*
- absence d'une lettre finale muette (*épisseri, tomat* - ici, le « t » peut en effet parfois se prononcer malgré l'absence de « -e » final comme dans BRUT, OUEST, etc.) ;
- ajout d'une lettre muette (*enthi*) ;
- segmentations non conformes, avec ou sans traits d'union (*en-ti, et pisseri*) ;
- réalisations qui témoignent de marques d'oralité comme l'élision du « e » muet dans ÉPICERIE (*épissrie*) ou l'assimilation du /i/ dans FIANCÉ (*fillancé*).

Bien que ne respectant pas la norme orthographique, nous avons considéré comme acceptables, les réalisations graphiques plausibles de chaque phonème (ex. /o/ : « eau », « au », « o », etc.), en tenant compte des règles de position ou de distribution spécifiques au français.

Nous présentons dans le tableau 7, toutes les réalisations présentes dans notre corpus « bas » qui ont été codées comme des **graphies phonographiquement « acceptables »**.

mots	graphies considérées comme phonographiquement acceptables	
PHARMACIE	[farmasi]	<i>farmaci(e), pharmaçi(e), farmaci, fharmacie, fharmaçie, pharmacy, farhmaci, farmaçie, farmassi.</i>
ANTI	[ãti]	<i>enti(e), antie, en ti, enthi, hanti, anthi, anty, enty, henti(e), anthie, antti, antis.</i>
RHUME	[rym]	<i>rum(e), ruhm(e), hrume, rhùme, rhumme, rumh, rumme.</i>
ALCOOL	[alcɔl]	<i>alcol(e), alccol(e), allcol, alcoll(e), alccool, alcòls, âlcol, àlcoole, al'coole, alkol(e), alcôl, allcole, allcolle, al col.</i>
ÉPICERIE	[episəri]	<i>epiceri, episserie, epiçerie, épisserie, epicery, épicerrie, épiceri, episseris, hépicerie.</i>
TOMATES	[tomat]	<i>tomatte, tommate, thomate, thomat, tomat, tomates, taumatte, tommatte.</i>
PAYS	[pei]	<i>péie.</i>
CONFITURE	[kɔ̃fityR]	<i>comfiture(s), confitur, comfitur, konfitur.</i>
CERISES	[səRiz]	<i>serise, serize, çerise, cerize(s), se rises, ceurises, cerrise.</i>
SEL	[sɛl], [sel]	<i>cel(e), celes, selle, sele, sell, cél.</i>
FROMAGES	[fRomaz]	<i>frommage, fromajé, fraumage, phromaje.</i>
SENTENT	[sãt], [sɔ̃]	<i>sante(s), son(s), cente, santte.</i>
FORT	[fɔR]	<i>for, ford, fore, fors, fohr, faure, for.</i>
LIBRAIRIE	[libReRi]	<i>libreri(e), libréri(e), librai, librèrie, librais, l'hibrerie, librairie, l'ibrarie.</i>
FIANCÉ	[fjãse]	<i>fiancer, fiançais, fience, fiencer, fiencé(e), fiancée, fiançee, fiançai(s), fiançai(e), fiançé(s), phiancée, fiançai, fiançais, fiançe, fiancai, fillancé(e), fillencée, fillansais⁸³.</i>
SOLENNEL	[solanel]	<i>solanel, solanelle(s), solannel, solanele, solannelle, sollanelle, solanèl(e), sollannel, solanéle, saulanel, solennele, solanell, sollanel, sollanell, sollenele, solanélle, saulanelle, sol à nel, solanail.</i>

Tableau 7. Graphies considérées comme phonographiquement acceptables chez les 441 scripteurs « bas ».

⁸³ Nous avons initialement considéré que les graphies *fillancé(e)*, *fillencée*, *fillansais* étaient phonographiquement approximantes. Mais il existe dans le lexique des formes qui se rapprochent de la forme phonologique de ces propositions (comme FILLE). Nous avons alors modifié le classement de ces graphies pour les considérer comme phonographiquement acceptables (les calculs n'ont pas été refaits dans ce sens, mais ces réponses ne concernent que 5 graphies parmi les 441 dictées).

3.2.2. LES GRAPHIES APPROXIMANTES

Certaines graphies ne respectent pas le principe phonographique, parce que le graphème associé au phonème est erroné (« ou » pour /o/ dans *toumate*) ou parce qu'il ne respecte pas les règles de position relatives à l'écriture du français (« s » pour /s/ alors qu'entouré de deux voyelles, il produit le phonème /z/, comme dans *épiseri*). Nous distinguons par une terminologie différente ces deux cas de figures.

Dans la catégorie des graphies « approximantes », nous utilisons le terme de phonographie dans une acception très large, qui consiste à observer les graphèmes ou les groupes de lettres utilisés par les scripteurs pour restituer la chaîne orale. En portant notre analyse sur le graphème, nous avons par exemple admis que le graphème « s » en français puisse avoir les deux valeurs phonologiques /s/ et /z/ (Catach, 1984), quelle que soit la distribution de ce graphème dans le mot (par exemple *épiserie* qui dans ce cas de figure se lirait /epizəRi/). Ce choix nous a amenée à prendre en compte pour cette analyse des graphies qui peuvent, au premier abord, poser problème au lecteur, telles que *rum* considérée comme une transcription de la forme orale /rym/ bien que le graphème « u » dans cette position devant « -m » final se prononce généralement /ɔ/ (*maximum*, *rhum*). Dans ce cas, la graphie a été codée graphie phonographiquement approximante. Enfin, les graphies qui ne restituent pas le « -e » final qui permet la prononciation de la dernière consonne dans CERISE (*ceris*) ont également été classées parmi les graphies approximantes. Pour le /z/ en finale de CERISE, nous sommes davantage dans une transcription phonétique qui accepte pour réalisation sonore, toutes les réalisations possibles du graphème « s » (dont /z/, même si ici le contexte vocalique n'est pas réuni).

Ces graphies qui approchent la forme phonographique attendue reflètent chez le scripteur des difficultés avec l'écrit de l'ordre de la connaissance de règles de fonctionnement inhérentes au français. Dans un but de communication, la lecture des graphies *rum* ou *épiserie*, qui relèvent davantage d'une transcription phonétique, provoquerait une gêne plus ou moins importante selon les cas, qui serait également fonction du niveau d'expertise du lecteur/scripteur.

Cette catégorie nous permet néanmoins de repérer les enquêtés à qui il manque une connaissance des règles de fonctionnement du français plus fine mais qui ont une connaissance certes partielle du système, mais existante.

Le choix de considérer une graphie comme approximante a souvent engendré des discussions au cas par cas, qui nous ont amenée à prendre des décisions (pour les analyses quantitatives), que nous considérons néanmoins comme discutables, la frontière entre ce qui peut être

acceptable et ce qui ne l'est pas étant parfois très mince. Nous illustrons notre propos en observant les graphies d'ALCOOL que nous avons classées comme approximantes. Nous indiquons entre parenthèses le nombre de graphies observées dans notre corpus :

- *alcohol* (2 occ.) : l'insertion d'une lettre muette entre le redoublement de voyelle a certes l'effet de produire deux fois le phonème /ɔ/, mais la présence du « h » muet témoigne néanmoins d'un recours à la voie d'adressage. La réalisation phonique /ɔ/ ou /o/ pour « oo » est en effet peu fréquente en français : zoo (deux graphèmes « o » avec leur réalisation phonique distincte), igloo (« oo » qui transcrit /ũ/) ;
- *alchol* (1 occ.) : « ch » pour « k » que l'on peut trouver dans ALMANACH, VARECH ;
- *alecoole* (1 occ.) : l'ajout du graphème « e » après le graphème « l » dont la réalisation peut être muette dans un autre environnement graphique ;
- *aalcool* (1 occ.) : dont le redoublement du premier phonème peut avoir été influencé par le redoublement de « o ». Même si le redoublement de « a » n'existe pas en français, cette erreur n'empêche pas de retrouver la graphie ALCOOL.

Nous présentons, dans le tableau 8, les formes graphiques rencontrées dans notre corpus « bas » qui ont été codées **graphies phonographiquement « approximantes »**. Nous avons retiré du tableau les items pour lesquels aucune des réponses des enquêtés n'a été considérée « approximante ».

mots	graphies considérées comme phonographiquement approximantes	
ANTI	[ãti]	<i>aenti.</i>
RHUME	[rym]	<i>rhum, reume, reum, rhûm, rheum, rueme.</i>
ALCOOL	[alcɔl]	<i>alcohol, alchol, alecoole, aalcool.</i>
ÉPICERIE	[episəri]	<i>episeri(e), et piseri(e), épiecrie, epicri(e), épiserie, épisrie, épicserie.</i>
PAYS	[pei] [peji]	<i>pei, payi(e), peyi, peie, pays, péyi.</i>
CONFITURE	[kɔ̃fityR]	<i>confiteure.</i>
CERISES	[səRiz]	<i>seriz, seris, ceriz, criz, ceris.</i>
SENTENT	[sãt], [sɔ̃]	<i>sent(s), sant, scent, cent.</i>
FIANCÉ	[fjãse], [fijãse]	<i>fianse, fiancei, fiyanse, fiensé, fiense,</i>
SOLENNEL	[solanɛl]	<i>solanaille, colanelle</i>

Tableau 8. Graphies considérées comme phonographiquement approximantes chez les 441 scripteurs « bas ».

Lorsque le graphème utilisé ne permet jamais d'associer le phonème souhaité, par exemple « p » pour transcrire le phonème /t/, alors la graphie a été codée **graphie erronée**.

3.2.3. RÉCAPITULATIF : LE CODAGE DES GRAPHIES QUI S'ÉCARTENT DE LA NORME

Dans cette analyse, nous proposons ainsi de classer les réponses des enquêtés sur les mots de la dictée en quatre catégories :

- **orthographe correcte** pour les graphies qui respectent la norme orthographique (*tomate(s)* pour TOMATES) ;
- **phonographie acceptable** pour les graphies qui ne respectent pas la norme orthographique mais qui sont conformes au principe phonographique (*épisserie* pour ÉPICERIE) ;
- **phonographie approximante** pour les graphies qui relèvent davantage d'une transcription phonétique parce qu'elles ne sont pas conformes au principe phonographique (leur réalisation ne tenant pas compte des règles de position et de distribution relatives à l'écriture du français) comme *épiserie* pour ÉPICERIE ;
- les **graphies erronées** pour les mots dont la forme graphique proposée ne peut (en aucune manière) restituer la forme phonologique du mot (*toumate* pour TOMATES).

Ce classement nous permet d'observer différents niveaux de compétences que nous estimons important de qualifier lorsqu'il s'agit d'analyser des graphies produites par des scripteurs en difficulté avec l'écrit. Cette analyse des graphies qui vise à valoriser aussi les compétences phonographiques dans la production de mots écrits, nous permettra de nous interroger sur l'impact du critère phonographique dans la prise en compte des difficultés à l'écrit. Enfin, les différents scores appliqués sur les dictées nous permettront de définir un profil de scripteur pour chaque enquêté.

3.3. LE CRITÈRE PHONOGRAPHIQUE APPLIQUÉ AUX PSEUDO-MOTS

Les pseudo-mots ont été comptabilisés selon leur conformité au principe phonographique. Ne disposant pas de données concernant la manière dont les items ont été dictés et notamment sur une éventuelle particularité de prononciation des enquêteurs, nous avons comptabilisé comme correctes des formes graphiques qui peuvent différer de la forme sonore attendue mais qui peuvent s'expliquer par la prononciation inhérente à l'enquêteur ou à l'enquêté. C'est ainsi que nous avons pris en considération toutes les graphies des voyelles antérieures /æ/ et /ə/ dont le point d'articulation est proche, et de la voyelle /ɔ/, plus postérieure, ces trois voyelles formant une série au sein d'un système vocalique particulièrement riche :

- *micateul* pour MICATOL ;
- *jek, jeuq* pour JOC.

Ces graphies ont été comptabilisées comme phonographiquement correctes. Par ailleurs, même si une forme graphique est plus probable qu'une autre en français (ex. *joque* plutôt que *jok*, *cirape* plutôt que *cirap*), les deux ont été comptabilisées comme phonographiquement correctes, aucune orthographe ne prédominant pour la graphie de pseudo-mots.

Nous présentons, dans le tableau 9, les formes que nous avons retenues comme phonographiquement acceptables dans notre corpus « bas ».

pseudo-mots	graphies phonographiquement acceptables
MICATOL	<i>micatol(e), mikatol(e), micathol, micatolle, mycatol(e), miquatol(e), mycathol(e), micatôle, mickatol(e), miccatol, mykatole, micatoll, miquatoll, mi ka tol, mikathol, mykattol, mi catol, mycatoll, miqatole, micateül, miekatele.</i>
SIRAP	<i>sirap(e), cirap(e), cyrape, syrap(e), sirappe, cyrap, çirape, cirappe, cirrap(e), si rape, cyrappe, siraps, sirrape.</i>
DUX	<i>dux(e), ducse, dukse, dusc(e), dûx, duxce, duçce, d'uxe, duxs, duckse, dukce, ducsse.</i>
GOBAR	<i>gobar(e), gobard, gobart, gaubar, gaubard, gobarre, gaubart, guobar, gobbar, gaubarre, guobard, guobare.</i>
JOC	<i>Jock€, joc, jok€, joque(s), jocque, jauque(s), jocques, jocq, joq€, joqu, jauc, joch, jokc, joks, jauke, jauq, geoc.</i>
PÉVANOR	<i>pevanor(e), pévanor(e), pevanord, pévanord, pévanort, pevanort, pèvanor(e), pevanores, pévannor, pévanaure, paivanore, p.vanor, pevanaure, pevanorr, pévanorre, paivanor, pévanaur, pe vanore, p.evanor, pevannor, p vanor, petvanor, pey va nor, pevanorre, pevanors, pe va nore, pevaneur.</i>

Tableau 9. Graphies des pseudo-mots considérées comme phonographiquement acceptables chez les 441 scripteurs « bas ».

Dans la partie suivante, nous présentons le protocole de l'enquête complémentaire, composé de la dictée d'IVQ augmentée du « rayon bricolage », d'un questionnaire et d'un entretien.

4. LE PROJET « COMPÉTENCES ORTHOGRAPHIQUES DES ADULTES ET CONTEXTES SOCIAUX D'USAGE »

4.1. OBJECTIFS ET INTÉRÊT DU PROJET POUR L'ANALYSE DE LA DICTÉE D'IVQ

L'enquête complémentaire a pour objectif de mieux interpréter les résultats de la dictée produite par les enquêtés repérés en difficulté avec l'écrit (dans IVQ 2011) et d'évaluer leur particularité en production écrite, en faisant passer cette même dictée à des enquêtés de 2011 qui n'avaient pas été interrogés en production d'écrit puis à comparer le résultat des deux populations.

En 2011, suite à l'enquête IVQ, l'INSEE avait proposé aux enquêtés qui n'avaient pas été considérés en difficulté avec l'écrit de participer à une extension de l'enquête IVQ. Nous avons eu accès en 2014 aux fichiers adresses de ceux qui avaient accepté, ce qui nous a permis de faire passer la dictée d'IVQ à des enquêtés qui n'avaient pas été évalués en production d'écrit et de disposer ainsi de deux échantillon de dictées qui oppose deux catégories de personnes. L'une considérée en difficulté suite au module « d'orientation » d'IVQ (et qui a passé la dictée en 2011) et l'autre qui n'avait pas été considérée en difficulté suite à leurs résultats à ce même module « d'orientation » (et à qui nous avons fait passer la dictée d'IVQ en 2014).

Ce second recueil de données a été rendu possible grâce à un projet complémentaire intitulé « Compétences orthographiques des adultes et contextes sociaux d'usage »⁸⁴, à l'initiative de J-P. Jeantheau (ANLCI). Mené dans les régions Rhône-Alpes et Haute-Normandie, ce projet a reçu le soutien financier de la DGLFLF (Délégation générale à la langue française et aux langues de France) ainsi que l'approbation du comité du label de la Statistique Publique et du Conseil National de l'Information Statistique⁸⁵.

Les enquêtés dirigés vers le module « haut » suite au module d'orientation avaient été dispensés de dictée. Comme le souligne Jeantheau (2010), conclure que ces derniers auraient obtenu 100% de réussite à l'exercice de production écrite (puisque'ils n'ont pas montré de difficulté particulière à l'écrit) serait une extrapolation un peu hâtive, notamment parce que certains items sont peu fréquents et/ou peu consistants. Nous pourrions en revanche penser qu'ils auraient obtenu un score compris entre celui obtenu par les personnes en difficulté et 100%.

Cette post-enquête représente la possibilité de tester les compétences à l'écrit de personnes dites « sans difficulté particulière » sur les mêmes critères que ceux qui ont servi à caractériser les écarts orthographiques et les lacunes en phonographie d'un public identifié comme probablement peu performant en matière d'écriture. Cela nous permet surtout de disposer d'un corpus de comparaison non plus seulement vis-à-vis de la norme mais également de l'usage. Nous pourrions alors porter un regard différent quant aux attentes en matière d'écriture, notamment lorsque les difficultés observées dans les écarts orthographiques sont partagées autant par les scripteurs ordinaires que par les autres.

⁸⁴ Se référer à l'annexe 4 pour la présentation de ce projet.

⁸⁵ Se référer aux annexes 5 et 6 pour voir ces courriers.

Cette post-enquête représente également une opportunité pour nous, de tester des personnes dites ordinaires sur d'autres difficultés orthographiques que celles de la dictée de l'INSEE et de cerner leurs pratiques d'écriture et leurs représentations de l'orthographe.

4.2. UNE ENQUÊTE DE TERRAIN

4.2.1. MISE EN PLACE DE L'ENQUÊTE ET BILAN

Afin que nous puissions réaliser l'enquête, l'INSEE nous a fourni les fiches adresses des 311 enquêtés rhônalpins⁸⁶ qui avaient donné leur accord en 2011 pour être réinterrogés (adresse du foyer en 2011, nom et année de naissance) : nous en avons retenu 250 sur la base du critère géographique. En effet, certaines adresses isolées, notamment dans des zones de montagne, ne permettaient pas de mener l'enquête dans de bonnes conditions.

Trois ans après l'enquête réalisée par l'INSEE, cette post-enquête a été particulièrement difficile à mettre en œuvre. Sur les 250 adresses de départ, nous avons retrouvé 184 personnes dans 42 communes des 8 départements de la région Rhône-Alpes, les 66 autres ayant déménagé (les pertes sont très sensibles dans les grandes agglomérations : à Lyon par exemple, 13 enquêtés sur les 26 interrogés en 2011 et ayant accepté d'être réinterrogés n'ont pas été retrouvés en 2014). Plusieurs passages au domicile⁸⁷ ont souvent été nécessaires avant de procéder à l'entretien, notamment pour la prise de rendez-vous qui a rarement été possible par téléphone car nous ne disposions pas des coordonnées téléphoniques des enquêtés de 2011. Bien que ceux-ci aient tous accepté en 2011 d'être réinterrogés, 53 personnes présentes à leur domicile ont refusé de participer à cette extension de l'enquête IVQ, pour des motifs variés : enquête non obligatoire (à la différence des enquêtes d'utilité publique), manque de temps à consacrer à une enquête, manque d'intérêt pour ce type d'enquête, etc. Enfin, 44 personnes n'ont jamais été trouvées lors des trois déplacements à leur domicile, n'ayant parfois jamais répondu aux différentes sollicitations (courriers, avis de passage dans les boîtes aux lettres, proposition de rendez-vous) ou n'étant pas disponibles dans les créneaux horaires proposés (travailleurs de nuit, personnes en déplacement professionnel la semaine, indisponibilités diverses). Dans certains cas, la frontière entre l'absentéisme et le refus de participer à l'enquête était très mince.

⁸⁶ En Haute-Normandie, pour le laboratoire DySoLa de Rouen, Clara Mortamet et Jeanne Conseil ont également obtenu les fiches adresses des enquêtés Haut-Normands, ce qui permis de recueillir pour cette région 109 dictées supplémentaires.

⁸⁷ Se référer à l'annexe 7 pour voir le courrier adressé aux enquêtés que nous souhaitons réinterroger.

Finalement, 28% des enquêtés retrouvés ont refusé de participer à cette nouvelle enquête et 27% n'étaient pas disponibles lors du recueil. Ce sont donc 87 personnes qui ont pu, en 2014, être interrogées dans 42 communes des 8 départements de la région Rhône-Alpes.

Pour mener les entretiens dans des lieux parfois très éloignés les uns des autres, C. Brissaud et C. Mortamet ont fait appel à une société privée⁸⁸. Ont également participé au recueil des données sur le terrain : C. Mortamet et J. Conseil du laboratoire *DySoLa* (pour l'enquête Haut Normandie) ainsi que J-P. Jeantheau de l'ANLCl, C. Brissaud et nous-même du laboratoire LIDILEM (pour l'enquête Rhônalpine) Pour la réalisation de l'enquête, un guide à l'attention des enquêteurs a été rédigé (qui décrit l'objectif de l'enquête, les consignes et le protocole précis) et une journée de formation a été organisée pour les enquêteurs.

4.2.2. LE PROTOCOLE DE L'ENQUÊTE COMPLÉMENTAIRE

Le protocole a été conçu au sein du laboratoire LIDILEM à Grenoble et du laboratoire DYSOLA à Rouen⁸⁹. Il est constitué d'une partie de production écrite (la dictée d'IVQ⁹⁰ à laquelle une vingtaine de mots a été rajoutée), d'un questionnaire qui porte sur les pratiques d'écriture à la maison et au travail, et d'un entretien semi-directif qui interroge les enquêtés sur leurs souvenirs d'apprentissage de l'orthographe, sur leur utilisation de l'écrit (notamment électronique) et sur leur rapport à l'écrit et à l'orthographe.

Dans les enquêtes destinées à évaluer les adultes, Degorre et Murat (2009) soulignent la difficulté à obtenir des déclarations de personnes qui craignent d'être stigmatisées avec pour limite, une tendance à minimiser leurs difficultés. Nous pourrions donc penser qu'interroger des personnes, qui d'après les résultats d'IVQ, ne présentent pas de difficulté à l'écrit lèverait cette difficulté. En réalité, l'enquête complémentaire montre que même chez ce public, la question de l'orthographe est également sensible et la crainte du jugement parfois tout autant présente.

Aussi, instaurer un climat de confiance s'est révélé tout aussi important (que lorsque l'on interroge des personnes en difficulté) pour mettre en place un cadre favorable au recueil d'informations et pour inciter les confidences.

⁸⁸ C'est l'entreprise CDA (dirigée par Catherine Delannoy & Associés) qui a été recrutée pour cette mission.

⁸⁹ Se référer à l'annexe 8 pour consulter le protocole de l'enquête complémentaire.

⁹⁰ Voir annexe 2, un exemple de dictée produite par un enquêté de cet échantillon de scripteurs qui n'ont pas été considérés en difficulté avec l'écrit dans IVQ 2011 (enquêté « haut »).

4.2.2.1 INTRODUCTION DE L'ENQUÊTE

La partie introductive de l'enquête revêt une importance capitale pour le déroulement de l'enquête, tout comme la posture de l'enquêteur tout au long du test et des entretiens. Nous avons en effet ressenti une certaine méfiance chez quelques enquêtés pouvant même prendre la forme d'une attitude agressive qui généralement s'est estompée au fur et à mesure de l'entretien.

Pour prendre contact avec l'enquêté, nous avons commencé par présenter l'objectif de notre venue et avons cherché à impliquer la personne en lui expliquant l'importance et la contribution que sa participation apporte à la recherche. Nous avons également souligné qu'ils avaient été sélectionnés parce qu'ils ont obtenu de bons résultats lors de l'enquête de 2011, ce qui permet de valoriser leurs compétences et d'instaurer un climat plus propice à des discussions autour de la question de l'orthographe. Certains étaient par ailleurs surpris d'avoir « réussi » les tests en 2011 et d'avoir été classés parmi les personnes « sans difficulté avec l'écrit ».

Pour aider les enquêteurs, cette phase d'approche a été standardisée et un texte a été préparé. Il était spécifié aux enquêteurs que ce texte n'était pas destiné à être répété textuellement, mais que ce sont les idées, et notamment les mots employés qui sont importants. Il a également été indiqué de préférer le terme de « questionnaire » comme dans une enquête classique et de ne jamais parler d'évaluation, d'épreuve ou de dictée, c'est-à-dire de tout ce qui peut rappeler une situation d'évaluation scolaire :

« C'est une enquête du laboratoire LIDILEM de l'Université de Grenoble qui fait suite à l'enquête sur l'information dans la vie quotidienne qui vous a été proposée par l'INSEE en 2011.

C'est une enquête qui vise à mieux connaître les capacités à l'écrit des personnes qui n'ont pas de difficultés particulières dans ce domaine.

Le test que vous allez passer a été proposé en 2011 aux personnes les plus en difficultés et nous souhaitons aujourd'hui proposer ce test aux personnes qui ne sont pas en difficultés pour :

- mieux évaluer la particularité dans ce domaine des personnes jugées en difficulté à l'écrit ; par comparaison avec les productions de personnes qui ne sont pas en difficultés

- évaluer la pertinence de la dictée pour mesurer les compétences en orthographe des personnes en difficulté à l'écrit, pour que celle-ci soit reprise dans de futures enquêtes.

Le but est de permettre ensuite de mieux agir dans le domaine de la formation des adultes. Nous avons vraiment besoin de vous pour cela. »

Il a été précisé aux enquêteurs que certains enquêtés peuvent avoir besoin d'éclaircissements sur les objectifs de l'enquête. Un texte a été préparé pour cela :

« La vie quotidienne et la vie professionnelle deviennent de plus en plus complexe. Quelles sont les tâches que l'on doit accomplir dans la vie quotidienne ? Quelles sont les difficultés que l'on rencontre lorsque l'on doit se repérer dans des modes d'emploi, des journaux, etc. Tous ces supports donnent beaucoup d'informations : Est-ce que ces informations sont adaptées ?

Pour l'utilisation des résultats :

« Ils serviront à orienter plus efficacement l'action du système de formation des adultes (la formation tout au long de la vie prend de plus en plus d'importance). »

« Ils serviront à rendre un peu plus concrets les savoirs que l'on enseigne, à les adapter à la vie quotidienne bref, de rajouter du concret. »

4.2.2.2 LE RAYON « BRICOLAGE »

L'enquête débute par la dictée. Nous avons utilisé le même mode de recueil que celui mis en place dans IVQ : les enquêtés ont rédigé sur papier libre, la liste de courses dictée par l'enquêteur. Celui-ci pouvait répéter les items autant que nécessaire ; les enquêtés disposant également d'un temps de relecture.

La consigne était la suivante :

« À présent, je vais vous demander d'écrire une liste de courses à remettre à un livreur de supermarché. Dans cette liste, il faut écrire le nom du rayon et les produits dont vous avez besoin. Vous allez entendre des mots que vous ne connaissez pas, ce sont des noms de marques que nous avons dû inventer car nous n'avons pas le droit de faire de la publicité. Écrivez-les comme vous pensez que cela doit s'écrire. Il n'y a pas de contrainte de temps, vous pouvez prendre votre temps. Je commence à vous lire les mots. »

La production écrite a été augmentée d'un « rayon bricolage », ajouté à la fin de la dictée d'IVQ. Les items du « rayon bricolage » ont été choisis pour présenter des difficultés lexicales et morphographiques, en nombre plus important que celles de la dictée d'IVQ. La thématique volontairement plus masculine est destinée à équilibrer l'intérêt de la dictée autant pour les femmes que pour les hommes.

(RAYON) BRICOLAGE

- DEUX PRODUITS PARFUMÉS QUI DÉGRAISSENT BIEN,
- QUATRE PAQUETS DE VIS PLATES,
- UN CISEAU À BOIS NORMAL,
- UNE BOITE D'ALLUMETTES,
- UNE PINCE POUR ARRACHER LES CLOUS,
- UN POT DE PEINTURE,
- QUELQUES MÈTRES DE TUYAU D'ARROSAGE.

Nous présentons ensuite le classement des mots de la dictée d'IVQ et des mots du « rayon bricolage » selon leur indice de consistance calculé dans Manulex-Infra.

CLASSEMENT DES MOTS DE LA DICTÉE D'IVQ ET DU RAYON BRICOLAGE SELON LEUR INDICE DE CONSISTANCE CALCULÉ DANS LA BASE MANULEX-INFRA

Pour rendre compte du degré de difficulté de la dictée, nous présentons dans le tableau 10, le classement des mots de la dictée d'IVQ et du « rayon bricolage » selon leur indice de consistance (1 étant le mot qui apparaît comme le plus consistant de la dictée) calculés dans la base Manulex-Infra. Notons que les items ANTI (dictée d'IVQ) et DÉGRAISSER y compris ses dérivés (« rayon bricolage ») sont absents de Manulex-Infra.

Les calculs sont effectués « par token »⁹¹, sur le mot entier (moyenne des différentes associations dans le mot) et à partir de tous les manuels de la base (moyenne du CP au CM2).

rayon bricolage	CPO	rang	dictée d'IVQ	CPO	rang
NORMAL	299.46	1	SOLENNEL	86.64	4
PRODUIT	130.07	2	TOMATE	82.99	5
VIS	113.38	3	FROMAGE	81.72	7
BRICOLAGE	82.01	6	SENTIR	81.04	9
PARFUMÉ	81.17	8	SEL	80.52	10
PLATE	79.75	11	PAYS	79.25	13
BOITE	79.51	12	FORT	78.31	16
BOIS	79.25	14	LIBRAIRIE	77.63	17
MÈTRE	79.07	15	CONFITURE	76.01	19
CLOU	77.38	18	CERISE	72.03	20
CISEAU	71.55	21	PHARMACIE	67.17	23
PEINTURE	70.01	22	ÉPICERIE	64.86	25
ARRACHER	65.68	24	RHUME	61.74	27
PAQUET	64.51	26	ALCOOL	61.47	28
ARROSAGE	59.54	29	FIANCÉ	59.19	30
POT	56.66	31			
ALLUMETTE	52.30	32			
TUYAU	51.52	33			
PINCE	48.21	34			

Tableau 10. Classement des mots de la dictée d'IVQ et du « rayon bricolage » selon leur degré de consistance (consistance PO totale, du CP au CM2, calcul par token) calculé dans Manulex-Infra-Infra.

Le classement des mots proposés aux enquêtés interrogés dans le cadre de l'enquête complémentaire selon leur indice de consistance montre que le rayon bricolage propose davantage de difficultés lexicales, avec 5 mots qui ont un indice de consistance inférieur à 60 : ARROSAGE, POT, ALLUMETTE, TUYAU, PINCE alors que dans la dictée d'IVQ seul FIANCÉ a un indice inférieur à 60. Le « rayon bricolage », comme dans la dictée d'IVQ possède également de nombreux mots consistants comme BRICOLAGE, VIS, PRODUIT, NORMAL.

⁹¹ Les calculs "par token" prennent en compte le nombre d'occurrences des mots dans les textes (c'est-à-dire leur fréquence réelle), ce qui permet de prendre en considération la rareté d'un mot. Sans cela, un mot qui apparaît fréquemment dans un texte aurait le même poids qu'un mot qui apparaît plus rarement.

CLASSEMENT DES MOTS DU RAYON BRICOLAGE SELON LEUR INDICE DE FRÉQUENCE CALCULÉ DANS LA BASE LEXIQUE 3

Dans le tableau 11, nous présentons l'indice de fréquence U^{92} des mots de la dictée complémentaire (« rayon bricolage ») calculé dans la base Lexique 3. Les mots sont présentés sous forme de lemme et avec la forme fléchie pour CISEAUX. Le rang indique le classement du mot par rapport aux autres mots de la dictée, selon l'indice de fréquence calculé dans la base Lexique (1 étant le mot qui apparaît comme le plus fréquent).

Nous utiliserons les indices de la base « livres » comme indices de référence : aussi, ceux-ci sont classés par ordre décroissant, du plus fréquent au moins fréquent, pour faciliter la lecture.

⁹² Pour rappel, U représente la fréquence estimée d'usage. Cet indice prend en compte la dispersion des mots dans le corpus ainsi que leur fréquence réelle, ce qui permet de disposer d'un indicateur plus fiable de la fréquence des mots (pour davantage d'explications, se référer au chap.3, section 1.2).

mots	U « livres »	rang	U « films »	rang
BOIS	299.46	1	115.56	2
MÈTRE	130.07	2	43.49	5
ARRACHER	113.38	3	54.19	3
PAQUET	89.19	4	44.96	4
PEINTURE	67.64	5	29.53	7
NORMAL	62.77	6	127.15	1
PLAT	61.76	7	14.66	12
POT	48.04	8	29.89	6
CLOU	27.50	9	13.83	13
PRODUIT	23.99	10	28.80	8
ALLUMETTE	20.88	11	15.65	10
TUYAU	20.61	12	26.73	9
GRAISSER	18.31	13	9.03	14
PINCE	14.73	14	9	15
CISEAU	13.85	15	8.72	16
CISEAUX	11.01		7.86	
PARFUMÉ	10.20	16	2.24	17
BOITE	7.77	17	15.43	11
ARROSAGE	1.89	18	1.72	18
BRICOLAGE	1.76	19	0.89	19
DÉGRAISSER	0.41	20	0.30	20

tableau 11. Classement des mots du « rayon bricolage » selon leur indice de fréquence calculé dans la base lexicale 3.

Comme nous l'avons observé avec les mots de la dictée d'IVQ, le classement par ordre de fréquence selon les indices calculés dans la base « livres » ou « films » sont plutôt proches. Nous observons un décalage sur les items NORMAL (qui apparaît comme le plus fréquent dans la base « films » mais à la 6^e place dans la base « livres ») et PLAT (qui est classé 7^e dans la base « livres » mais 12^e dans la base « films »). Ces items étant consistants (95⁹³ pour NORMAL et 79 pour PLATE), la différence d'indice de fréquence n'aura que peu d'impact sur leur réussite orthographique. Aussi, chez ce public, que nous pouvons supposer lecteur, nous observerons les indices de fréquence proposés par la base Lexique « livres ».

Les mots rencontrés les plus fréquemment dans les livres sont : BOIS, MÈTRES, ARRACHER (avec un indice U compris entre 299 et 113). L'item ARRACHER est en revanche de consistance

⁹³ Dans notre texte, pour plus de lisibilité, nous ne conservons pas les chiffres après la virgule sur les indices de consistance et de fréquence (nous ne les avons pas arrondis), sauf lorsque les chiffres après la virgule permettent de distinguer deux mots : par exemple .ARROSAGE (U « livres » = 1,89) et BRICOLAGE (U « livres » = 1.76).

moyenne (CPO = 65, rang 24/33). D'autres sont très peu fréquents : ARROSAGE, BRICOLAGE, DÉGRAISSE (indice U entre 1.89 et 0.41).

Les quatre mots les moins consistants du rayon bricolage sont de fréquence moyenne dans la base « livres » (et de la dictée d'IVQ) : POT (CPO = 56 ; U = 48, rang 8), ALLUMETTE (CPO = 52 ; U = 20.88, rang 11), TUYAU (CPO = 51.52 ; U = 20.61, rang 12), PINCE (CPO = 48 ; U = 14, rang 14). Nous observerons plus particulièrement le comportement des scripteurs sur ces mots, pour rendre compte de l'impact de la fréquence sur la restitution orthographique de mots peu consistants (à la différence des items SOLENNEL et ARROSAGE qui sont à la fois peu consistants et peu fréquents).

LES MOTS DU « RAYON BRICOLAGE » PRÉSENTÉS SELON LEUR SCORE DANS EOLE

Nous présentons dans le tableau 12, les scores donnés dans EOLE pour les mots du « rayon bricolage » (dictée complémentaire). Certains mots de nos dictées n'y apparaissent pas car ils ne font pas partie des 12000 termes qui ont été retenus dans EOLE pour être les plus courants du français écrit des années 2000.

Nous avons ajouté des remarques sur la forme attendue dans la dictée, qui ne testait pas toujours uniquement la forme lexicale.

mots du « rayon bricolage »	CP	CE1	CE2	CM1	CM2
ALLUMETTE	4	30	67	76	78
ARRACHER	0	68	44	55	67
ARROSAGE	0	0	27	38	41
BIEN ⁹⁴	61	83	98	100	84
BOIS	48	60	89	93	100
BOITE ⁹⁵	59	51	50	76	82
BRICOLAGE	50	72	83	79	97
CISEAU	7	31	49	69	72
CLOU	61	56	89	95	85
DÉGRAISSER	0	9	33	34	55
DEUX	74	97	100	100	100
METRES	38	50	67	70	62
NORMAL	51	86	91	95	90
PAQUET	11	53	33	66	65
PARFUMER	78	79	84	90	95
PEINTURE	10	45	87	90	87
PINCE	53	62	82	93	91
PLATE	75	76	74	85	91
POT	15	72	81	90	90
PRODUIT	6	38	89	75	97
QUATRE	72	95	90	96	97
QUELQUE	21	27	75	53	84
TUYAU	0	8	28	53	59

Tableau 12. Pourcentages de réussite de mots du "rayon bricolage" d'après EOLE.

Les données issues d'EOLE complètent les données sur la fréquence et la consistance orthographique qui sont fournies par les bases Manulex-Infra (Lété, Sprenger-Charolles, Colé, 2004) et Lexique. Il améliore notamment ces outils qui permettent de prédire la réussite ou la difficulté mais qui ne fournissent pas en revanche de données sur les productions effectives des élèves. Il reste néanmoins limité pour une analyse plus fine des réussites et des difficultés

⁹⁴ Sur ce mot, la catégorie grammaticale du mot utilisé dans la dictée n'est pas précisée : nom, adjectif ou adverbe.

⁹⁵ C'est l'item BÔÎTE qui est indiqué dans EOLE et il n'est pas spécifié si la réponse attendue pouvait également posséder un « i » sans accent circonflexe, comme le recommandent les rectifications de 90.

orthographiques parce qu'il n'intègre pas les mots-outils et ne tient pas compte de manière systématique des classes grammaticales et des formes fléchies.

Le fait de disposer de données sur les productions reste un atout majeur car les indices de réussite montrent que l'acquisition est en règle générale linéaire. Le nombre de termes atteignant un score de 75 % de réussite est de 347 au CP, 1013 au CE1 ; 2188 au CE2, 3606 au CM1, 5054 au CM2). En revanche, les nouveaux apprentissages orthographiques introduisent aussi des doutes et des erreurs qui n'apparaissent pas auparavant, que nous observons dans les mots des dictées, avec ARRACHER, PAQUET et VIS.

Pothier et Pothier (2003) estiment que les mots qui atteignent 75% de réussite peuvent être considérés comme acquis par l'ensemble d'une classe d'âge (ou comme devant être acquis pour ceux qui n'en maîtrisent l'orthographe).

Nous observons à présent dans la figure 6, les taux de réussite d'élèves de CM2, sur les mots du « rayon bricolage ». Comme dans la figure précédente, les mots qui sont réussis à plus de 75% sont indiqués en gris clair, ceux qui sont réussis à moins de 75% sont en gris foncé.

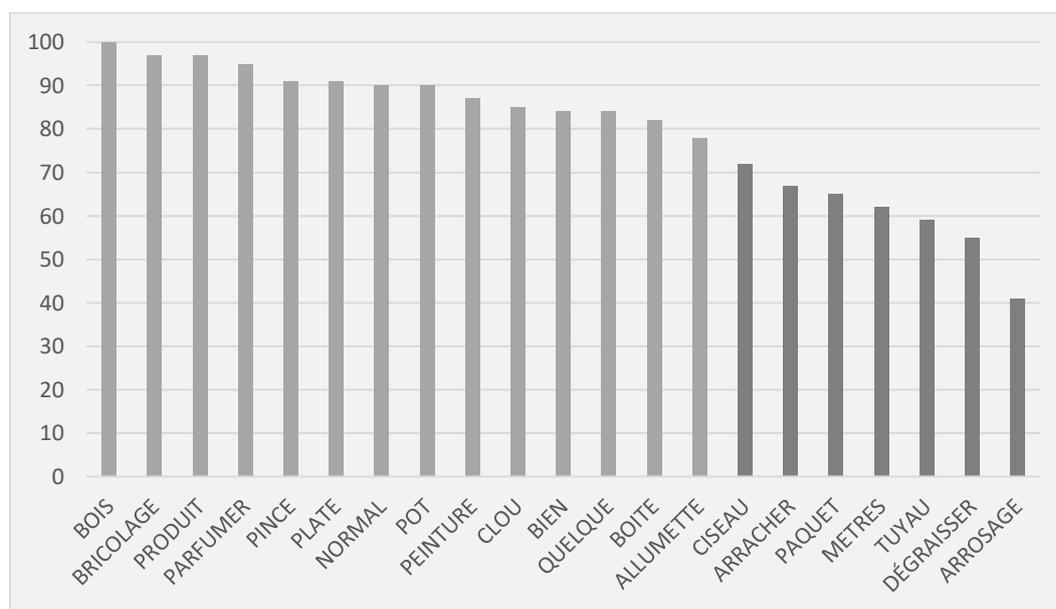


Figure 6. Classement des mots du « rayon bricolage » selon leur pourcentage de réussite dans EOLE, chez des élèves de CM2.

Sur les mots du « rayon bricolage », nous observons également de très bons taux de réussite chez des élèves de CM2 sur certains mots : BOIS, BRICOLAGE, PRODUIT, PARFUMER. En revanche, 7 mots obtiennent moins de 75% de restitution orthographique mais ils ont tous, sauf ARROSAGE (41%), des taux de réussite supérieurs à 50%. Il apparaît en premier lieu que le lexique utilisé dans la dictée d'IVQ poserait davantage de problèmes que celui du « rayon bricolage » s'il était dicté à des élèves de CM2 (avec davantage de mots qui sont réussis par moins d'un élève sur deux).

4.2.2.3 LE QUESTIONNAIRE

Lorsque l'exercice de dictée est terminé, l'enquêté répond au questionnaire fermé⁹⁶. Il était indiqué aux enquêteurs de poser les questions dans l'ordre. L'enquêteur devait cocher la réponse choisie par l'enquêté et avait la possibilité d'ouvrir une remarque pour préciser une réponse « originale ».

Les questions fermées sont classées en quatre catégories que nous présentons ci-dessous.

Écriture à la maison

Cette première partie est destinée à connaître les habitudes d'écriture (manuscrite) des enquêtés lorsqu'ils sont chez eux (questions Q1 ; Q2) ainsi que les équipements dont ils disposent avec leur fréquence d'utilisation : ordinateur (Q3, Q3.1, Q3.2 ; Q4) tablette (Q4b, Q4c). Nous interrogeons ensuite les enquêtés sur leur usage d'internet (Q5) et nous cherchons à savoir s'ils ont des pratiques d'écriture électronique ou s'ils ne font que naviguer (Q5.1). Les enquêtés définissent ensuite les raisons pour lesquelles ils écrivent lorsqu'ils utilisent internet (Q6 : mails, réseaux sociaux, forums, chats) et la fréquence à laquelle ils écrivent (Q7).

Ordinateur au travail

Dans cette partie, nous cherchons à connaître l'exposition des enquêtés à l'écriture électronique lorsqu'ils sont au travail : accès à un ordinateur (Q8), fréquence d'utilisation (Q9) et type d'utilisation (Q10).

Orthographe

Cette partie s'intéresse à l'orthographe, de manière générale. Pour définir le rapport à l'orthographe des enquêtés, nous recueillons leur analyse, leur ressenti quant à leur rapport à l'orthographe : nous cherchons à savoir s'il leur arrive d'éprouver des difficultés (Q11 ; Q13) et lesquelles (Q14 : ici nous laissons la personne déclarer spontanément ses points faibles et nous les listons). Si l'enquêté déclare avoir des difficultés, les questions Q15 et Q16 nous permettent de savoir s'il cherche à résoudre ses difficultés et de quelle manière. Leur attitude vis-à-vis de la norme est ensuite recueillie dans la question Q12. Pour conclure cette partie sur l'orthographe, nous leur demandons si, finalement, ils trouvent facile de ne pas faire de fautes lorsqu'ils écrivent (Q17).

⁹⁶ Voir le protocole de l'enquête complémentaire, annexe 8.

Téléphone portable

C'est ici l'utilisation de l'écriture comme moyen de communication qui nous intéresse : utilisation de SMS (Q18 ; Q19), fréquence (Q20) et manière de les écrire, à savoir l'utilisation de l'écriture intuitive (Q21), des abréviations ou du langage texto (Q21.1, Q21.2).

4.2.2.4 L'ENTRETIEN : UN MOMENT PRIVILÉGIÉ

Les questions ouvertes sont posées à la suite des questions fermées et ont été enregistrées. Elles viennent conclure l'enquête et représentent un moment privilégié avec l'enquêté, destiné à faire émerger ses représentations de l'écriture et de l'orthographe.

Pour introduire l'entretien, nous demandons aux enquêtés de fournir spontanément trois mots qui leur viennent à l'esprit pour qualifier l'orthographe (Q1) puis ce qu'ils pensent de l'orthographe du français (Q2).

La suite de l'entretien s'intéresse à la question de **l'apprentissage de l'orthographe**.

Nous les interrogeons d'abord sur leurs souvenirs d'apprentissage (Q3). Cette question faisant appel à des souvenirs parfois très lointains, nous laissons d'abord le temps à la personne de répondre mais si l'enquêté n'arrive pas à formuler de réponse, nous le relançons :

- « *Vous avez peut être un bon ou un mauvais souvenir ?* »
- « *Vous rappelez-vous d'exercices ou de leçons que vous faisiez ? De manuels que vous utilisiez ?* »

La question suivante (Q4) leur demande d'évaluer la difficulté à apprendre l'orthographe. Si la réponse fournie n'est pas étayée, nous relançons : « *Qu'est-ce qui est facile/ difficile ?* ».

Leur implication dans une action d'apprentissage de l'orthographe est ensuite définie dans les questions 5 et 6. Nous leur demandons s'il leur est déjà arrivé d'aider quelqu'un à apprendre l'orthographe et de définir auprès de qui (enfants pour l'école, membres de la famille, adultes au travail, autre).

La question 16 permet de rendre compte de la manière dont ils inscrivent la question de la norme orthographique au sein de la société. Nous leur demandons sans détours ce qu'ils pensent des personnes qui font des erreurs d'orthographe. Cette question cruciale intervient sur la fin de l'entretien car nous espérons que le climat favorable mis en place tout au long de l'enquête et la réflexion qui a eu le temps de se construire au fur et à mesure des questions, nous permettra de recueillir des réponses sincères et étayées.

La dernière question de l'entretien (Q17 : « *Quand vous écrivez, est-ce qu'il y a d'autres choses que l'orthographe qui sont difficiles pour vous ?* ») intervient en guise de conclusion, pour recueillir des informations supplémentaires que l'enquêté souhaiterait nous livrer.

4.3. LE TRAITEMENT DES DONNÉES DE L'ENQUÊTE COMPLÉMENTAIRE

4.3.1. CALCUL DES SCORES SUR LE « RAYON BRICOLAGE »

Pour calculer les scores obtenus par les enquêtés « haut » sur le « rayon bricolage », nous avons attribué 1 point pour chaque item qui respecte le critère utilisé⁹⁷ (orthographique et morphographique) avec 22 mots notés sur 22.

4.3.1.1 LE SCORE ORTHOGRAPHIQUE

Le barème que nous appelons « orthographique », consiste à noter la dictée sur 22 points et reflète l'ensemble des compétences attendues dans l'exercice : maîtrise de l'orthographe lexicale et de la morphographie. À l'inverse de la dictée d'IVQ, la morphographie n'a pas été notée à part de l'orthographe lexicale : aussi, pour réussir les items, les scripteurs doivent retrouver la forme orthographique du mot (orthographe lexicale) ainsi que la morphographie (MÈTRE + S).

Les items qui ont été pris en compte dans le barème sont indiqués en majuscules :

(rayon) BRICOLAGE, DEUX PRODUITS PARFUMÉS qui DÉGRAISSENT bien, quatre PAQUETS de VIS PLATES, un CISEAU à BOIS NORMAL, une BOITE d'ALLUMETTES, une PINCE pour ARRACHER les CLOUS, un POT de PEINTURE, QUELQUES MÈTRES de TUYAU d'ARROSAGE.

4.3.1.2 LE SCORE LEXICAL

Le barème d'orthographe lexical nous permet d'observer les compétences lexicales des enquêtés « haut » sur 21 mots du rayon bricolage. Pour calculer le score de réussite sur chacun des mots, nous attribuons 1 point lorsque la réponse est conforme à la norme orthographique. Nous ne prenons pas en compte la morphographie dans cette analyse.

Dans le tableau 13 nous listons les graphies des mots issues des dictées, que nous avons considérées comme conformes d'un point de vue orthographique.

⁹⁷ Se référer à l'annexe 3 et 3bis pour un relevé complet des graphies des enquêtés « haut » sur la dictée d'IVQ et sur le « rayon bricolage ».

mots	graphies considérées comme orthographiquement conformes
BRICOLAGE	<i>bricolage(s)</i>
PRODUITS	<i>produit(s)</i>
PARFUMÉS	<i>parfume(é)(er)(és)</i>
DÉGRAISSENT	<i>dégraisse(nt), degraisse(nt)</i>
PAQUETS	<i>paquet(s)</i>
VIS	<i>vis</i>
PLATES	<i>plate(s)</i>
CISEAU	<i>ciseau(x)</i>
BOIS	<i>bois</i>
NORMAL	<i>normal(e), droit</i>
BOITE	<i>boite, boîte</i>
ALLUMETTES	<i>allumette(s)</i>
PINCE	<i>pince</i>
ARRACHER	<i>arracher(é)(ées), accrocher</i>
CLOUS	<i>clou(s)</i>
POT	<i>pot</i>
PEINTURE	<i>peinture(s)</i>
QUELQUES	<i>quelque(s), quel que</i>
MÈTRES	<i>mètre(s), mètre(s)</i>
TUYAU	<i>tuyau(x)</i>
ARROSAGE	<i>arrosage</i>

Tableau 13. Relevé des graphies considérées comme orthographiquement correctes dans le score d'orthographe lexicale (sur 16 mots).

4.3.1.3 LE SCORE MORPHOGRAPHIQUE

Nous avons évalué la compétence en morphologie des enquêtés « haut » à partir des 11 marques de nombre et de trois absences de marque présentes dans l'intégralité de la dictée :

- accord de 6 noms : FROMAGES, PRODUITS, PAQUETS, ALLUMETTES, CLOUS, MÈTRES ;
- accord de 2 verbes : SENTENT, DÉGRAISSENT ;
- accord d'un déterminant : QUELQUES ;
- accord de 2 adjectifs : PARFUMÉS, PLATES.

4.3.2. LA TRANSCRIPTION DES ENTRETIENS

Les réponses aux questions fermées ont été saisies dans Excel pour permettre une analyse quantitative des données et les entretiens ont été transcrits intégralement à partir des fichiers

audio⁹⁸. Nous n'avons pas utilisé de conventions de transcriptions particulières, ni de logiciel car nous souhaitons analyser qualitativement les réponses, en nous attachant au fond et à non à la forme.

Nous avons donc saisi le plus simplement possible les réponses des enquêtés, en retranscrivant textuellement leurs déclarations (et les interventions des enquêteurs), y compris les mots de liaison utilisés : « ben, heu, ha, hum, etc. ». Nous n'avons pas retranscrit les marques d'oralité (ex. « *je r'viens* » pour « *je reviens* ») mais avons rédigé les mots en entier (« *je reviens* » pour « *je r'viens* »). Nous n'avons conservé que « *y'a* » pour « *il y a* », car il était utilisé fréquemment. Nous avons respecté l'absence de double négation : ex. « *je suis pas bon* », « *j'aime pas* ».

Nous avons utilisé le moins de marques typographiques possibles, pour simplifier la saisie :

- « » : pour indiquer un discours et marquer un changement de locuteur (alternance enquêteur / enquêté ou intervention extérieur) ;
- ... : pour indiquer la présence d'une pause dans le discours ou pour indiquer que l'enquêté réfléchissait ;
- l'interrogation (?) et l'exclamation (!) : seules les marques de ponctuation clairement indiquées dans le ton utilisé par l'enquêté ont été notées. En revanche, nous n'avons pas noté la prosodie ;
- [bruit] : pour indiquer qu'un bruit nous a empêché d'entendre l'enquêté ;
- (inaudible) si l'un des intervenants a pris la parole mais que son discours est incompréhensible ;
- [rires] : pour indiquer la présence de rires.

La question posée par l'enquêteur est transcrite telle qu'elle apparaît dans le protocole. Pour plus de lisibilité, la réponse de l'enquêté est isolée dans un paragraphe indépendant. Si l'enquêteur est intervenu pendant la réponse de l'enquêté, son intervention est indiquée entre crochets : ex. [*Enquêteur : et si vous voyez des lettres avec des fautes ?*]. Les transcriptions suivent l'entretien et l'ordre des questions. Les réponses sont segmentées en fonction des tours de parole notamment si la réponse a pris la tournure d'une discussion et que l'enquêteur est intervenu.

Enfin, les interventions sont anonymes : les enquêtés et les enquêteurs portent un numéro.

⁹⁸ Voir un exemple de transcription d'un entretien, annexe 9.

Exemple de transcription, enquêté n°26, enquêteur n°2 :

[Enquêteur : Qu'est-ce que vous pensez de l'orthographe du français ?]

« Comme une autre langue, heu ..., l'orthographe du français c'est un peu comme l'espagnol, c'est un peu comme l'italien, plus difficile que l'anglais mais ... »

[Enquêteur : Plus difficile que l'anglais ?]

« Ah ben oui, y'a beaucoup plus d'accords, plus de temps et puis des règles grammaticales un peu compliquées au niveau du complément d'objet On a toujours des doutes. »

[Enquêteur : Que pensez-vous des personnes qui font des erreurs d'orthographe ?]

« Je dirais que ça arrive, mais qu'après ça dépend comment ça leur a été enseigné et quelles difficultés ils avaient à la base, donc quelqu'un qui vit dans une famille qui lit en particulier ou qui est curieux, forcément assimilera plus facilement l'orthographe. Internet, c'est aussi une manière d'apprendre l'orthographe, pouvoir se servir de ce qu'il écrit à un endroit pour pouvoir s'en servir et pour le mémoriser et comme ça on apprend je dirais. »

[Enquêteur : Et si vous voyez des lettres avec des erreurs d'orthographe ?]

« Oui, ben y'a des moments ça me fait surtout sursauter quand c'est des administrations. Y'a des moments pas mal, j'ai eu le cas du ministère de la justice, où des problèmes de phrases, incompréhensibles au-delà des termes juridiques ou alors ça peut être d'autres administrations comme la CAF ou la CPAM quoi. Ou sinon ça peut être oral, au niveau de la télé, des erreurs de vocabulaire, d'utilisation d'un vocabulaire qui n'est pas adapté ... »

[coupé par l'enquêteur (inaudible)]

« Oui, au niveau de l'oralité, du vocabulaire qui n'est pas adapté au thème ou au sujet. »

4.3.2.1 LE CODAGE DES RÉPONSES

Certaines réponses présentes dans les entretiens ont également fait l'objet d'un codage, pour pouvoir réaliser des analyses quantitatives. Pour le calcul des pourcentages nous avons, comme lors de l'analyse des dictées, rapporté la valeur à calculer selon le nombre de réponses recueillies (et non selon le nombre d'enquêtés).

Nous détaillons ci-dessous les réponses qui ont été codées.

REPRÉSENTATIONS DE L'ORTHOGRAPHE

La première question de l'entretien⁹⁹ a permis de recueillir les trois mots qui venaient spontanément aux enquêtés pour définir l'orthographe. Nous avons relevé ces mots dans un fichier Excel avec un code 0 « sans réponse » pour les non-réponses. :

- Trois mots pour qualifier l'orthographe :
 - o mot 1 ;
 - o mot 2 ;
 - o mot 3.

La deuxième question¹⁰⁰ appelait à un jugement :

- « *qu'est-ce que vous pensez de* » et nous avons relevé dans les réponses, les adjectifs fournissant l'opinion des enquêtés autour de trois thématiques principales :

1/ Le niveau de difficulté

Spontanément, la plupart des enquêtés ont apporté un jugement qui traite de la question de la difficulté de l'orthographe. Nous avons donc recensé les adjectifs fournis par les enquêtés pour définir le niveau de difficulté de l'orthographe et avons créé quatre catégories de difficulté (facile – plutôt facile – difficile – très difficile). Nous présentons ci-dessous les différentes réponses recueillies et le code attribué :

- 1 : plutôt facile (« assez facile », « relativement simple », « pas trop compliquée ») ;
- 2 : facile (« facile », « simple », « naturelle ») ;
- 3 : difficile (« compliquée », « difficile », « pas facile », « complexe », « dure », « assez compliquée ») ;
- 4 : très compliquée (« très compliquée », « très difficile », « archi compliquée », « très dure », « une des orthographe les plus compliquées qui soit », « la plus dure ») ;
- 5 : sans avis ;
- 6 : autre ;
- 0 : non-réponse.

⁹⁹ Q1 : « *Quels sont les trois premiers mots qui vous viennent à l'esprit si je vous dis « orthographe ?* ».

¹⁰⁰ Q2 : « *Qu'est-ce que vous pensez de l'orthographe du français ?* ».

Certains enquêtés ont abordé la question de l'efficacité de l'orthographe (*bonne - bien faite*), de sa richesse ou de critère esthétique (la beauté de la langue française) mais nous aborderons ces aspects de manière qualitative.

LA QUESTION DE L'APPRENTISSAGE DE L'ORTHOGRAPHE

La question faisant appel aux souvenirs d'apprentissage de l'orthographe (Q3) a permis de coder la présence ou non de souvenirs et l'évaluation que les enquêtés en ont fait (relance 1).

2/ La présence ou l'absence de souvenirs

- 1 : oui, lorsque les enquêtés répondait « *oui* » ou un équivalent « *je me rappelle des dictées* » ;
- 2 : non, lorsque les enquêtés répondait « *non* » ou un équivalent « *c'est trop lointain* » ;
- 0 : non réponse.

3/ L'évaluation que l'enquêté fait de ses souvenirs

- 1 : bons ou plutôt bons
- 2 : mauvais souvenir
- 3 : neutre, pas de bon ni de mauvais souvenir (« pas particulièrement bon » ou « c'est pas des mauvais souvenirs ») ;
- 0 : non réponse.

La relance 2 permet de savoir si l'enquêté garde des souvenirs précis de son apprentissage de l'orthographe avec exercices en particulier. Nous observons notamment si l'apprentissage de l'orthographe est associé à l'exercice de dictée.

4/ Présence ou absence de souvenir d'exercices ou de leçons

- 1 : oui ;
- 2 : non ;
- 3 : pas vraiment ;
- 0 : non réponse.

Nous avons enfin listé les exercices proposés par les enquêtés lorsqu'ils ont étayé leurs réponses (A-Q3-list-ex) : dictée, etc.

La réponse à la question suivante (Q4 : « *Est-ce facile d'apprendre l'orthographe du français ?* ») pouvait être tranchée (« *oui* » ou « *non* ») ou mitigée chez les enquêtés qui ont

nuancé leur réponse (« *pas vraiment* »). Un enquêté n'est pas parvenu à se prononcer « *il paraît que non* ». Nous avons donc codé les réponses en quatre catégories :

- 1 : oui ;
- 2 : non ;
- 3 : mitigé ;
- 4 : sans avis ;
- 0 : non réponse.

5/ Intervention pour aider à apprendre l'orthographe

La question 5¹⁰¹ nous permet de savoir si les enquêtés sont intervenus auprès d'un tiers pour aider à apprendre l'orthographe :

- 1 : oui ;
- 2 : non ;
- 0 : non réponse.
- Relance¹⁰² « Avec des membres de votre famille, vos enfants (petits enfants), des voisins, des amis ? »

Cette relance permet aux enquêteurs de faire préciser les réponses obtenues car spontanément, les enquêtés faisaient référence à l'aide aux devoirs des enfants et il a fallu les relancer pour qu'ils fournissent d'autres exemples. Cette relance a également permis aux enquêtés de modifier leur réponse. En effet, certains qui avaient le sentiment de ne jamais avoir aidé personne à apprendre l'orthographe, ont modifié leur réponse en se rendant compte qu'ils avaient notamment aidé leurs enfants à apprendre leurs leçons. Enfin, la fin de la relance était destinée à voir si intervenir sur la question de l'orthographe dans le cadre professionnel est considéré comme un tabou « *Au travail, ça vous est arrivé ?* » :

- 1 : oui ;
- 2 : non ;
- 0 : non réponse.

Les justifications des enquêtés qui répondent « non » font également l'objet d'une analyse qualitative.

¹⁰¹ Question 5 : « *Avez-vous eu l'occasion d'aider des personnes à apprendre l'orthographe ?* ».

¹⁰² Relances question 5 : « *Avec des membres de votre famille, vos enfants (petits enfants), des voisins, des amis ?* » ; « *Au travail, ça vous est arrivé ?* ».

6/ Les personnes qui font des erreurs orthographiques

La question 6¹⁰³ interroge les enquêtés sur leur perception des erreurs d'orthographe produites chez les autres. C'est l'analyse qualitative de leurs réponses qui nous ont permis de définir différents types de réactions que nous avons regroupées en trois catégories. Nous présentons les différentes catégories avec des exemples issus des discours. Les points saillants sont soulignés :

- 1 : avis négatif sur les personnes qui ne maîtrisent pas l'orthographe, discours plutôt moralisateurs et normatifs ;

L'enquêté exprime un jugement négatif sur les personnes qui font des erreurs d'orthographe : dans les discours, nous relevons une terminologie péjorative, qui porte atteinte à l'image, au manque de crédibilité professionnel. Nous avons également classé dans cette catégorie, les discours mitigés qui comme dans l'exemple 3 ci-dessous, se disent compréhensifs tout en n'admettant pas que les personnes qui ont des responsabilités professionnelles puissent faire des erreurs orthographiques.

¹⁰³ Question 6 : « *Que pensez-vous des personnes qui font des erreurs d'orthographe ?* ».

Exemple 1 : « J'estime que ce n'est pas "normal", on va à l'école, on apprend, y'a des gens qui sont là qui ben qui s'investissent pour l'apprentissage de la langue, c'est pas logique que derrière on écrive avec des fautes d'orthographe. C'est une mauvaise image de la personne. » (enquête n°72) ;

Exemple 2 : « Si y'a 1 faute ou 2 dans un texte c'est pas gênant, s'il y a beaucoup de fautes je me dis oui c'est quelqu'un qui écrit pas correctement le français. Ça peut donner un mauvais a priori sur la personne, ce qui est un peu bête parce que ça empêche pas ses qualités mais une lettre avec des fautes d'orthographe ou un texte avec des fautes d'orthographe, surtout si c'est le cas d'une personne qui est un peu haut placée, enfin c'est pas ça mais responsable de quelque chose, encore pire, ça fait pas très professionnel du coup. » (enquête n°57) ;

Exemple 3 : « Selon le niveau de la personne ça a tendance à m'interpeller, c'est à dire que si je côtoie des jeunes ingénieurs et je reçois des mails ou des courriers de jeunes, donc des jeunes avec des diplômes un peu élevés où y'a pas mal de fautes, ça ça m'interpelle un peu quand même, alors que j'en ferais pas attention à ça si ça vient de, d'un ouvrier quoi. » (enquête n°67).

- 2 : tolérance et compréhension ;

À l'inverse de la catégorie précédente, certains discours sont compréhensifs, tolérants et ne portent pas de jugement.

Exemple 4 : « J'en pense rien, parce que tout dépend comment ces gens-là ont été à l'école, même quelqu'un qui sait pas lire. Moi une fois je croise quelqu'un qui me demande des renseignements et j'ai vu qu'il savait pas lire les panneaux, bon ben je lui ai expliqué et puis c'est tout, j'ai pas à juger la personne hein. » (enquête n°58) ;

Exemple 5 : « Ben c'est pas de leur faute, après...on peut dire qu'ils ont pas appris à l'école mais bon... c'est pas par rapport à eux. »

[Enquêteur : « disons que c'est plutôt que ça vous énerve quand vous en voyez dans certains courriers ? »]

« Oui mais y'a pas de jugement, je juge pas. » (enquête n°42).

- 3 : sans avis et autres.

Nous avons classé dans cette catégorie, les discours des enquêtés qui ne donnaient pas d'avis et restaient dans une position neutre ainsi ceux qui ne répondent pas vraiment à la question.

Exemple 6 : « Des personnes je pense rien. J'ai du mal à voir une faute d'orthographe devant moi, après... »

[Enquêteur : les personnes qui font beaucoup de fautes ?]

« J'ai pas d'avis particulier sur elles. » (enquête n°1) ;

Exemple 7 : « Ils sont comme moi, pas très bons. » (enquête n°49)

Enfin, nous avons procédé à un recueil des discours qui portent spécifiquement sur la faute d'orthographe.

4.3.2.2 L'ANALYSE DES DISCOURS SUR LA FAUTE DES ENQUÊTÉS « HAUT »

Les discours recueillis dans le cadre de cette enquête doivent permettre de faire émerger le rapport que les scripteurs entretiennent avec la norme orthographique, notamment lorsque nous les avons questionnés sur ce qu'ils pensent de l'orthographe du français¹⁰⁴, ainsi que sur les images que véhiculent pour eux les personnes qui font des erreurs d'orthographe¹⁰⁵. C'est principalement dans les discours sur la faute que les enquêtés se positionnent et révèlent, en creux, leur attitude vis-à-vis de la norme orthographique.

Dans les réponses fournies par les enquêtés, trois types de discours émergent et nous avons typé ces discours en trois grandes tendances : les discours plutôt moralisateurs sur l'orthographe, qui relèvent d'une attitude normative vis-à-vis de la langue (voire hyper normative), des discours plutôt compréhensifs et des discours contradictoires ou mitigés de la part d'enquêtés qui déclarent ne pas se positionner sur ces questions.

Nous avons construit ces catégories à partir de faits de discours qui nous ont semblé pertinents et qui permettent de cerner les différentes réactions des personnes interrogées tout comme Millet et al. dans l'enquête de 1990, qui avaient également relevé différentes attitudes dans les discours des enquêtés, dont la grande majorité se situait dans ce que les auteurs ont appelé l'« attitude enthousiaste » et l'« attitude contradictoire ». « L'attitude enthousiaste » pourrait être rapprochée des discours à tendance moralisatrice de notre enquête.

¹⁰⁴ Question 2 : « Qu'est-ce que vous pensez de l'orthographe du français ? ».

¹⁰⁵ Question 6 : « Que pensez-vous des personnes qui font des erreurs d'orthographe ? ».

Pour cette analyse, nous avons donc procédé de manière qualitative, en effectuant un relevé de tous les faits de discours qui abordent la « faute » d'orthographe qui nous ont permis ensuite de déterminer si l'enquêtés avait une attitude plutôt normative ou non.

4.3.2.3 LES ENQUÊTÉS QUI TIENNENT DES DISCOURS MORALISATEURS SONT-ILS MEILLEURS QUE LES AUTRES EN ORTHOGRAPHE ?

Nous nous sommes ensuite interrogée sur l'existence d'un lien entre le rapport à la norme recueilli dans les discours des enquêtés et leurs compétences orthographiques observées dans la dictée. Nous regardons grâce au test statistique du Khi2 s'il existe une corrélation entre les performances orthographiques observées dans les dictées et le type de discours tenu par l'enquêté.

Nous vérifierons si les enquêtés qui ont tenu des discours moralisateurs sur la norme sont également ceux qui obtiennent les meilleurs résultats sur l'intégralité de la dictée et plus précisément en morphographie, compétence, nous l'avons vu, particulièrement discriminante.

Nous observerons de la même manière, les compétences orthographiques des enquêtés qui ont tenu des discours plutôt compréhensifs ou qui étaient mitigés.

5. SÉLECTION DU CORPUS POUR NOS ANALYSES

Pour déterminer les compétences à l'écrit des scripteurs repérés en difficulté dans l'enquête IVQ, nous disposons de 441 textes recueillis en 2011 et représentant le quart sud-Est de la France (régions Auvergne, Bourgogne, Provence-Alpes-Côte d'Azur et Rhône-Alpes). Suite à un travail de comparaison des résultats obtenus par les enquêtés dans les régions Haute-Normandie et Rhône-Alpes, nous avons fait le choix de ne pas utiliser le corpus complet des dictées. Cette comparaison des données, qui a fait l'objet d'une publication, avait conclu qu'il n'existait pas de différences significatives entre les régions comparées.

Dans les quatre régions sélectionnées pour notre corpus, 447 personnes (sur les 3237 interrogées) ont été dirigées vers le module bas et parmi elles, 6 enquêtés n'ont pas produit de dictée. Certains enquêtés n'ont écrit que quelques items avant d'abandonner. Nous disposons ainsi de 441 textes (complets ou partiels) produits par des enquêtés repérés en difficulté avec l'écrit dans l'enquête (enquêtés « bas ») :

- Auvergne : 28 dictées ;
- Bourgogne : 22 dictées ;
- Provence Alpes Côte d'Azur : 285 dictées ;
- Rhône-Alpes : 106 dictées.

Pour mieux interpréter les résultats de la dictée produite par les enquêtés repérés en difficulté avec l'écrit (dans IVQ 2011) et évaluer leur particularité en production écrite, nous disposons d'un second corpus de dictées qui ont été produites par 87 enquêtés qui avaient été considérés sans difficulté avec l'écrit dans l'enquête IVQ de 2011. Cette dictée est composée de la dictée d'IVQ et d'un « rayon bricolage » ajouté pour disposer de davantage de données.

Ce second recueil de données a été rendu possible grâce à un projet complémentaire intitulé « Compétences orthographiques des adultes et contextes sociaux d'usage » mené dans les régions Rhône-Alpes et Haute-Normandie. Ce projet a reçu le soutien financier de la DGLFLF (Délégation générale à la langue française et aux langues de France) et l'approbation du comité du label de la statistique publique. Il a consisté à faire passer une dictée (composée de la dictée d'IVQ augmentée de quelques items), un questionnaire et un entretien qui porte sur les représentations de l'orthographe.

Nous disposons ainsi de 87 textes (complets ou partiels), questionnaires et discours produits par des enquêtés considérés sans difficulté avec l'écrit dans l'enquête IVQ de 2011 (enquêtés « haut ») :

- Rhône-Alpes : 87 dictées.

Nous présentons à présent nos résultats dans les deux chapitres qui suivent.

PRÉSENTATION DES RÉSULTATS ET ANALYSES

CHAPITRE 5. ANALYSE DE 441 DICTÉES DE SCRIPTEURS EN DIFFICULTÉ DANS IVQ 2011

1. PRÉSENTATION DU CORPUS « BAS »	236
2. RÉSULTATS DE LA DICTÉE D'IVQ.....	239
2.1. RÉPARTITION DES SCORES	242
2.2. COMPÉTENCES EN ORTHOGRAPHE LEXICALE : DES ÉCARTS IMPORTANTS D'UN MOT À L'AUTRE	248
2.3. COMPÉTENCES EN PHONOGRAPHE OBSERVÉES SUR LES PSEUDO-MOTS ET SUR LES MOTS ..	253
2.4. COMPÉTENCES EN MORPHOGRAPHIE OBSERVÉES SUR LES MOTS.....	285
2.5. ANALYSE DE QUELQUES MOTS CHOISIS POUR LEUR COMPLEXITÉ ORTHOGRAPHIQUE.....	298
2.6. CONCLUSION GÉNÉRALE SUR LES 447 ENQUÊTÉS « BAS »	319

Nous présentons dans ce chapitre, une analyse des dictées produites par 441 enquêtés repérés comme étant en difficulté avec l'écrit au terme du module d'orientation de l'enquête IVQ de 2011.

Pour rappel, notre corpus est issu des données recueillies par l'INSEE dans l'enquête Information et Vie Quotidienne de 2011. Nous avons sélectionné quatre régions (selon l'ancien découpage administratif), ce qui représente 447 personnes (parmi les 3237 interrogées dans ces quatre régions, soit 13,8%) repérées en difficulté avec l'écrit suite au module « d'orientation » et dirigées vers le module « bas ». Nous précisons que parmi elles, 6 enquêtés n'ont pas produit de dictée, à cause de difficultés à l'écrit trop importantes. Nous disposons ainsi de 441 textes (complets ou partiels)

Dans un premier temps, nous analysons d'une part les écarts orthographiques entre les formes produites et les formes attendues et observons d'autre part la restitution phonographique des pseudo-mots. Dans un deuxième temps, après avoir écarté les formes qui respectent la norme, nous décrivons finement les choix graphiques des scripteurs en nous positionnant au niveau du phonème et de sa transcription (le graphème). Cette analyse nous permettra de voir si les enquêtés qui n'ont pas su restituer la forme normée ont pu en revanche produire une graphie de type phonographique qui rende le message lisible.

L'écriture sera ici analysée comme un système de représentation de la parole au-delà de la norme, ce qui nous amènera à repenser le concept de maîtrise de l'écriture (et surtout de l'orthographe) comme le principal critère de compétence à l'écrit pour déterminer le taux d'illettrisme. L'INSEE a en effet comptabilisé les compétences en phonographie à partir de 6 pseudo-mots et les compétences en morphographie d'après deux accords en nombre. Néanmoins, nous l'avons vu dans la partie méthodologie (chapitre 4), le score total prend surtout en considération l'orthographe lexicale et grammaticale. La prise en compte d'autres critères de compétence nous permettra de nous demander dans quelle mesure l'orthographe peut être prise en compte lorsque l'on souhaite évaluer l'efficacité d'un message écrit dans un objectif de communication, pour évaluer les compétences en littéracie.

1. PRÉSENTATION DU CORPUS « BAS »

Notre corpus est composé des productions des 441 enquêtés qui sont ceux qui ont produit une dictée parce qu'ils ont été considérés en difficulté avec l'écrit suite à leurs résultats au module « d'orientation » de l'enquête. Ce module « d'orientation » est composé d'un test de lecture de mots, de questions de compréhension orale et d'un test de numératie. Dans notre travail, nous appellerons ce corpus de scripteurs considérés en difficulté (à l'issue du module

d'orientation) : corpus « bas¹⁰⁶ », par contraste avec ceux qui ont été dirigés vers le module « haut » (parce qu'ils n'ont pas été considérés en difficulté au terme du module « d'orientation » de l'enquête) et que nous appellerons corpus « haut ».

Dans les quatre régions sélectionnées pour notre corpus, 447 personnes ont été dirigées vers le module « bas » et parmi elles, 6 enquêtés n'ont pas produit de dictée. Nous disposons ainsi de 441 textes (complets ou partiels) dont la répartition régionale est la suivante :

- Auvergne : 28 dictées ;
- Bourgogne : 22 dictées ;
- Provence Alpes Côte d'Azur : 285 dictées ;
- Rhône-Alpes : 106 dictées ;

Au terme du module « bas », l'INSEE, avec son barème, a classé les enquêtés dans quatre catégories, selon les critères suivants :

- de graves difficultés dans au moins un des domaines fondamentaux de l'écrit ;

Sont classés dans cette catégorie, les enquêtés qui ont obtenu moins de 40% à au moins une des épreuves qui composent le module « bas »¹⁰⁷ : 145¹⁰⁸ enquêtés de notre échantillon ont été classés dans cette catégorie (soit 32%).

- de fortes difficultés dans au moins un des trois domaines fondamentaux de l'écrit ;

Il s'agit des enquêtés qui ont obtenu entre 40 et 60% de réussite au minimum à au moins une des épreuves du module « bas ». Cette catégorie concerne 107 enquêtés de notre échantillon (soit 24%).

- des difficultés (mais ni graves ni fortes) dans au moins un des trois domaines fondamentaux de l'écrit ;

Cette catégorie concerne les personnes qui ont obtenu entre 60 et 80% de réussite au minimum à au moins une des épreuves du module « bas ». 174 enquêtés de notre échantillon ont été classés dans cette catégorie (soit 39%).

¹⁰⁶ Nous reprenons le terme « bas » du module « bas » destiné aux personnes repérées en difficulté avec l'écrit au terme du module d'orientation de l'enquête IVQ.

¹⁰⁷ Pour rappel ce module est composé d'un test de lecture de mots et de pseudo-mots, d'un test de compréhension écrite et d'un test de production écrite de mots et de pseudo-mots (dictée).

¹⁰⁸ Parmi ces enquêtés sont compris ceux qui n'ont produit aucune dictée.

- pas de difficultés dans les trois domaines fondamentaux de l'écrit ;

Pour entrer dans cette catégorie, les enquêtés doivent avoir réussi plus de 80% de chaque exercice du module « bas ». Dans notre échantillon, il apparaît que 21 enquêtés (5%) qui avaient pourtant été dirigés vers le module « bas » d'IVQ n'ont finalement pas été considérés en difficulté avec l'écrit au terme de l'enquête. Cela signifie qu'ils ont obtenu plus de 80% de réussite dans chacun des trois domaines évalués par le module « bas ».

Au niveau national, les enquêtés qui entrent dans les deux premières catégories (graves difficultés et fortes difficultés) ainsi que les enquêtés qui n'ont pas pu passer les épreuves à cause d'une maîtrise insuffisante de la lecture ou du français (1% des 18-65 ans en 2011 selon Jonas, 2012) constituent pour l'INSEE des situations « préoccupantes » face à l'écrit, et ont été identifiées comme personnes qui relèveraient de l'illettrisme (7% au terme de l'enquête IVQ de 2011 contre 9% en 2004).

Cette répartition est illustrée pour notre échantillon par la figure 7 :

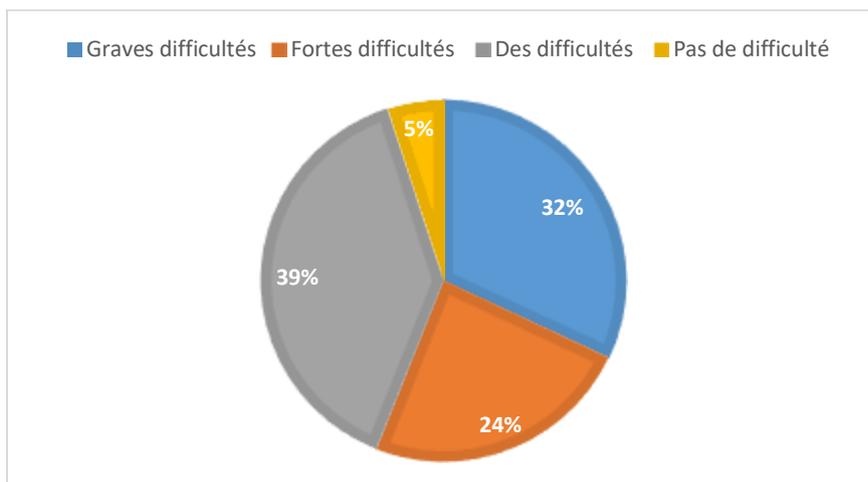


Figure 7 : Répartition des 447 enquêtés « bas » du quart Sud-Est par type de difficulté, sur l'intégralité du module « bas », selon l'INSEE

Selon les catégories définies par l'INSEE, 44% de notre échantillon ne se trouve pas dans une « situation préoccupante » à l'écrit, en ayant réussi plus de 60% des trois épreuves qui permettent d'évaluer le niveau de compétences dans les domaines fondamentaux de l'écrit. Il s'agit de ceux qui ont des difficultés ni graves ni fortes : 39% et ceux qui n'ont pas de difficulté : 5%.

En revanche, d'après ces mêmes critères, 145 enquêtés (32%) ont de graves difficultés dans au moins un des domaines fondamentaux de l'écrit et 107 enquêtés (24%) ont de fortes difficultés dans au moins un des domaines fondamentaux de l'écrit. Ce sont ainsi plus de la moitié des enquêtés de notre échantillon (252 personnes, soit 56%) qui selon la définition de

l'illettrisme définie par l'ANLCI en 2003, sont susceptibles d'être considérés comme illettrés¹⁰⁹.

Selon cette définition :

« on parle d'illettrisme pour des personnes qui, après avoir été scolarisées en France, n'ont pas acquis une maîtrise suffisante de la lecture, de l'écriture, du calcul, des compétences de base, pour être autonomes dans les situations simples de la vie courante. [...] » (ANLCI, 2003).

Selon la définition de l'ANLCI, pour connaître précisément qui est considéré comme illettré dans notre échantillon, il faudrait :

- isoler les enquêtés qui ont montré de fortes ou de graves difficultés dans au moins un des domaines fondamentaux de l'écrit (soit ceux qui ont obtenu moins de 60% de réussite à l'une des épreuves du module « bas ») ;
- et sélectionner parmi eux ceux qui ont été scolarisés en France ; ce critère permettant d'exclure les enquêtés pour qui le français n'a pas été enseigné¹¹⁰.

Nous n'avons pas procédé à ce calcul, car notre travail consiste avant tout à décrire les graphies produites par les enquêtés, quelle qu'ait été leur scolarité.

2. RÉSULTATS DE LA DICTÉE D'IVQ

Notre corpus « bas » est donc composé de 441 textes (complets ou incomplets) car 6 enquêtés n'ont pas été en mesure de produire d'écrit à cause de trop grandes difficultés dans ce domaine. Nous leur avons néanmoins attribué le score de 0 sur 24 à la dictée, car nous voulions que leurs difficultés dans le domaine de l'écriture apparaissent dans les résultats. Leur score a donc été pris en compte dans le calcul du score global sur 24. Leur profil est par ailleurs présenté dans la section 2.1.1 qui traite des dictées qui ont obtenu un score de 0.

En revanche, dans les analyses suivantes (scores de réussite en orthographe lexicale sur les mots, en morphographie, en phonographie sur les pseudo-mots et sur les mots), nous avons écarté ces 6 enquêtés de nos résultats : notre échantillon est alors composé de 441 dictées.

Dans le calcul des différents scores (hormis le score global sur 24), nous avons fait le choix de ne prendre en considération que les réponses réelles : c'est-à-dire les graphies produites par

¹⁰⁹ Le critère des résultats n'est pas le seul. D'après la définition proposée par l'ANLCI (2003), il faut également que la personne ait été scolarisée en France.

¹¹⁰ Nous précisons que la base de données qui nous a été fournie par l'ANLCI pour mener à bien ce travail, ne nous permet pas de savoir quel enquêté a été considéré comme illettré au terme de l'enquête, ni quelles variables précisément ont été utilisées pour effectuer le calcul.

les scripteurs et que nous avons pu informatiser (vs graphies illisibles). Nous rappelons en effet que l'INSEE a fait le choix d'imputer des réponses lorsque les enquêtés n'ont pas produit de graphies.

Dans une démarche qui se veut davantage qualitative que celle de l'INSEE (qui est essentiellement quantitative), nous ne voulions pas imputer de réponses. En effet, nous ne pouvons savoir si les enquêtés qui n'ont pas produit de réponse auraient produit une forme orthographiquement non conforme (ni si leurs réponses, à défaut d'être conformes à la norme orthographique, auraient néanmoins été phonographiquement acceptables).

Nous avons en effet relevé un certain nombre de non-réponses dans les dictées, même sur celles qui ont été produites par des enquêtés qui n'ont pas montré de difficultés particulières sur les autres items. Pour illustrer nos propos, nous présentons une dictée produite par l'enquêté 2010, qui ne contient qu'une seule absence de réponse, ici l'item CERISE :

Pharmacie¹¹¹, anti-rume, micatol, sirap, Alcool 90° DUX, Épicerie, tomates de pays, confiture, sel gobart, 3 fromages qui ne sentent pas for, Librairie, le fiancé solenel de jok pévanor.

Nous constatons que cette dictée ne contient que deux graphies non conformes à la norme orthographique : rume et *solenel*. Aussi, au vue de la dictée produite par cet enquêté, l'absence de réponse sur CERISE n'est probablement pas liée à une trop grande difficulté sur la transcription de ce mot et il apparaît peu probable que cet enquêté aurait produit une graphie non conforme s'il avait écrit le mot.

Nous présentons une autre dictée dont la graphie CERISE est absente (voir le scan de cette dictée produite par l'enquêté 1960¹¹² :

Farmace, antie reme, mikatoul et sirap, alcool a 90 dexe, Ipsore, tomat de peie, convtor, sal cobar, 3 fromaj ke n sant pas for, Librire, le fyanse soulanal des jouk pevaner.

Dans cette dictée, il ne manque que l'item CERISE. En revanche, aucun des mots n'a été orthographié conformément à la norme et seul le pseudo-mot SIRAP possède une graphie phonographiquement acceptable. Ici, nous pouvons supposer que l'enquêté aurait également produit une graphie non conforme de CERISE s'il avait écrit le mot.

¹¹¹ Dans ce manuscrit, les graphies produites par les enquêtés sont reproduites en minuscules et en italique, par contraste avec les graphies en majuscules, qui présentent les mots de la dictée ou des exemples qui ne sont pas issues des graphies de notre échantillon.

¹¹² Pour voir le scan de de cette dictée, se référer à l'annexe 10.

Ces deux exemples qui laissent penser qu'il s'agit davantage d'un oubli que d'une volonté de ne pas transcrire un des mots de la dictée (parce qu'il serait trop difficile) montrent bien que l'absence d'un item peut révéler deux réalités bien différentes. C'est pourquoi nous avons tenu compte du nombre de réponses effectives recueillies, sans imputation.

Dans le tableau 14, nous présentons pour chaque item pris en compte, le nombre de réponses recueilli ; ce nombre étant celui qui a été utilisé pour calculer les différents pourcentages de réussite de nos analyses. Nous précisons également le nombre d'absences de réponses et le nombre de graphies illisibles. Les items sont classés de celui pour lequel nous avons recueilli le plus de réponses (TOMATE-S) à celui pour lequel nous en avons recueilli le moins (SENTENT).

items	graphies recueillies	graphies illisibles	non-réponses
TOMATE-S	432	0	9
CONFITURE	431	1	9
SEL	429	0	12
PHARMACIE	428	1	12
PAYS	426	1	14
GOBAR	425	1	15
CERISE-S	424	2	15
ANTI	423	1	17
FIANCÉ	423	4	14
FORT	423	0	18
ALCOOL	421	0	20
ÉPICERIE	421	1	19
JOC	421	1	19
PÉVANOR	421	1	19
RHUME	420	1	20
SIRAPE	420	0	21
SOLENNEL	420	3	18
LIBRAIRIE	416	2	23
MICATOL	414	2	25
DUX	413	1	27
SENTENT	405	1	35

Tableau 14. Nombre de réponses, d'absences de réponses et de graphies illisibles recueillies pour chaque mot et pseudo-mot de la dictée d'IVQ dans les 441 dictées produites par les enquêtés « bas ».

2.1. RÉPARTITION DES SCORES

Nous présentons dans la figure 8, le score global obtenu par les 447 enquêtés « bas » sur la dictée, recalculé selon nos critères¹¹³. Nous rappelons que pour notre analyse, les items ont été codés selon le respect de la forme orthographique pour les 16 mots et le respect de la forme phonographique pour les 6 pseudo-mots¹¹⁴. La dictée a été notée sur 24 points, avec 1 point attribué par item réussi (16 mots et 6 pseudo-mots) et 2 points pour la morphographie (1 point pour le « -s » de FROMAGES et 1 point pour le « -nt » de SENTENT).

Selon les mots et pseudo-mots, nous constatons une grande disparité dans le nombre de non réponses. Nous précisons que le nombre de graphies illisibles est très faible, avec 4 graphies maximum sur FIANCÉ. Le nombre de réponses recueillies sur les mots semble lié au degré de consistance des mots (calculé dans Manulex-Infra¹¹⁵) : ce sont en effet les mots les plus consistants de la dictée (TOMATE-S, CONFITURE) qui obtiennent le plus grand nombre de réponses, ainsi que les mots fréquents selon Lexique¹¹⁶ (PAYS). Sur SENTENT qui est l'item qui comporte le plus grand nombre d'absence de réponse (35 non-réponses contre 9 pour TOMATE-S), c'est peut-être sa place au sein d'un syntagme qui contient une expression relative (alors que la plupart des items sont des mots isolés) qui a gêné les scripteurs.

¹¹³ Se référer au chapitre 4 section 3.1 pour la méthodologie utilisée.

¹¹⁴ Sur les mots, les graphies que nous avons considérées comme acceptables et approximantes d'un point de vue phonographique n'ont pas été comptabilisées dans ce calcul.

¹¹⁵ Se référer au tableau 3 section 1.2.1 du chapitre 4 pour une présentation des indices de consistance calculés dans la base Manulex pour chaque mot.

¹¹⁶ Se référer au tableau 4 section 1.2.2 du chapitre 4 pour une présentation des indices de fréquence calculés dans la base Lexique pour chaque mot.

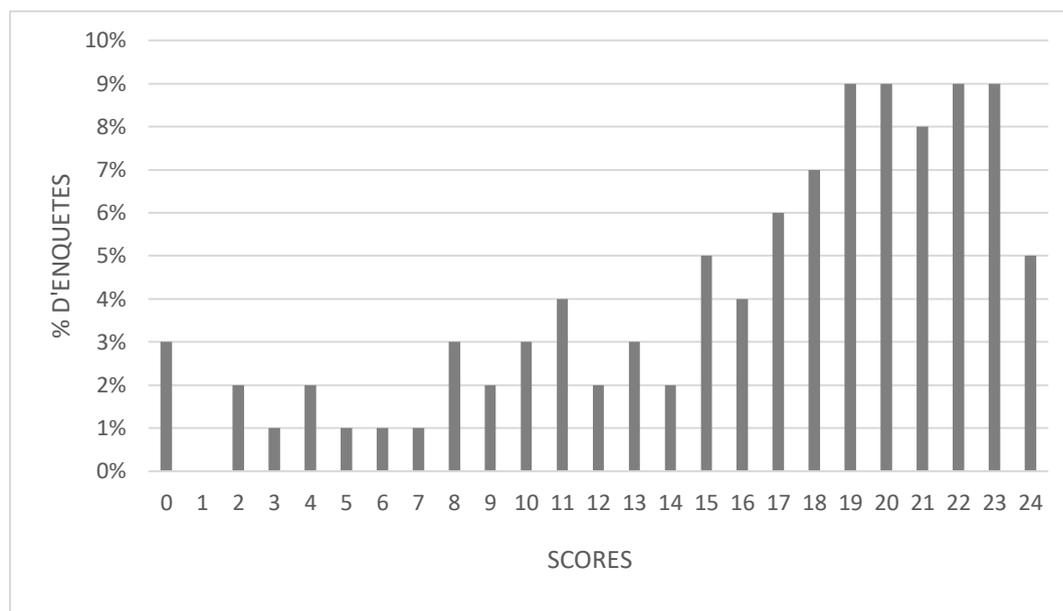


Figure 8. Répartition des scores (sur 24) obtenus à la dictée d'IVQ par les 447 enquêtés « bas ».

La figure 8 permet d'observer une dispersion maximale des scores, répartis de 0 à 24 sur 24, qui montre une hétérogénéité des compétences à l'écrit chez ces scripteurs. Il apparaît néanmoins que la majorité des enquêtés a plutôt bien réussi l'exercice : la moyenne se trouve à 16 sur 24 et la médiane située à 18 sur 24 est plutôt élevée avec une concentration des scores entre 19 et 23.

Alors que ce public a été sélectionné dans l'enquête IVQ pour ses faibles compétences en compréhension de l'oral, en lecture et en numératie (tests qui composent le module d'orientation), il apparaît que 5% des enquêtés (21 personnes) n'ont pas été en difficulté sur cet exercice en parvenant à 100% de réussite et 9% (40 personnes) n'ont produit qu'une seule erreur. En ayant obtenu plus de 80% de réussite à la dictée, ce sont 40% des enquêtés « bas » à qui cet exercice de production écrite n'a pas posé de problème particulier. Selon les critères et le calcul effectués par l'INSEE en 2011, ce sont 37% des enquêtés qui avaient obtenu plus de 80% de réussite sur cet exercice¹¹⁷. Ce faible décalage s'explique par l'application de critères moins stricts dans notre calcul des scores, avec la prise en compte de trois items de plus dont FROMAGE(S) ET SENTE(NT) qui sont considérés comme plutôt consistants par Manulex-Infra et d'un barème qui pénalise moins les enquêtés pour évaluer la morphographie

¹¹⁷ Base de données de l'INSEE, champ « Profil P : résultats en production de mots écrits ».

(les deux marques de nombre ont été évaluées sur 4 points par l'INSEE contre 2 points dans notre barème).

Ce résultat montre qu'un faible niveau de compétences en lecture de mots, en compréhension écrite ou en numération tels que révélés par le test d'orientation pour les enquêtés « bas », n'affecte pas nécessairement les compétences orthographiques (évaluées par les 16 mots et les accords) et phonographiques (évaluées par les 6 pseudo-mots). Le module « d'orientation » se révèle donc insuffisant pour affirmer qu'il existe un lien entre les compétences orthographiques et le reste des compétences évaluées.

La dispersion des scores que l'on peut observer parmi les enquêtés « bas » est révélatrice en effet d'une réalité multiforme des compétences orthographiques. Les résultats montrent ici que la diversité des profils orthographiques est également présente chez un public qui a pourtant été repéré en difficulté avec l'écrit et dont on pourrait penser qu'il formerait un groupe homogène de personnes ne maîtrisant pas la norme orthographique.

2.1.1. LES DICTÉES QUI ONT UN SCORE DE « 0 »

Parmi les 447 enquêtés repérés en difficulté avec l'écrit, 14 ont obtenu un score de 0 sur 24 (soit 3% des personnes interrogées). Nous observons ici leur test de production écrite pour voir si ce score provient d'une dictée entièrement erronée ou d'une absence de réponse.

Parmi les enquêtés qui ont obtenu un score de « 0 », six n'ont écrit aucun item¹¹⁸. Le détail du fichier INSEE indique que ces 6 enquêtés ont tous passé le module d'orientation en entier pour être dirigés directement vers le module « bas ». Par ailleurs deux sur les six n'ont également répondu que partiellement aux autres tests du module « bas » (lecture, compréhension et numération) et semblent avoir de grandes difficultés avec l'écrit, que les enquêteurs ont commentées :

*« L'enquêté ne savait pas lire et trouvait les exercices trop difficiles »
(enquête 1603)*

*« Ne voulait pas écrire car cela demande beaucoup de temps et d'énergie »
(enquête 1989)*

Leur profil montre de grandes disparités dans les parcours personnels et le type de difficultés à l'écrit :

¹¹⁸ Enquêtés 1603, 1755, 1878, 1968, 1989, 2003.

L'enquêté 1603 est un homme né en France en 1946. Sa langue maternelle ainsi que la langue parlée à la maison est le français. Il déclare avoir été scolarisé de l'âge de 6 ans à l'âge de 14 ans mais ne pas savoir « bien » lire le français. Il estime être en difficulté scolaire depuis le CP. Ses résultats dans les autres modules de l'enquête révèlent des difficultés dans tous les domaines de l'écrit, avec moins de 40% de réussite dans tous les exercices (compréhension d'un texte simple, lecture de mots, numération).

L'enquêtée 1989 est une femme née en 1953. Elle est arrivée en France à l'âge de 12 ans et sa langue maternelle est le créole antillais. La langue de scolarisation est en revanche le français. Elle a été scolarisée en France jusqu'à l'âge de 16 ans et a arrêté sa scolarité en classe de 3^{ème}. Ses difficultés en français semblent se concentrer plus particulièrement sur l'écriture puisque les scores obtenus dans les autres exercices sont supérieurs à 60% de réussite : entre 60 et 80% de réussite sur la compréhension d'un texte simple et plus de 80% de réussite en lecture de mots.

Parmi les autres enquêtés qui ont eu un score de « 0 » à l'exercice de dictée, deux enquêtés (enquêté 1636 et enquêté 2015¹¹⁹) ont fait l'effort de produire l'intégralité de la dictée malgré leur grande difficulté à l'écrit : aucun des mots produits par ces deux personnes n'a été correctement orthographié et aucun des pseudo-mots ne présente une forme phonographiquement acceptable¹²⁰. L'étude du profil de ces deux enquêtés montre que leur langue de première scolarisation ainsi que leur langue maternelle est l'arabe.

Enfin, quatre n'ont écrit que le premier (PHARMACIE) ou le deuxième item (ANTI-RHUME) avant d'abandonner :

Enquêté 1644 : *pharmeri* (homme né en 1979 au Vietnam, dont la langue maternelle est le vietnamien, qui est arrivé en France à l'âge de 26 ans).

Enquêté 1767 : *famasie* (femme née en France en 1957, scolarisée en France et dont la langue parlée à la maison est le français).

¹¹⁹ Pour voir le scan de cette dictée, se référer à l'annexe 11.

¹²⁰ Par souci d'économie et de lisibilité, nous utilisons le terme « acceptable » lorsque nous parlons des graphies des pseudo-mots qui permettent de restituer la forme phonique (ex. *micathole* pour /mikatol/ vs *mécatal*).

Enquêté 1781 : *ph* (homme né en 1955 en Afrique¹²¹, arrivé en France à l'âge de 42 ans et dont la langue maternelle est le portugais).

Enquêté 2017 : *nhamacia, antem* (femme née en 1973 en Afrique, arrivée en France à l'âge de 28 ans avec pour langue maternelle le créole du cap vert).

Les graphies proposées par ces enquêtés, dont aucune ne restitue ni la forme orthographique ni la forme phonographique des items, révèlent de grandes difficultés avec l'écriture du français. L'observation du profil de ces quatre enquêtés permet de se rendre compte que pour trois d'entre eux, les difficultés à l'écrit relèvent d'une difficulté d'apprentissage de la langue française, qui n'est ni leur langue maternelle, ni leur langue de scolarisation. En revanche, l'enquêtée 1767 présente un profil différent qui nous invite à présent, à regarder de plus près ce qu'il en est.

L'enquêtée 1767, âgée de 54 ans au moment de l'enquête, est née en France et a été scolarisée en France. Selon les critères définis par l'INSEE, elle possède le profil pour être considérée comme illettrée. Le score obtenu à la dictée d'IVQ et la spécificité de son profil par rapport aux autres personnes ayant obtenu ce score vont dans ce même sens.

Cette personne n'a produit qu'une seule graphie : *famasie* pour PHARMACIE, qui restitue partiellement la forme orthographique attendue :

- réussite phonographique : la forme proposée restitue 6 phonèmes sur les 7 attendus (le /R/ est manquant) ;
- choix des graphèmes : nous pouvons observer qu'elle n'a choisi que des graphèmes de base (Catach et al., 1980) dont deux (le « f » et le « s ») sont erronés. Le choix du graphème « s » dans ce contexte montre une méconnaissance des règles de distribution spécifiques au français, ou, tout du moins, une difficulté à les appliquer ;
- le -e muet final : l'enquêtée a en revanche correctement restitué le -e muet final.

L'analyse de l'unique graphie produite par cette enquêtée montre que sa connaissance du système orthographique du français est partielle mais bien réelle. L'erreur sur le choix du graphème pour transcrire /s/ et l'absence du /R/ montrent des compétences insuffisantes pour proposer une forme qui restitue la forme phonologique du mot. Elle a néanmoins conscience de certaines régularités du français, comme le montre la présence du -e muet final.

¹²¹ Nous ne connaissons pas le pays exact, le code INSEE indiquant « reste de l'Afrique ».

Ainsi, dans la dictée d'IVQ, les enquêtés qui n'ont pas produit d'écrit (ou seulement un ou deux items erronés) ont des profils variés et leurs difficultés en orthographe semble relever de différentes problématiques : le français comme langue étrangère (lorsque le français n'est ni la langue maternelle, ni la langue de scolarisation) ou l'illettrisme selon la définition qui en est proposée par l'ANLCI (2003). Les difficultés peuvent porter sur une ou sur toutes les composantes de l'écrit et il est intéressant d'observer les réussites plutôt que de se centrer uniquement sur les erreurs, comme le cas du « -e final » correctement restitué, cité précédemment. La diversité des profils semble également révéler un rapport à l'écriture différent selon les enquêtés : si certains ont produit une dictée intégralement erronée, d'autres semblent préférer ne rien produire alors que leurs résultats dans les autres composantes de l'écrit montrent qu'ils ont pourtant des compétences à l'écrit. En effet, il semble peu probable que l'enquêtée 1989 (qui nous le rappelons a obtenu entre 60 et 80% de réussite sur la compréhension d'un texte simple et plus de 80% de réussite en lecture de mots) soit dans l'incapacité d'(ortho)graphier un seul des mots de la dictée.

2.1.2. RETOUR SUR LA CLASSIFICATION DES ENQUÊTÉS À L'ISSUE DE LA DICTÉE

Les scores obtenus à la dictée d'IVQ nous permettent de revenir sur la classification des enquêtés proposée par l'INSEE et que nous avons détaillée dans la présentation du corpus (section 1, figure 7).

En utilisant notre barème de correction pour l'exercice de dictée et en répartissant les enquêtés selon la classification de l'INSEE, nous obtenons pour notre échantillon, la répartition des enquêtés « bas » suivante :

- en grave difficulté avec moins de 40% de réussite à la dictée (scores compris entre 0 et 9 sur 24) = 17% ;
- en forte difficulté avec un taux de réussite compris entre 40 et 60% (scores compris entre 10 et 14 sur 24) = 13% ;
- des difficultés partielles, mais ni graves ni fortes avec un taux de réussite compris entre 60 et 80% (scores compris entre 15 et 19 sur 24) = 30% ;
- pas de difficulté avec plus de 80% de réussite (scores à partir de 20 sur 24) = 40%.

Cette classification est illustrée par la figure 9 :

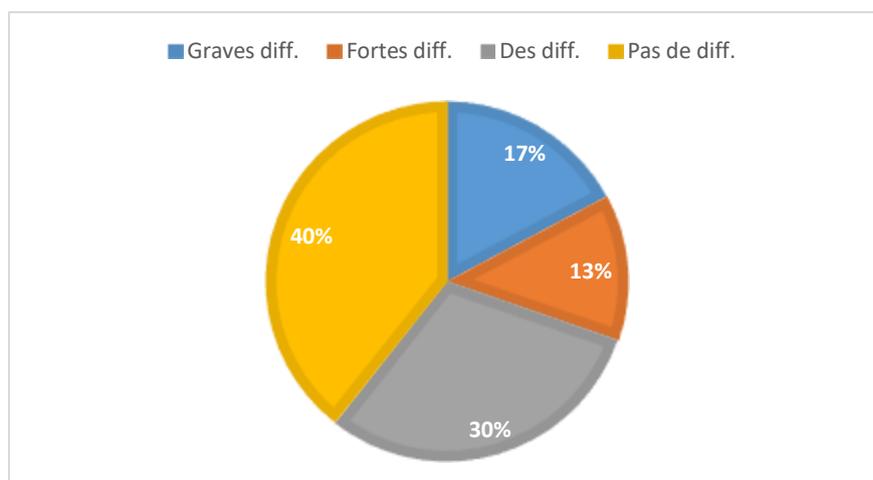


Figure 9. Répartition des 447 enquêtés « bas » du quart Sud-Est par type de difficulté, sur la dictée.

Nous observons que 40% des enquêtés obtiennent plus de 80% de réussite à la dictée d'IVQ, et ne sont pas en difficulté avec l'écriture (et l'orthographe). À l'issue du module « bas », l'INSEE considère que seulement 5% de ces enquêtés ne sont pas en difficulté dans les domaines fondamentaux de l'écrit (parce qu'ils ont obtenu plus de 80% de réussite dans les trois tests du module « bas », lecture, compréhension écrite et écriture). Pour les 35% restants, les difficultés se concentrent donc sur les autres composantes de l'écrit évaluées : la compréhension écrite ou la lecture de mots.

Dans les autres catégories d'enquêtés (difficultés partielles, difficultés fortes, graves difficultés), les difficultés ne se concentrent pas seulement sur l'écriture mais aussi sur la compréhension écrite ou la lecture. En effet, 24% des enquêtés sont considérés en « fortes difficultés » à l'issue du module « bas », dans les domaines fondamentaux de l'écrit alors qu'ils ne sont que 13% à l'être sur la dictée. De même, 32% ont été considérés en « graves difficultés » à l'issue du module « bas », alors que seuls 17% des enquêtés le sont sur la dictée.

Pour une analyse détaillée des résultats obtenus dans la dictée, nous présentons ensuite les compétences orthographiques et phonographiques de 441 scripteurs (ceux pour lesquels nous disposons d'écrits). Les analyses détaillent les résultats en fonction du type de compétences mobilisées : lexicales, phonographiques et morphographiques.

2.2. COMPÉTENCES EN ORTHOGRAPHE LEXICALE : DES ÉCARTS IMPORTANTS D'UN MOT À L'AUTRE

Nos analyses, dans cette section, portent spécifiquement sur les compétences en orthographe lexicale des enquêtés « bas » : c'est-à-dire sur leur capacité à produire des graphies conformes aux attentes orthographiques. Nous classons dans un premier temps les graphies selon leur

ordre de réussite, pour observer les mots les mieux restitués et ceux sur lesquels les enquêtés « bas » ont eu le plus de difficultés. Puis nous proposons de comparer leurs difficultés avec celles rencontrées par des élèves de CM2 sur ces mêmes mots, dont les résultats sont issus d'EOLE (Pothier et Pothier, 2003). Pour rappel, EOLE permet de disposer de repères sur le niveau d'acquisition de l'orthographe lexicale des élèves, du CP au CM2.

2.2.1. CLASSEMENT DES MOTS DE LA DICTÉE D'IVQ CHEZ LES ENQUÊTÉS « BAS »

La figure 10 présente les compétences en orthographe lexicale des 441 enquêtés « bas » (pour rappel, ceux qui ont été orientés vers la dictée à la suite du module d'orientation) sur la dictée d'IVQ à partir de l'analyse des réponses sur 16 mots de la dictée. Les scores ont été calculés en fonction du nombre de réponses recueillies¹²². Les mots sont classés selon leur pourcentage de réussite, du mot le mieux réussi au mot le moins bien réussi. Nous rappelons que dans cette analyse, les marques morphographiques ne sont pas prises en considération.

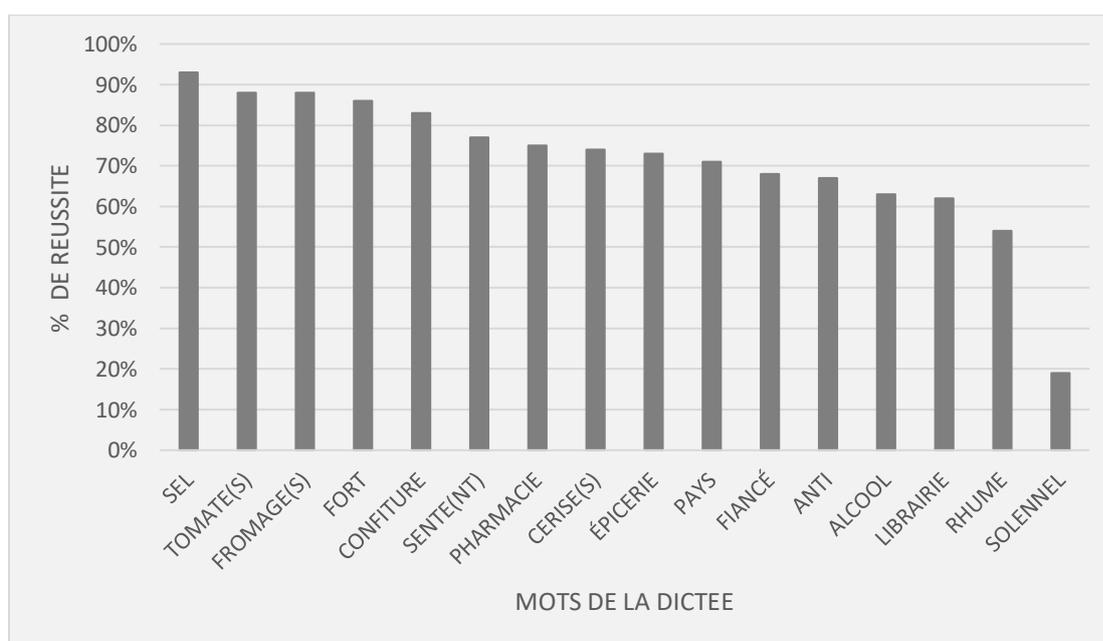


Figure 10. Compétences en orthographe lexicale des 441 enquêtés « bas » sur 16 mots de la dictée d'IVQ.

Ce graphique nous permet d'observer des écarts de réussite importants selon les mots avec 93% de restitution correcte pour SEL contre seulement 19% pour SOLENNEL. Ces deux mots sont de longueur très différente, ce qui pourrait expliquer les différences de réussite : 3 phonèmes pour SEL contre 7 pour SOLENNEL. Cependant, à l'inverse, on observe des écarts de réussite importants entre FROMAGE(S) qui possède 6 phonèmes et RHUME qui n'en

¹²² Se référer au tableau 15, au début de la section 2 pour voir le nombre de réponses recueillies pour chaque mot dans les dictées des enquêtés « bas ».

contient que 3. Cela nous invite à nous interroger sur l'impact de la consistance et de la fréquence sur la réussite.

Il apparaît en effet que ce sont les mots particulièrement consistants (avec un indice > à 80 selon Manulex-Infra)¹²³ et dont les correspondances phonie-graphie sont les plus régulières du corpus qui ont été les mieux réussis : les trois mots les mieux réussis chez les enquêtés « bas » (SEL, TOMATE et FROMAGE-S), sont en effet les plus consistants du corpus (selon Manulex-Infra).

La consistance (Lété, 2004) semble avoir un impact particulièrement important sur la réussite, ce qui semble confirmer que chez les scripteurs considérés en difficulté avec l'écrit, la régularité des correspondances phonie-graphie est un facteur de réussite orthographique.

Mais ce facteur ne semble pas être le seul : certains mots dont le degré de consistance est moindre ont également obtenu une bonne restitution de la forme orthographique. Parmi ces mots moins consistants, se trouvent ceux qui combinent des graphèmes moins fréquents (tels que « ph » et « c » de PHARMACIE) ou ceux qui possèdent une lettre finale muette (tel que le « t » de FORT) : ces deux mots ont été bien restitués, avec 75% de réussite observée sur PHARMACIE et 86% pour FORT. Ces deux mots sont également plutôt fréquents dans Manulex-Infra, ce qui semble confirmer l'importance de la fréquence sur la réussite orthographique chez ces scripteurs en difficulté avec l'écrit, comme indiqué aussi dans la partie théorique.

Ainsi, pour résumer, il apparaît que l'item PHARMACIE, peu consistant¹²⁴, combine en effet plusieurs difficultés : la graphie de /f/ (transcrit « ph ») et de /s/ (transcrit « c ») qui ne font pas partie des archigraphèmes ainsi que la présence d'une finale muette « -e ». Cet item est cependant bien restitué (75% de réussite), ce qui semble indiquer que c'est plutôt l'effet de fréquence qui serait ici un facteur important de réussite. Par rapport aux items de la dictée, PHARMACIE est peu consistant dans Manulex-Infra (indice de consistance 67, rang 11 parmi les mots de la dictée d'IVQ) et également peu fréquent (indice de fréquence 16, rang 10). Nous émettons ici une limite de la base Manulex-Infra lorsqu'elle est utilisée pour évaluer le degré d'exposition aux mots des adultes. Cet item devrait posséder un indice de fréquence plus élevé, mais nous ne disposons d'aucune base qui fournisse des indices issus de tous les supports auquel un adulte est confronté au quotidien. Dans le cas de PHARMACIE, il peut s'agir

¹²³ Voir la section 1 du chapitre 3 pour une présentation de l'outil Manulex et la section 1.2. du chapitre 4 pour une présentation des indices de consistance des mots de la dictée calculés dans Manulex-Infra.

¹²⁴ PHARMACIE, indice de consistance 67.

d'une exposition aux enseignes des pharmacies, dans la rue, mais aussi sur le support qui permet de ranger les ordonnances, généralement fourni par les pharmacies ou encore les emballages qui portent la mention *pharmacie* ou *pharmacien*.

L'item PAYS, réussi à 71% par les enquêtés « bas » apparaît en revanche comme le plus fréquent du corpus (selon Manulex-Infra) tout en possédant une structure rare, qui demande nécessairement au scripteur de l'avoir intégré en mémoire pour le restituer (le transcodage des relations phonèmes-graphèmes ne pouvant aboutir à la graphie correcte). Par ailleurs, même parmi les graphies erronées, nous observons des traces de la structure complexe de ce mot : *pay*, *payi*, *pais*.

Ces caractéristiques permettent de prédire un cumul des difficultés, ce qui se trouve confirmé par les graphies obtenues. Pour SOLENNEL par exemple, 45% des réponses sont phonographiquement acceptables mais 83% sont orthographiquement non conformes (*solanel*, *solannel*, *saulanel*, *sollanel*).

Inversement, nous constatons que les mots dont le degré de consistance est moindre (RHUME, ÉPICERIE, PAYS) ou dont l'indice de fréquence est particulièrement faible (FIANCÉ, SOLENNEL) sont également les moins bien restitués de la dictée. Certains mots sont à la fois peu consistants et peu fréquents : SOLENNEL, ALCOOL, RHUME, FIANCÉ. Ces caractéristiques permettent de prédire un cumul des difficultés, ce qui se trouve confirmé par les graphies obtenues. Nous y revenons plus loin.

2.2.2. COMPARAISON DES DIFFICULTÉS ENTRE LES ENQUÊTÉS « BAS » ET DES ÉLÈVES DE CM2

Nous comparons à présent les difficultés des enquêtés « bas » sur les mots de la dictée d'IVQ aux difficultés des élèves de CM2 dans EOLE. Pour faciliter la comparaison, nous présentons dans la figure suivante les mots présents dans les deux dictées, classés du mieux restitué par les enquêtés « bas » (SEL) au moins bien restitué d'un point de vue orthographique (SOLENNEL).

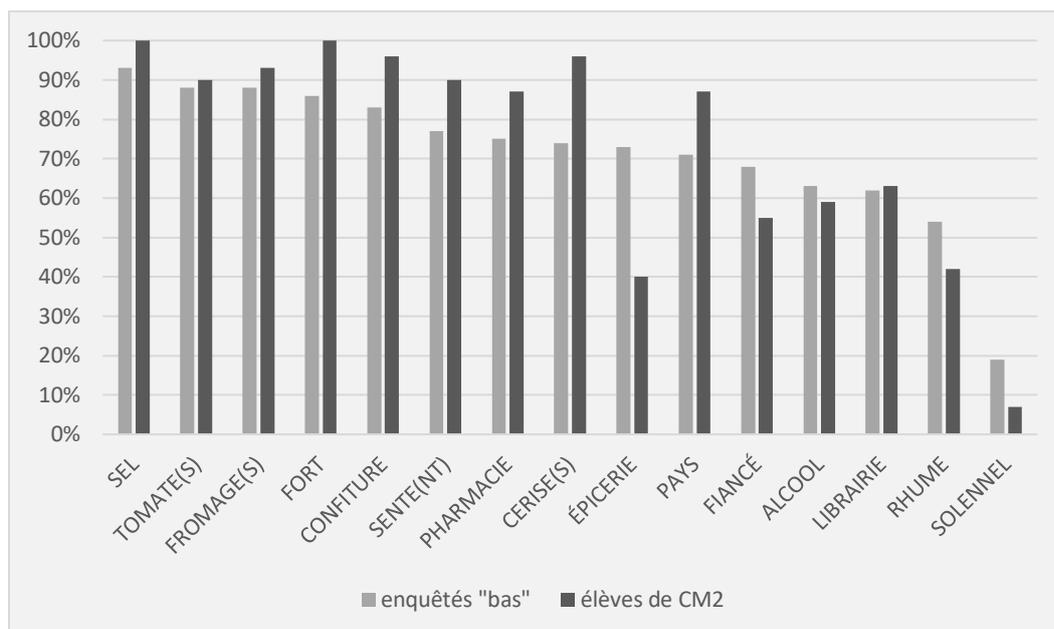


Figure 11. Comparaison des scores de réussite des mots de la dictée d'IVQ, par les 441 enquêtés "bas" et par des élèves de CM2 (EOLE).

Dans le chapitre 4 (méthodologie, section 1.2.2), nous avons présenté les taux de réussite des mots de la dictée d'IVQ qui sont présents dans EOLE (tableau 5). Nous avons relevé que certains mots de la dictée d'IVQ obtiennent de très bons taux de réussite chez des élèves de CM2 (FORT, SEL, CERISE, CONFITURE, FROMAGE) mais que 6 mots obtiennent moins de 75% de restitution orthographique dont en particulier RHUME (42%), ÉPICERIE (40%) et SOLENNEL (7%) qui sont réussis par moins d'un élève sur deux. Pothier et Pothier (2003) ont utilisé la barre des 75% pour indiquer le lexique acquis par les élèves, que nous avons reproduit à titre d'indication sur les mots d'IVQ (voir tableau 5). Au-delà de cette barre des 75%, la figure 11 ci-dessus nous permet de voir que l'ordre des difficultés éprouvés par les enquêtés « bas » et les élèves de CM2 (EOLE) est assez similaire. Les 6 mots qui ne sont pas acquis en CM2 à 75% : ÉPICERIE, FIANCÉ, ALCOOL, LIBRAIRIE, RHUME et SOLENNEL sont à peu près les mêmes que ceux qui ont posé le plus de difficultés aux enquêtés « bas ». Seul PAYS est bien mieux restitué par des élèves de CM2 (87%). Les pourcentages de réussite sur ce mot montrent par ailleurs que son acquisition ne fait qu'augmenter en fonction des classes (comme la plupart des mots) mais qu'il est déjà restitué par un peu plus d'un élève sur deux dès le CE1 (52%) alors que seuls 5% des élèves de CP ont su l'orthographier. Pour rappel, ce mot est celui qui est le plus fréquent dans Manulex parmi les mots de la dictée d'IVQ (U = 309), ce qui semble conforter ici le lien entre la fréquence d'exposition et la compétence orthographique.

Le mot SOLENNEL, très peu réussi à la fois par les élèves de CM2 (7%) et par les enquêtés « bas » (19%) est caractéristique de ce classement des mots par ordre de difficulté : entre la réussite orthographique en fin de cycle primaire et chez des adultes : ce qui est difficile pour les adultes, l'est déjà pour les élèves de 11 ans.

2.3. COMPÉTENCES EN PHONOGRAPHIE OBSERVÉES SUR LES PSEUDO-MOTS ET SUR LES MOTS

La dictée d'IVQ comporte 16 mots, dont les 6 items retenus pour l'analyse menée dans la section 2.5.1 de ce chapitre par rapport aux critères de consistance et de fréquence, et 6 pseudo-mots. Les 6 pseudo-mots sont utilisés dans le protocole d'IVQ pour préciser les compétences phonographiques des enquêtés confrontés à la transcription de mots qui n'existent pas.

Dans cette partie, nous détaillons les compétences en phonographie des enquêtés « bas », dans un premier temps à partir de ce que permettent de dire les graphies produites pour les 6 pseudo-mots. Puis, nous proposons de compléter cette analyse des compétences phonographiques avec l'observation des graphies non conformes des mots existants (d'un point de vue orthographique) en distinguant les graphies non conformes qui permettent de restituer la forme orale du mot (phonographie acceptable et approximante) de celles qui ne la restituent pas (phonographie non acceptable).

Les pseudo-mots et les mots produits par les scripteurs dans le cadre de la dictée ont donc été codés selon leur conformité au principe phonographique. Nous appellerons les réalisations graphiques qui ne modifient pas la forme sonore du mot malgré les erreurs orthographiques qu'elles contiennent : graphies **phonographiquement acceptables** (par exemple *rume* pour RHUME, *en-ti* pour ANTI). Nous montrons que certaines graphies se situent à la frontière de ce qui est acceptable. Nous les appelons graphies **approximantes**.

Les graphies approximantes sont celles qui :

- proposent des associations de graphèmes qui n'existent pas en français, tout du moins dans certaines positions : par exemple *ssirape* pour le pseudo-mot /sirap/ ;
- ne respectent pas les règles de position ou de distribution spécifiques au français : par exemple pour ÉPICERIE, la graphie *épiseri* pour qui le « s » dans cette position (entouré de deux voyelles) aurait la réalisation phonique /z/ et non /s/ ;
- qui témoignent de marques d'oralité comme l'élision du « e » muet dans ÉPICERIE : par exemple *épissri*.

Par cette approche, nous cherchons à mieux distinguer les composantes de la maîtrise de l'orthographe et à préciser ce qu'il en est de la capacité des scripteurs à coder les phonèmes en proposant une forme graphique des mots qui permet, dans un objectif de communication, de transmettre un message.

2.3.1. COMPÉTENCES EN PHONOGRAPHIE SUR LES PSEUDO-MOTS

2.3.1.1 SCORES DE RÉUSSITE SUR LES GRAPHIES ACCEPTABLES ET APPROXIMANTES

Nous débutons cette analyse par le classement des pseudo-mots selon leur pourcentage de réussite phonographique, présenté dans la figure 12. Nous avons calculé les scores à partir des réalisations graphiques que nous avons qualifiées d'acceptables et d'approximantes. Comme pour les mots, les pourcentages de réussite ont été calculés en fonction du nombre de réponses obtenues, en mettant de côté les cas d'absences de réponse (ou de réponses illisibles)¹²⁵.

Pour faciliter les analyses, les pseudo-mots sont classés du mieux réussi au moins bien réussi par les enquêtés « bas ».

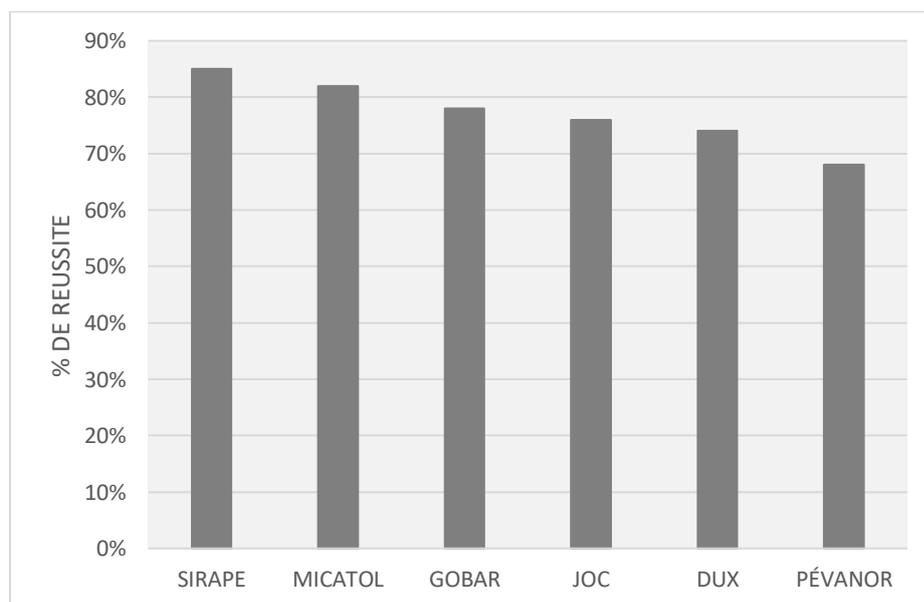


Figure 12. Proportion de réponses phonographiquement acceptables et approximantes obtenues par les 441 enquêtés « bas » pour chaque pseudo-mot.

Les taux de réussite sur les pseudo-mots sont compris entre 85% et 68%. Le résultat moyen est un peu plus élevé que celui obtenu sur les mots (77% contre 72% pour les mots) mais

¹²⁵ Se référer au tableau 14, au début de la section 2 de ce chapitre pour voir le nombre de réponses recueillies pour chaque pseudo-mot dans les dictées des enquêtés « bas ».

aucun pseudo-mot n'a obtenu un score de réussite aussi élevé que certains mots¹²⁶ (par exemple SEL réussi par 93% des enquêtés). En revanche nous n'observons pas de score inférieur à 68% de réussite (pour PÉVANOR) contrairement à 4 mots : ALCOOL (63%), LIBRAIRIE (62%), RHUME (54%) et SOLENNEL qui n'obtient que 19% de réussite. Il faut rappeler que pour les pseudo-mots, nous nous intéressons à la phonographie uniquement. En effet, il n'est pas possible d'analyser la maîtrise de l'orthographe de mots qui n'existent pas, alors que les scores pour les mots existants concernent leur orthographe lexicale (et pour certains aussi grammaticale). Pour les pseudo-mots, comptabilisons la proportion de réponses acceptables et approximantes que nous opposons aux graphies qui ne restituent pas la forme phonologique des pseudo-mots (et que nous qualifions de non acceptables).

Comme pour les mots, on note des écarts de scores entre les pseudo-mots. Pour les mots existants, nous avons eu recours aux critères de fréquence et de consistance pour analyser ces écarts. Nous allons à présent détailler les raisons possibles de ces écarts pour les pseudo-mots.

2.3.1.2 HYPOTHÈSES SUR LES ÉCARTS ENTRE LES PSEUDO-MOTS

Nous nous interrogeons ici sur les raisons qui pourraient expliquer la différence de résultats entre les pseudo-mots : notamment entre SIRAPE, le mieux réussi (85%) et PÉVANOR, le moins bien réussi (68%). Nous avançons quatre hypothèses : la place des pseudo-mots dans la dictée (hypothèse liée à une surcharge cognitive), le choix des pseudo-mots, la proximité de certains avec des mots existants), l'impact de la structure syllabique ou du nombre de syllabes sur la réussite, ou encore un biais possible lié aux conditions de l'enquête.

LA PLACE DES PSEUDO-MOTS DANS LA DICTÉE : SURCHARGE COGNITIVE ?

La première explication pourrait être liée à la surcharge cognitive, avec notamment en cause la place des pseudo-mots dans la dictée. L'ordre de réussite des pseudo-mots est quasiment identique à celui de leur apparition dans la dictée. Par ailleurs, la dernière phrase (qui est le titre d'un roman et son auteur) cumule les difficultés : LE FIANCÉ SOLENNEL DE JOC PÉVANOR, avec deux mots difficiles et deux pseudo-mots consécutifs (avec PÉVANOR, 6^e et dernier pseudo-mot, le moins bien restitué des six).

¹²⁶ Voir la figure 10 (section 2.2, chapitre 5) pour le détail de la réussite des enquêtés « bas » en orthographe lexicale, sur chaque mot.

LE CHOIX DES PSEUDO-MOTS

Nous pouvons également évoquer le choix des pseudo-mots. Deux pseudo-mots ressemblent à des noms existants : SIRAPE, proche de SIROP (d'autant plus qu'il est présenté comme une marque de sirop contre la toux) et JOC qui ressemble au prénom JACQUES.

La proximité du pseudo-mot JOC avec le prénom JACQUES peut expliquer les 18 occurrences associées à la prononciation /ʒak/ dans les réponses des enquêtés (18 occurrences) : *jacques* (5), *jacque* (5), *jaque* (4), *geaque* (1), *gac* (1), *jack* (1), *jake* (1). Nous retrouvons également le prénom GEORGES (2 occurrences) : *georges* (1), *gorge* (1).

La structure du mot « sirop » est quant à elle observée dans les graphies *sirope* (3), *syrop* (1), *sirop* (1), *cirope* (1).

Dans SIRAPE, la proximité avec un mot existant aurait été facilitatrice, alors que pour JOC elle aurait été plus perturbante, comme c'est le cas aussi pour DUX qui peut être assimilé à LUXE ou aux chiffres DEUX (6 occurrences) et DIX (2 occurrences).

LA STRUCTURE SYLLABIQUE DES PSEUDO-MOTS¹²⁷

Une autre hypothèse pour expliquer les différences de réussite entre les pseudo-mots pourrait être la structure syllabique des pseudo-mots.

SIRAPE et MICATOL, les deux pseudo-mots les mieux réussis, présentent une structure syllabique différente (respectivement CV-CVC et CV-CV-CVC) avec un nombre différent de syllabes. PÉVANOR, le pseudo-mot le moins bien réussi, possède une structure identique à celle de MICATOL (CV-CV-CVC). Enfin, SIRAPE et GOBAR, qui possèdent la même structure CV-CVC, présentent un taux de réussite assez proche (avec respectivement 85% et 78% de réussite) alors que MICATOL et PÉVANOR qui possèdent tous deux la structure CV-CVCVC présentent un taux de réussite très différent (respectivement 82% et 68%). Par ailleurs tous les pseudo-mots commencent par une structure CV et se terminent par une structure CVC (sauf DUX qui, phonologiquement, possède une structure CVCC).

La structure syllabique des pseudo-mots ne semble pas être une explication convaincante : la différence de réussite proviendrait plutôt de la composition phonémique de ces pseudo-mots, avec des correspondances phonie-graphie plus faciles à restituer que d'autres, peut-être

¹²⁷ Cet aspect des choix linguistiques opérés a été développé dans une communication faite au congrès mondial de linguistique française en 2016. Texte disponible en ligne : <https://doi.org/10.1051/shsconf/20162707003>

parce qu'elles sont plus faciles à saisir à l'oral ou parce qu'elles relèvent de schémas plus fréquents à l'écrit et correspondent à des régularités enregistrées par les scripteurs.

LA PRONONCIATION DES PSEUDO-MOTS ET LES CONDITIONS DE PASSATION : UN POSSIBLE BIAIS

Certaines erreurs pourraient également provenir d'un biais du protocole de l'enquête lié à la prononciation de l'enquêteur ou aux conditions de passation du test (bruits pouvant perturber l'audition, etc.). Les résultats des enquêtés « haut » (voir section 3.2, chapitre 6) montrent en effet que les erreurs produites sur les mots sont majoritairement d'ordre orthographique, très peu d'erreurs d'origine phonographique ayant été relevées chez ces scripteurs. Sur les pseudo-mots en revanche, de nombreuses erreurs ont été relevées, notamment sur PÉVANOR (11% des graphies phonographiquement incorrectes : *pavanord, bevanor, pevarnor*). La présence de telles graphies chez des scripteurs qui n'ont pas montré de difficulté particulière sur le reste de la dictée invite à s'interroger sur un possible biais du protocole.

Les graphies *djoc* et *djoque* nous questionnent également sur la manière dont certains enquêteurs ont prononcé l'item JOC : /ʒɔk/ ou /dʒɔk/? Nous ne savons pas quelles instructions ont été données aux enquêteurs qui ont réalisé l'enquête de 2011 concernant la prononciation de ces pseudo-mots. Dans l'enquête complémentaire de 2014, nous avons également relevé un taux de réussite plus bas sur les pseudo-mots chez les enquêtés interrogés par un des enquêteurs (alors que la majorité de ces enquêtés - 6 sur les 11 interrogés - ont pourtant obtenu de très bons scores à la dictée), que l'on pourrait attribuer à la manière dont il a prononcé les items. Pour l'enquête complémentaire, les enquêteurs ont participé à une journée de formation pendant laquelle la dictée leur a été présentée et lue. Dans les consignes écrites qui portent spécifiquement sur la dictée, il est indiqué :

« Commencer à lire les mots que l'enquêté doit inscrire sur la feuille, en lui indiquant si besoin du doigt l'endroit où il doit écrire le mot que vous venez de lire. »

« Lisez lentement et le plus distinctement possible en adaptant votre lecture au rythme d'écriture de l'enquêté. Répétez la phrase si nécessaire. »

« Si l'enquêté semble ne pas bien comprendre les mots inventés, répétez-les-lui distinctement en articulant chaque syllabe et en lui rappelant que c'est une marque inventée. »

Pour autant, les enquêteurs avaient-ils conscience de l'importance de la précision dans la prononciation de ces mots inventés ?

Si l'on souhaite évaluer les compétences orthographiques et phonographiques à partir de l'oral, les paramètres liés à l'oralité sont à considérer : pour limiter le biais observé ici, on pourrait envisager une dictée sous forme enregistrée et proposée dans de bonnes conditions d'audition (bruits extérieurs, qualité du son, etc.), avec des arrêts possibles et des segmentations de l'enregistrement pour permettre des réécoutes mot par mot.

Par ailleurs, les pseudo-mots ont entraîné des scores de réussite relativement importants y compris chez les enquêtés « bas » et s'avèrent donc peu discriminants. On peut dès lors se questionner sur leur intérêt dans une dictée pour évaluer les compétences en phonographie et repérer des difficultés dans ce domaine. Nous l'avons vu, les compétences en phonographie des enquêtés « bas » peuvent être saisies dans l'analyse des graphies erronées des mots de la dictée, ce qui permet d'identifier les enquêtés qui ont des difficultés dans ce domaine, au-delà du respect de la norme orthographique (ex. RHUME : *rime, rune, rom*). Nous émettons enfin une réserve sur l'utilisation du pseudo-mot JOC, trop proche d'un prénom et de DUX, proche de LUXE (2 occurrences dans le corpus) ou des chiffres DEUX (6 occurrences) et DIX (2 occurrences). DUX ne permet par ailleurs d'observer que des compétences phonographiques restreintes : 91% des enquêtés « bas » ont correctement restitué la forme phonologique du début du mot /dy/, les erreurs portant sur la transcription de la suite /ks/.

2.3.1.3 VARIATION SUR LES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES

Pour les pseudo-mots, il était attendu du scripteur qu'il transcrive la forme sonore entendue en choisissant parmi les correspondances phonie-graphie du français. Certains phonèmes pouvaient donc être transcrits par plusieurs graphèmes (ex. /i/ : « i », « y ») alors que d'autres ne pouvaient être transcrits que d'une seule manière (ex. /m/ à l'initiale : « m »).

Le tableau 15 ci-dessous précise la variation observée sur les formes produites. Il présente :

- les réponses que nous avons considérées comme acceptables du point de vue phonographique : le nombre de formes graphiques différentes = les variantes (colonne 1) et le pourcentage de réponses (colonne 2) ;
- les réponses que nous avons considérées comme approximantes (« approx. » dans le tableau) : le nombre de formes graphiques différentes = les variantes (colonne 3) et le pourcentage de réponses (colonne 4) ;
- les réponses que nous avons considérées comme non acceptables phonographiquement : le nombre de formes graphiques différentes = les variantes (colonne 5) et le pourcentage de réponses (colonne 6) ;

Nous précisons que conformément à la méthodologie annoncée (voir chapitre 4), nous avons neutralisé l'accentuation pour considérer comme une seule variante les formes accentuées et non accentuées (ex. : *pevanor*, *pévanor*, *pèvanor* = une même variante PÉVANOR). Dans le tableau 15, les pseudo-mots sont présentés selon leur taux de réussite, du mieux réussi au moins bien réussi du point de vue phonographique.

	1	2	3	4	5	6
pseudo-mots	variantes acceptées	% de réponses acceptées	variantes approx.	% de réponses approx.	variantes non acceptées	% de réponses non acc.
SIRAPE	18	85%	5	1%	50	14%
MICATOL	28	82%	0	0%	55	18%
GOBAR	13	78%	0	0%	58	22%
JOC	24	76%	9	5%	54	18%
DUX	14	74%	5	7%	55	19%
PÉVANOR	30	68%	0	0%	103	32%

Tableau 15. Détail de la variation sur les pseudo-mots chez les 441 enquêtés « bas ».

Sur les pseudo-mots, les possibilités de variantes sont importantes pour la transcription d'un même item (choix différents de graphèmes, redoublements de consonnes, ajout de lettres muettes, segmentation). Le nombre de variantes est cependant restreint pour les formes acceptables ou approximantes d'un point de vue phonographique, par rapport au nombre de variantes différentes que nous considérons comme non acceptables phonographiquement.

Parmi les graphies considérées comme acceptables sur le plan phonographique, on retrouve peu d'écart dans le nombre de variantes entre MICATOL et PÉVANOR, les deux pseudo-mots qui contiennent 3 syllabes (avec respectivement 28 et 30 formes différentes). Les autres pseudo-mots présentent également un nombre de variantes acceptables relativement similaire (entre 13 et 18 formes). Seul JOC a suscité un nombre plus élevé de propositions avec 24 formes acceptables et 9 formes approximantes.

Pour 5 des 6 pseudo-mots, le nombre de graphies non acceptables est assez similaire avec un nombre de variantes compris entre 50 (pour SIRAPE, le mieux réussi) et 58 (pour GOBAR). En revanche, PÉVANOR, le pseudo-mot le moins bien restitué présente un nombre de variantes deux fois plus important que les autres pseudo-mots avec 103 formes différentes.

L'analyse des formes proposées par les enquêtés permettra d'observer les réponses récurrentes et de comprendre ce qui pose problème au niveau phonographique, pour les enquêtés qui ont à transcrire ces pseudo-mots.

Le tableau 16 présente pour chaque pseudo-mot, à titre d'exemples, 20 graphies (parmi lesquelles se trouvent les plus récurrentes), que nous avons considérées comme phonographiquement non acceptables. Nous indiquons entre parenthèses le nombre d'occurrences. Les graphies, pour chaque pseudo-mot, sont classées selon le nombre d'occurrences.

MICATOL	SIRAPE	DUX	GOBAR	JOC	PÉVANOR
nicatol (8)	siraque (3)	duce (7)	cobar (12)	jox (8)	pivanore (3)
mecatol (5)	sirope (3)	deux (6)	cobare (10)	jacque (5)	prévanor (3)
mécato (2)	cérape (2)	dus (5)	goubar (6)	jacques (5)	pevanol (3)
mecatole (2)	cirat (2)	dusque (4)	goubare (3)	jque (4)	evanor (3)
mécato (2)	sirate (2)	dix (2)	goba (3)	choc (3)	eranol (3)
mekatol (2)	sirup (2)	dups (2)	coba (2)	jouque (2)	péranor (3)
micadole (2)	surape (2)	dusse (2)	gombar (2)	jos (2)	puvanore (2)
micadol (2)	cerape (2)	doux (2)	copar (2)	jouk (2)	prévanord (2)
makatol (1)	cirapare (1)	duques (2)	gabard (2)	soce (1)	pavanor (2)
mecatile (1)	cirapeure (1)	luxé (2)	cobard (2)	foc (1)	pavanore (2)
mecalole (1)	cypar (1)	dixse (1)	quobar (2)	geod (1)	penavor (2)
medical (1)	cyrak (1)	dks (1)	cobarre (2)	geaque (1)	pénavol (2)
makatale (1)	siraph (1)	dics (1)	cobart (2)	georges (1)	penavore (2)
megatole (1)	cirapre (1)	dexe (1)	coubare (1)	gogue (1)	pevanare (2)
mequatole (1)	ciralé (1)	digux (1)	caubare (1)	gopenor (1)	pévanol (2)
méquatolle (1)	cirope (1)	diquxe (1)	caubar (1)	gonque (1)	pevanoir (2)
mercatole (1)	dicirppe (1)	diquesi (1)	cobas (1)	gorge (1)	péfanor (1)
licatale (1)	firape (1)	disc (1)	copa (1)	gouch (1)	pefanero (1)
miatol (1)	shyrape (1)	dexu (1)	combar (1)	jace (1)	péfanor (1)
micacol (1)	cirabe (1)	dixe (1)	codar (1)	gque (1)	péfanore (1)

Tableau 16. Relevé de 20 graphies de chaque pseudo-mot considérées comme phonographiquement non acceptables chez les 441 enquêtés « bas ».

Nous présentons ensuite dans le tableau 17, l'intégralité des formes graphiques considérées comme acceptables qui ont été recueillies dans les 441 dictées, pour chaque pseudo-mot. Les

formes mises en italique sont des formes accentuées qui ont également été proposées sans accent. Le fait qu'elles soient en italique indique donc qu'il faut les rapprocher de la même forme non accentuée car elles ne constituent pas une variante à part entière : comme nous l'avons dit plus haut, nous considérons que *pevanor*, *pévanor*, *pèvanor* sont une même variante de la forme PÉVANOR.

Le nombre d'occurrences est indiqué entre parenthèses et les réponses présentées de la plus fréquente à la moins fréquente.

MICATOL	SIRAPE	DUX	GOBAR	JOC	PÉVANOR
micatol (166)	sirap (111)	dux (203)	gobar (139)	jock (83)	pevanor (90)
micatole (56)	sirape (101)	duxe (76)	gobard (102)	joc (77)	<i>pévanor (48)</i>
mikatol (38)	cirape (54)	ducse (7)	gobare (32)	jok (54)	pevanore (37)
micathol (16)	cirap (33)	dukse (4)	gobart (27)	joque (43)	pevanord (30)
mikatole (8)	cyrape (10)	du ^{sc} ¹²⁸ (3)	gaubar (10)	jocque (16)	<i>pévanord (19)</i>
micatolle (7)	syrape (9)	du ^{sc} (3)	gaubard (7)	jauque (11)	<i>pévanore (16)</i>
mycatol (7)	syrap (9)	dûx (1)	gobarre (4)	joke (9)	<i>pévanort (11)</i>
micathole (4)	sirappe (8)	duxce (1)	gaubart (3)	jocques (6)	pevanort (6)
miquatol (4)	cyrap (5)	duqce (1)	guobar (2)	jocq (6)	<i>pèvanor (3)</i>
mycathol (4)	sirape (5)	d'uxe (1)	gobbar (1)	joq (2)	<i>pèvanore (3)</i>
micatôle (3)	cirape 3)	duxs (1)	gaubarre (1)	joqu (2)	<i>pevanores (2)</i>
miquatole (3)	cirappe (3)	duckse (1)	guobard (1)	jauc (2)	<i>pévannor (2)</i>
mycathole (3)	cirrape (2)	dukce (1)	guobare (1)	jocke (1)	<i>pévanaure (2)</i>
mycatole (3)	si rape (2)	ducse (1)		joch (1)	paivanore (2)
mickatol (2)	cyrappe (1)			jock (1)	p.vanor (2)
miccatol (2)	cyrrape (1)			joques (1)	pevanaure (1)
mykatole (2)	siraps (1)			jauques (1)	pevanorr (1)
micatoll (1)	cirrap (1)			joqe (1)	<i>pévanorre (1)</i>

¹²⁸ Nous avons considéré la graphie « sc » comme une manière de réaliser le graphème « x ». En effet, au niveau de l'écriture manuscrite il n'est pas toujours possible de savoir si la lettre formée correspond à un « x » ou un « sc ».

mickatole (1)	sirrape (1)	joks (1)	paivanor (1)
miquatolle (1)		jauke (1)	pévanaur (1)
mi ka tol (1)		jauq (1)	pe vanore (1)
mikathol (1)		geoc (1)	p.evanor (1)
mykattol (1)			pevanor (1)
mi catol (1)			p vanor (1)
mycatoll (1)			petvanor (1)
miqatole (1)			pey va nor (1)
micateùl (1)			pevanorre (1)
miekatele (1)			pevanors (1)
			pe va nore (1)
			pevaneur (1)

Tableau 17. Relevé des graphies considérées comme phonographiquement acceptables pour chaque pseudo-mot chez les 441 enquêtés « bas ».

Nous observons pour chaque pseudo-mot entre une et trois formes récurrentes, dont certaines le sont particulièrement. Nous présentons ci-dessous, les réponses recueillies pour chaque pseudo-mot.

PRÉSENTATION DES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES DE MICATOL

Pour **MICATOL**, 77% des propositions considérées acceptables se présentent sous la forme de 3 graphies :

- *micatol* (166 occurrences, soit 49% des graphies acceptables) ou sa forme avec un « -e » muet final *micatole* (56 occurrences, soit 17% des graphies acceptables) ;
- *mikatol* (38 occurrences, soit 11% des graphies acceptables).

La graphie la plus utilisée *micatol(e)* représente 66% des réponses acceptables et 54% du nombre total de réponses, tous types de graphies confondus (formes approximantes et non acceptables). Elle est composée d'archigraphèmes : « i » pour /i/, « c » pour /k/, « o » pour /o/ et possède une forme minimale, sans ajout de lettre muette ou redoublement de consonnes. La graphie *mikatol* ne diffère de la précédente que dans le choix du graphème pour transcrire /k/.

L'observation des autres graphies qui représentent 23% des graphies obtenues montre en revanche de nombreuses variantes sur les graphèmes :

- « i » et « y » pour /i/ ;
- « c », « k », « qu », « q », « cc », « ck » pour /k/ ;
- « t », « tt », « th » pour /t/ ;
- « o » et « ô » pour /ɔ/. Notons que pour ce phonème, aucun « au » n'a été relevé, ce qui rejoint les observations faites pour les graphies obtenues pour l'item TOMATE dans la section 2.5.

Nous rappelons par ailleurs que les graphies permettant de réaliser la forme /mikatoel/, dont le phonème /oe/ est proche du /ɔ/ normalement attendu, ont été considérées comme acceptables car possiblement liées à des facteurs phonétiques de prononciation de l'enquêteur ou de réception par l'enquêté.

Nous relevons deux graphèmes qui, utilisés en l'état, ne sont pas plausibles dans le système graphique du français mais permettent néanmoins de transcrire phonétiquement l'item :

- « q » utilisé dans *miqatole* (et qui devrait être suivi de « u ») ;
- l'ajout d'un « e » muet dans *miqatele*.

Ces graphies peuvent témoigner de plusieurs phénomènes que nous ne pouvons que supposer :

- une connaissance partielle du système graphique du français ;
- une liberté vis-à-vis des règles de fonctionnement de l'orthographe, pour transcrire des noms propres ;
- l'influence d'autres systèmes d'écriture, par exemple en allemand (« mie » /mi/), ou de prononciations distinguant un /i/ court et long.

Dans ce cas, un entretien métagraphique aurait éventuellement permis de mieux comprendre l'origine de ces graphies, mais nous notons néanmoins que la très grande majorité des graphies acceptables proposent d'une part une forme qui respecte les règles de fonctionnement du français et d'autre part une forme qui utilise en majorité les graphèmes les plus fréquents.

L'observation de la lettre finale montre que 70% des formes graphiques choisies se terminent par le graphème « l », sans ajout d'un « -e » muet final¹²⁹.

Ainsi, la majorité des enquêtés qui ont produit une forme graphique acceptable de MICATOL, ont privilégié une graphie minimale, en choisissant des archigraphèmes.

PRÉSENTATION DES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES DE SIRAP

Pour **SIRAPE**, deux graphies sont très récurrentes : *sirape* (et sa forme avec un « -e » muet final *sirape*) et *cirape* (et sa forme sans « -e » muet final *cirap*), que nous présentons ci-dessous :

- les graphies *sirap* (111 occurrences, soit 31% des propositions acceptables) et sa forme avec un « -e » muet final *sirape* (96 occurrences, soit 27% des graphies acceptables) représentent 58% des propositions acceptables et 49% du nombre total de réponses, tout type de graphies confondues (formes approximantes et non acceptables).

La forme *sirap* rappelle la graphie du mot SIROP et a probablement influencé la transcription de ce pseudo-mot (qui désigne la marque d'un sirop contre le rhume : ANTI-RHUME MICATOL OU SIRAPE).

- *cirape* (54 occurrences, soit 15% des graphies acceptables) et sa forme sans « -e » muet final *cirap* (33 occurrences, soit 9% des graphies acceptables) représente 24% des réponses acceptables et 21% du nombre total de réponses, tout type de graphies confondues.

Lorsque l'on observe l'intégralité des graphies acceptables, 47% n'ont pas de « -e » muet final.

Le graphème « s » à l'initiale a été choisi par 68% des scripteurs. Ce choix correspond à la fréquence de ce graphème à l'initiale des mots du français.

Comme pour MICATOL, nous relevons parmi les propositions de graphies acceptables, toutes les combinaisons de graphèmes possibles :

- « s » et « c » pour /s/ ;
- « i » et « y » pour /i/ ;
- des redoublements de consonnes « r », « rr » pour /r/, « p », « pp » pour /p/.

¹²⁹ Une analyse portant spécifiquement sur le choix des finales des pseudo-mots MICATOL, SIRAP, GOBAR et PÉVANOR est proposée chapitre 6, section 1.5.2.3.

Comme pour MICATOL, ce sont les graphies minimales et les graphèmes les plus fréquents (comme le « s » à l'initiale ») qui ont été privilégiés.

PRÉSENTATION DES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES DE DUX

DUX, dont les possibilités de graphies sont limitées, est le pseudo-mot qui présente à la fois le moins de variantes et une forme graphique très fréquente : nous relevons 206 occurrences de *dux* et *duc* (soit 68% des réponses acceptables) ainsi que 79 occurrences de cette forme avec ajout d'un « -e » muet final *duxe* et *duce* (soit 26% des réponses acceptables). Nous rappelons que nous avons considéré la graphie « sc » comme une réalisation graphique de « x », par rapport à la proximité de formes que présentent ces graphèmes dans l'écriture manuscrite.

La forme *dux(e)* représente ainsi 94% des graphies phonographiquement acceptables et 69% du nombre total de réponses, tout type de graphies confondues.

La réalisation graphique de /ks/ semble avoir posé problème à 17 enquêtés, qui ont cherché des graphiques plausibles : « cs », « ks », « xce », « qce », « xs », « ckse », « kce », « csse ».

96 enquêtés (30%) ont ajouté un « -e » muet final, parfois indispensable par rapport aux possibilités de l'orthographe du français comme dans *ducse*, *duqce* ou accessoire avec *duxe*, mais entre les deux formes les plus fréquentes, c'est la graphie minimale *dux*, sans lettre muette qui a été préférée.

PRÉSENTATION DES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES DE GOBAR

Les 4 propositions les plus fréquentes de **GOBAR** ont pour différence le choix de la finale muette avec, comme pour les pseudo-mots analysés précédemment, la graphie minimale *gobar* qui a été privilégiée :

- *gobar* (139 occurrences, soit 42% des réponses acceptables) ;
- *gobard* (102 occurrences, soit 31% des réponses acceptables) ;
- *gobare* (32 occurrences, soit 10% des réponses acceptables) ;
- *gobart* (27 occurrences, soit 8% des réponses acceptables).

Parmi les propositions de graphies acceptables, nous relevons deux graphies différentes sur l'archigraphème « o » avec « au » proposé dans *gaubar* (10), *gaubard* (7), *gaubart* (3), *gaubarre* (1). Les enquêtés ont donc privilégié le graphème le plus fréquent « o » (93% contre 6% pour « au »).

On trouve également la proposition « gu » pour /g/ dans seulement 4 graphies : *guobar* (2), *guobard* (1), *guobare* (1) et un redoublement de consonnes sur le « r » dans *gobarre* (4) et *gaubarre* (1).

Pour écrire cet item (comme pour les pseudo-mots analysés précédemment), les enquêtés ont donc privilégié les graphèmes les plus fréquents. Par ailleurs, 178 graphies proposent une lettre finale muette « -e », « -d » ou « -t », soit 54% des propositions acceptables.

PRÉSENTATION DES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES DE JOC

Nous observons sur le pseudo-mot JOC qui ne comporte que trois phonèmes, presque autant de formes différentes que sur l'item PÉVANOR :

- *jock* (83 occurrences, soit 26% des réponses acceptables) ;
- *joc* (77 occurrences, soit 24% des réponses acceptables) ;
- *jok* (54 occurrences, soit 17% des réponses acceptables) ;
- *joque* (43 occurrences, soit 13% des réponses acceptables).

La graphie la plus fréquemment utilisée est *jock* (26%), qui contrairement aux autres pseudo-mots, n'est pas la graphie minimale. Elle se rapproche en revanche du prénom JACK dans sa version anglo-saxonne, qui a probablement influencé les scripteurs (prénom dont nous retrouvons la structure dans : *jocke* (1), *jokc* (1) avec une inversion des graphèmes « k » et « c »). Les traces de ce prénom dans sa version française JACQUES apparaissent également dans les graphies : *joque* (42), *jocque* (16), *jauque* (11), *jocques* (6), *joques* (1), *jauques* (1) et peut-être *jocq* (6). Nous soulignons la présence d'un « -s » final qui rappelle particulièrement celui présent dans le prénom.

Pour transcrire le premier phonème /ʒ/, un seul scripteur a utilisé le graphème « g » en ajoutant un « -e » muet pour produire le phonème attendu /ʒ/ avec la graphie *geoc* (1). Ce graphème a davantage été utilisé par les enquêtés¹³⁰ qui ne sont pas parvenus à produire une forme graphique plausible, parce qu'ils n'ont pas ajouté de « -e » muet après « g » : *goc* (6), *goque* (5), *gok* (4), *goq* (2), *goke* (2), *gock* (1), etc.

Nous observons enfin la restitution du dernier phonème /k/ :

- « ck » : *jock* (83). L'utilisation de ce graphème par 26% des scripteurs est le plus fréquent. Nous pouvons probablement l'imputer à sa proximité avec le prénom JACK ;

¹³⁰ 27 graphies contiennent cette erreur phonographique.

- « que » : *joque* (43), *jocque* (16), *jauque* (11), *jocques* (6), *joques* (1), *jauques* (1). Parmi les réponses acceptables, la finale en « -que » a été utilisée par 24% des scripteurs ;
- « c » (25%) : *joc* (77), *jauc* (2), *geoc* (1) ;
- « k » (17%) : *jok* (54) ;
- « ke » : *joke* (9), *jocke* (1), *jauke* (1) ;
- « q » : *jauq* (1), *joq* (1). Seuls deux scripteurs ont utilisé le graphème « q » qui est un graphème très peu fréquent en finale des mots (ex. « cinq », « coq ») tout comme l'utilisation du graphème « ch », rare en finale et utilisé par un scripteur : *joch* (1).

Nous relevons 9 graphies qui proposent en finale, des graphèmes qui n'existent pas en l'état en français : *joqu* (1), *joqe* (1) et enfin *jokc* (1), du moins pas dans cette position pour *jocq* (6) dont le nombre d'occurrences suggère que « cq » est identifié comme un graphème possible de /k/, comme dans ACQUIS, ou comme évoqué par rapport au prénom « JACQUES ».

PRÉSENTATION DES GRAPHIES PHONOGRAPHIQUEMENT ACCEPTABLES DE PÉVANOR

PÉVANOR est le pseudo-mot qui a été le moins bien restitué avec 68% de graphies acceptables. Comme pour GOBAR, les propositions les plus fréquentes ont pour différence le choix de la finale muette :

- *pevanor* (avec ou sans accent sur le « e » de la première syllabe, 141 occurrences, soit 49% des réponses acceptables)
- *pevanore* (avec ou sans accent, 56 occurrences, soit 19% des réponses acceptables) ;
- *pevanord* (avec ou sans accent, 49 occurrences, soit 17% des réponses acceptables) ;
- *pevanort* (avec ou sans accent : 17 occurrences, soit 6% des réponses acceptables).

Près de la moitié des enquêtés qui ont fourni une graphie phonographiquement acceptable de PÉVANOR (49%) ont choisi la forme graphique minimale *pevanor* (avec ou sans accent). Les autres propositions récurrentes ne se distinguent que par le choix de la finale muette : *pevanore* (avec ou sans accent), *pevanord* (avec ou sans accent), *pevanort* (avec ou sans accent). Comme pour GOBAR, ce sont ainsi 263 graphies (soit 91% des propositions acceptables et 62% du nombre total de réponses, tous types de graphies confondues) qui présentent une graphie minimale pour les 7 phonèmes de ce pseudo-mot : /pevanɔr/ = *pevanor*, avec ou sans ajout d'une finale muette.

L'observation des graphies montre des variantes graphiques sur les phonèmes /p/, /e/, /n/, /ɔ/ et /R/ :

- /e/ transcrit « e » ou « é » dans la majorité des graphies (90) contre « ai » (3), « et » (1), « ey » (1). On trouve également 4 graphies qui utilisent la valeur épellative du graphème « p » de manière isolée comme une abréviation matérialisée par un espace (*p vanor*) ou par un point (*p.vanor*), ce qui à la lecture fournit la forme phonologique attendue /pevanɔR/ ;
- /ɔ/, transcrit majoritairement « o » avec seulement 4 graphies « au » (4) ;
- on trouve enfin 3 redoublements de consonne sur le « n » et 2 sur le « r ».

Il apparaît que la variation graphique sur cet item est restreinte : la majorité des graphèmes choisis par les enquêtés sont des archigraphèmes : « p », « e » ou « é », « v », « a », « n », « o », « r ». La variation observée sur les graphies se situe majoritairement sur le choix d'une finale muette.

2.3.1.4 BILAN SUR LES COMPÉTENCES EN PHONOGRAPHIE SUR LES PSEUDO-MOTS

L'analyse des graphies proposées par les enquêtés « bas » pour transcrire les pseudo-mots montre un nombre important de variantes, qui s'explique par la diversité des combinaisons de graphèmes qu'offre le français. Parmi celles que nous avons considérées comme acceptables d'un point de vue phonographique (parce qu'elles permettent de restituer la forme phonologique de l'item), nous relevons entre 14 (pour DUX) et 30 (pour PÉVANOR) formes graphiques différentes. En revanche, selon les pseudo-mots, il apparaît qu'entre 1 et 3 graphies ont été préférées par les scripteurs. Pour 5 des 6 pseudo-mots, nous relevons une graphie (qui contient à minima 2 formes qui ne varient que par la présence d'un « -e » muet final) particulièrement récurrente :

- *micatol(e)* : 66% des réponses considérées comme phonographiquement acceptables et 54% du nombre total de réponses (formes approximantes et formes phonographiquement non acceptables) ;
- *sirap(e)* : 58% des propositions acceptables et 49% du nombre total de réponses ;
- *dux(e)* : 94% des graphies phonographiquement acceptables et 69% du nombre total de réponses ;
- *gobar* : 42% des réponses acceptables avec 3 autres formes graphiques qui ne varient que par l'ajout d'une finale muette (« -d », « -e », « -t »), ce qui représente pour la proposition *gobar(d)(e)(t)* 91% des réponses acceptables et 71% du nombre total de réponses ;

- *pevanor* : 49% des réponses acceptables avec 3 autres formes graphiques proches, qui ne varient que par l'ajout d'une finale muette (« -e », « -d », « -t »). La graphie *pevanor(e)(d)(t)* représente ainsi 91% des propositions acceptables et 62% du nombre total de réponses.

Pour ces 5 pseudo-mots, la graphie la plus fréquemment utilisée représente entre 58% et 91% des réponses phonographiquement acceptables et entre 49% et 71% du nombre total de réponses, tous types de graphies confondues. Il apparaît également que cette graphie est uniquement composée d'archigraphèmes (et donc des graphèmes les plus fréquents), ce qui représente une forme graphique que l'on peut qualifier de « minimale » (sans ajout de lettre muette ou de redoublement de consonnes), permettant de restituer de manière efficace la forme phonologique attendue.

En revanche, le pseudo-mot JOC ne fonctionne pas de la même manière. En effet, nous n'observons pas de forme graphique très fréquente, mais 3 propositions récurrentes : *jock*, *joc*, *jok*. De même, la graphie la plus fréquemment utilisée, *jock*, n'est pas constituée des graphèmes les plus fréquents du français. La proximité avec le prénom JACK, semble avoir influencé la graphie de ce pseudo-mot. La finale « -ck » peut également provenir de connaissances culturelles (influence anglo-saxonne probablement).

L'observation des autres propositions graphiques pour chaque pseudo-mot, montre peu de récurrence, car il y a de nombreuses variantes dont la plupart n'ont été produites que par un seul enquêté. Ces variantes utilisent ainsi les différentes possibilités graphiques offertes par le français, dont certaines dites « approximantes » se situent à la frontière de ce que l'on peut considérer comme phonographiquement acceptable.

Parmi les formes graphiques acceptables, le fonctionnement des finales muettes observées sur GOBAR et PÉVANOR retient notre attention et nous amène à proposer une analyse qui porte spécifiquement sur le comportement des scripteurs vis-à-vis de la possibilité d'ajouter une finale muette qui n'a pas un caractère obligatoire pour assurer le fonctionnement phonographique de la consonne finale¹³¹.

¹³¹ Une analyse portant spécifiquement sur le choix des finales est proposée section 1.5.2.3 du chapitre 5.

2.3.2. COMPÉTENCES EN PHONOGRAPHIE SUR LES MOTS

Après avoir analysé les graphies produites pour les 6 pseudo-mots d'un point de vue phonographique (section 2.3.1), nous nous centrons sur les compétences phonographiques des enquêtés « bas » observées sur 16 mots de la dictée. Nous comparons la proportion d'enquêtés qui proposent une graphie qui restitue la forme phonologique des mots (graphies phonographiquement acceptables additionnées aux graphies conformes aux attentes orthographiques) à ceux qui ont restitué la forme orthographique attendue. Comme pour les pseudo-mots, nous voulons cerner la capacité des scripteurs à coder les phonèmes de manière conforme au système phonographique du français, pour mieux distinguer les différentes composantes de la maîtrise de l'orthographe.

2.3.2.1 SCORES DE RÉUSSITE

Pour observer l'apport des différentes compétences (orthographiques, phonographiques - au sens strict et au sens large) sur la réussite, la figure 13 présente la répartition des réponses des enquêtés « bas » selon les différents critères : graphies orthographiquement correctes, phonographiquement acceptables, phonographiquement approximantes et graphies erronées¹³², indiquées « non conformes » dans la figure).

Nous rappelons que conformément au principe phonographique que nous avons défini dans la méthodologie (chapitre 4, section 3.2), les graphies comme *rume* ou *ruhme* ont été considérées comme acceptables du point de vue de la phonographie bien qu'elles soient éloignées de la norme orthographique, car elles permettent de restituer la forme orale des mots tout en tenant compte des règles de position et de distribution spécifiques au français. Ont été également acceptées les formes contenant des segmentations non conformes (*en-ti*, *et pisseri*, etc.) avec ou sans trait d'union, ces segmentations ne modifiant pas la forme sonore du mot, ainsi que les réalisations qui témoignent de marques d'oralité comme l'élision du « e » muet dans ÉPICERIE (*épissri*).

Ont également été prises en compte les réalisations graphiques plausibles de chaque phonème. Les propositions qui ne tiennent pas compte des règles de position ou de distribution spécifiques au français, sont classées dans les graphies **approximantes**. Ces dernières peuvent permettre au scripteur de faire passer un message : par exemple la graphie *rhum*, dans le contexte d'un énoncé (ici, à la PHARMACIE acheter un ANTI-RHUME) permet au

¹³²Les graphies considérées comme erronées sont celles qui ne respectent ni le principe orthographique ni le principe phonographique, au sens strict comme au sens large.

lecteur de comprendre qu'il s'agit bien d'un anti-rhume et non d'un antidote à la boisson (RHUM).

Dans la figure 13, les pourcentages de réussite pour chaque catégorie sont calculés selon le nombre de réponses recueillies et les mots sont classés par ordre de réussite orthographique, du mieux restitué (SEL) au moins bien restitué (SOLENNEL).

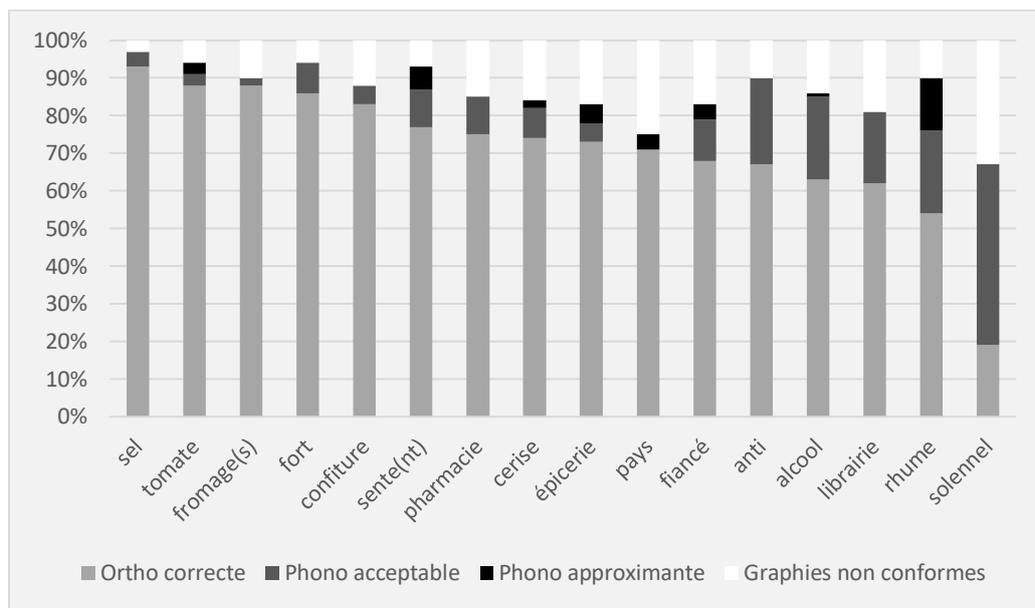


Figure 13. Classement des mots de la dictée d'IVQ chez les 441 enquêtés « bas » selon le type de réponse : orthographe conforme, phonographie acceptable, phonographie approximante, graphies non conformes.

L'observation de la restitution des mots selon le critère phonographique révèle des taux de réussite plutôt élevés, compris entre 67% pour SOLENNEL et 97% pour SEL alors qu'avec la prise en compte du critère orthographique seulement, le pourcentage de réussite en orthographe lexicale est compris entre 19% sur le mot SOLENNEL et 93% sur le mot SEL. La prise en compte du critère phonographique (graphies acceptables) fait augmenter ainsi le score moyen sur chaque mot et fait augmenter la moyenne de réussite sur l'ensemble des mots de 13 points (85% contre 72% avec la prise en compte du critère orthographique seulement). Sur certains mots, la prise en compte du critère phonographique fait augmenter les résultats de manière parfois très nette, notamment sur les mots les moins bien restitués du point de vue orthographique comme ALCOOL, LIBRAIRIE, RHUME et surtout SOLENNEL. Cet item, qui a particulièrement posé problème aux enquêtés « bas » (19% de graphies correctes) passe à 67% de restitution correcte (plus 48 points) lorsque l'on prend en compte le critère phonographique. Par ailleurs, nous n'observons que deux graphies approximantes : *solanille* et *colanille*. Toutes les autres graphies que nous avons qualifiées de phonographiquement acceptables respectent les règles de position et de distribution spécifiques au français : *saulanel*, *solannel*, *sollanel*, etc. Ainsi, aucun mot n'obtient de score inférieur à 67%.

Une prise en compte large du principe phonographique (graphies approximantes) ne change pas les résultats moyens : le score moyen est légèrement supérieur et gagne 2 points (87%). La prise en compte de ce critère n'a en effet d'incidence que sur peu de mots : SENTE(NT) (6%), CERISE (2%), ÉPICERIE (5%), PAYS (4%), FIANCÉ (4%) et RHUME (14%).

Sur RHUME, qui contient le plus de propositions approximantes, c'est la proximité de la forme *rhum* (que nous n'avons pu considérer comme une forme acceptable à cause de l'absence d'un « -e » muet final qui transforme la valeur phonologique du graphème « u ») qui fait augmenter significativement la proportion de réponses dites « approximantes » par rapport aux autres mots. La graphie *rhum* a en effet été proposée par 51 scripteurs alors que pour les autres mots, les propositions approximantes, lorsqu'il y en a, ne concernent qu'entre 9 et 23 occurrences (soit 2% et 6% des réponses).

Dans les analyses suivantes, nous observons en particulier les graphies non conformes aux attentes orthographiques de ÉPICERIE et LIBRAIRIE, pour rendre compte précisément des difficultés d'ordre phonographique que les enquêtés ont rencontrés sur ces deux mots.

2.3.2.2 ANALYSE DES GRAPHIES NON CONFORMES PRODUITES SUR LES ITEMS ÉPICERIE ET LIBRAIRIE

Nous observons ainsi les compétences en phonographie des scripteurs du corpus « bas » sur quelques difficultés d'ordre phonographique d'après les réponses produites sur deux mots de la dictée : ÉPICERIE et LIBRAIRIE.

- L'item ÉPICERIE nous permet d'observer le comportement des scripteurs face à une difficulté d'ordre phonographique importante en français : la transcription du phonème /s/ par un graphème qui tienne compte du contexte intervocalique ;
- L'item LIBRAIRIE fait partie des mots qui ont été les moins bien restitués (62% de réponses orthographiquement correctes). Cette analyse nous permettra de déterminer ce qui a posé problème aux scripteurs sur cet item.

Nous présentons les réponses non conformes aux attentes orthographiques recueillies dans ce corpus dans deux tableaux distincts, un pour chaque mot.

LA RESTITUTION DU PHONÈME /S/ DANS ÉPICERIE

Pour rendre compte des compétences phonographiques des enquêtés, nous observons dans le tableau 18 une des difficultés de l'item ÉPICERIE : la restitution du phonème /s/. Nous détaillons notre manière de procéder avant d'en venir au tableau. Les réponses sont classées en fonction du phonème utilisé pour transcrire le /s/ attendu dans cet item : /epis(ə)ri/.

- pour les graphies dans lesquelles il est observé une **restitution correcte de /s/** : les graphèmes proposés forment /s/ et correspondent à la forme attendue /s/. Nous classons les réponses en deux catégories :
 - (1) les réponses qui permettent d'obtenir une graphie de l'item dont la phonographie est acceptable (ex. *epiceri* = /episəri/). Parmi ces graphies, nous avons pris en compte les formes qui témoignent à l'écrit de l'assimilation du /ə/ (ex. *episri* = /episri/);
 - (2) les réponses qui ne restituent pas la forme phonographique attendue de l'item, malgré la restitution correcte de /s/ : ex. *epecerie* = que nous pouvons lire /epəsəri/).
- pour les graphies dans lesquelles il est observé une **restitution incorrecte de /s/** qui est la forme attendue, parce que les graphèmes utilisés produisent la forme /z/. Les réponses sont classées en deux catégories qui distinguent les graphies dont la forme phonographique ne présente comme erreur que la transcription phonographique du phonème /s/ comme dans *episeri* de celles qui présentent plus d'une erreur phonographique (ex. *episiri*) :
 - (3) 1 erreur phonographique sur le /s/ attendu = /z/ : ex. *episerie* ;
 - (4) plus d'une erreur phonographique : ex. *episiri* ;
 - (5) une dernière colonne recense enfin les quelques graphies qui présentent une autre transcription phonographique du /s/ attendu : /k/ et /ʃ/ : ex. *picherie*.

Pour chaque graphie, le nombre d'occurrences est indiqué entre parenthèses.

1	2	3	4	5
phonographie acceptable ou approximante	phonographie non acceptable	phonographie non acceptable 1 erreur	phonographie non acceptable 2 erreurs ou plus	phonographie non acceptable
/s/ = /s/	/s/ = /s/	/s/ = /z/	/s/ = /z/	/s/ = /k/, /ʃ/
epiceri (8)	epecerie (4)	episerie (8)	episiri (2)	épiecrie (2)
episserie (4)	epecirie (2)	episeri (4)	épicserie (1)	epicrie (1)
epiçerie (2)	epicirie (2)	et piserie (2)	epeiseire (1)	epicri (1)
épisserie (2)	epcerie (2)	épiserie (1)	pisr (1)	picherie (1)
epicery (1)	epicérie (2)	et piseri (1)	epesorie (1)	epècrer (1)
épicerrie (1)	ipiceri (2)		pisarie (1)	epcorio (1)
épiceri (1)	epecieri (2)		episary (1)	epucrai (1)
episseris (1)	epecèrie (1)		episéri (1)	et piro (1)
hépicerie (1)	éperciei (1)		épouséri (1)	picri (1)
épisrie (1)	épercié (1)		ibisiri (1)	
	epefferi (1)		ipiseri (1)	
	epefferies (1)		picsirie (1)	
	epicerie (1)			
	epésserie (1)			
	éphisserie (1)			
	epicaïri (1)			
	épi ssiri (1)			
	épice (1)			
	epiceraie (1)			
	épiceri (1)			
	eipiceraï (1)			
	epicir (1)			
	epicire (1)			
	epiciri (1)			
	eppure (1)			
	épicide (1)			
	épicerie (1)			
	epircérie (1)			
	et pusseris (1)			
	ale pissiri (1)			
	épicerie (1)			
	epicérie (1)			
	épisséure (1)			

	ipecerie (1)			
	ibsouri (1)			
	ibsori (1)			
	éupece (1)			
	èpicierie (1)			
	ipicierie (1)			
	épecerie (1)			
	ipicerie (1)			
	ipisseri (1)			
	ipsore (1)			
	les pesrir (1)			
	pisri (1)			
22	54	16	13	10

Tableau 18. Relevé des graphies orthographiquement non conformes du mot ÉPICERIE chez les 441 enquêtés « bas », classées selon la restitution du phonème /s/ et selon le nombre d'erreurs phonographiques.

Dans le corpus « bas », 115 enquêtés (27%¹³³) n'ont pas réussi à restituer la forme orthographique du mot ÉPICERIE¹³⁴ et ont proposé une forme que nous avons qualifiée de phonographiquement acceptable, d'approximante ou de non acceptable d'un point de vue phonographique.

Seules 22 graphies (colonne 1) parmi les réponses erronées restituent la forme phonographique de l'item, comme dans *epiceri* (8), *episserie* (4). 16 graphies (colonne 3) sont qualifiées d'approximantes à cause de la valeur phonologique du graphème choisi pour transcrire le phonème /s/ : ex. *episerie* (8), et *piseri* (1) (/epizəri/).

L'observation de la restitution du phonème /s/ montre que parmi toutes les 115 réponses orthographiquement non conformes, 70 enquêtés (soit 61% des 115 qui ont commis une erreur) ont néanmoins proposé un graphème plausible « ç », « c » ou « ss » : *epiçerie*, *epiceri*, *episserie* (phonographie acceptable) ou *epecerie*, *ipisseri* (phonographie non acceptable).

Les enquêtés qui n'ont pas réussi à proposer un graphème permettant de restituer le phonème /s/ attendu ont majoritairement utilisé le graphème « s » sans prendre en compte

¹³³ Le pourcentage a été calculé en fonction du nombre de réponses recueillies (et non en fonction du nombre de dictées).

¹³⁴ Nous rappelons que pour cet item, nous n'avons pas tenu compte de la présence d'un accent sur le « e » à l'initiale (voir chapitre 4 dédié à la méthodologie, section 3).

la règle de position de ce graphème, qui entouré de deux voyelles, prend la valeur phonologique /z/ : par exemple dans *episerie*.

Ainsi les graphèmes utilisés (« s », « ss », « c », « ç ») pour transcrire /s/ reproduisent majoritairement deux phonèmes : le /s/ attendu et le /z/ lorsque le choix du graphème n'a pas tenu compte du contexte intervocalique. Quelques enquêtés ont omis le « -e » muet après le graphème « c », transformant ainsi la valeur phonologique de /s/ en /k/ : *epicri* que nous lisons /epiKri/ alors que nous avons comptabilisé la graphie *épissri* que nous lisons /episri/ comme phonographiquement acceptable.

Finalement, la majorité des erreurs phonographiques observées sur l'item ÉPICERIE se porte sur d'autres zones de ce mot : en effet, parmi les 115 réponses non conformes (d'un point de vue orthographique), seules 16 graphies (colonne 3) ne comportent comme seule erreur phonographique que la graphie du phonème /s/ : *episerie* (8), *episeri* (4), *episerie* (2), *épiserie* (1), *et piseri* (1). Les autres erreurs sont observées sur chacun des phonèmes, même si certains phonèmes semblent poser davantage problème que d'autres :

- /i/ à l'initiale : *ipiseri* ;
- /p/ : *ephicerie*, *ibsouri* ;
- /i/ : *epecerie*, *epucrai* ;
- /ə/ : *ipicierie*, *epeiseire* ;
- /i/ en finale : *epcorio*, *epucrai*).

Nous observons enfin quelques graphies très éloignées de la forme phonographique attendue et qui se concentrent dans la 5^e et dernière colonne, parmi les graphies qui présentent une autre valeur phonologique que /s/ ou /z/ pour le phonème /s/ attendu : *picherie*, *epècrer*, *epcorio*, *epucrai*, *et piro*, etc. Ces graphies révèlent des difficultés d'ordre phonographique très importantes, qu'il faudrait analyser parallèlement au profil de l'enquêté.

LES GRAPHIES ORTHOGRAPHIQUEMENT NON CONFORMES RELEVÉES SUR LIBRAIRIE

Nous observons à présent les réponses fournies pour transcrire LIBRAIRIE, qui fait partie des mots ayant posé le plus de difficultés aux enquêtés « bas » d'un point de vue orthographique mais également phonographique.

Le tableau 19 permet de détailler les erreurs qui ont été produites sur cet item. La forme de ce tableau est différente de celle du tableau précédent, car à la différence du mot ÉPICERIE, dont la principale difficulté correspond à la transcription du phonème /s/, pour LIBRAIRIE nous cherchons à rendre compte des difficultés rencontrées par les scripteurs et qui ne semblent

pas centrées autour d'une difficulté en particulier. Dans ce tableau, les graphies ont été classées par rapport au critère phonographique : phonographie acceptable (colonne 1) et phonographie non acceptable (colonnes 2 et 3). Pour faciliter l'analyse, les graphies qui ne sont pas acceptables ont été classées selon le nombre d'erreurs : 1 phonème erroné (colonne 2) et 2 phonèmes erronés ou plus (colonne 3).

1	2	3
phonographie acceptable	phonographie non acceptable, 1 phonème erroné	phonographie non acceptable, 2 phonèmes erronés ou plus
librerie (43)	liberie (6)	librire (3)
librerie (11)	librarie (5)	lebriri (2)
libreri (10)	libririe (4)	livrirer (1)
librairi (6)	libraire (3)	lévarie (1)
libréri (4)	libairie (2)	les brire (1)
librière (2)	library (2)	luiprerie (1)
librairis (1)	libriri (2)	libererie (1)
l'hibrerie (1)	livrairie (2)	libere (1)
libraierie (1)	l'iberie (1)	colebé (1)
l'ibrairie (1)	librere (1)	libourie (1)
	librére (1)	libeuri (1)
	librairée (1)	liberiryai (1)
	lebrairie (1)	libracrie (1)
	librérié (1)	elrairie (1)
	libreuri (1)	librier (1)
	laibrerie (1)	libregi (1)
	libririt (1)	epririe (1)
	liprerie (1)	libre (1)
	lebrerie (1)	libraiser (1)
	lirerie (1)	epréi (1)
	livreri (1)	l'abirerie (1)
		imprepie (1)
		libreririe (1)
		les birire (1)
		librille (1)
		libriré (1)
		lébraire (1)
		librri (1)
		liphagi (1)

		l'imprimerie (1)
		liperi (1)
		lipriri (1)
		livrarir (1)
		livraraire (1)
		lubierie (1)
		livreire (1)
		mipari (1)
80	39	40

Tableau 19. Relevé des graphies orthographiquement non conformes du mot LIBRAIRIE chez les 441 enquêtés « bas » classées selon le nombre d'erreurs phonographiques : phonographie acceptable, 1 erreur phonographique, 2 erreurs phonographiques ou plus.

Parmi les enquêtés « bas », 257 scripteurs ont orthographié correctement l'item LIBRAIRIE (soit 57%), au rang 14 sur 16, avant RHUME et SOLENNEL (voir figure 13). Parmi les réponses erronées (d'un point de vue orthographique), 80 graphies (soit 19% des réponses) sont considérées comme acceptables du point de vue phonographique (aucune graphie n'a été classée dans les graphies dites approximantes) et 79 graphies ne sont pas phonographiquement acceptables car elles ne permettent pas de restituer la forme phonologique de l'item : *liperi*, *lubierie*, *librairée*, etc.

Parmi les graphies phonographiquement acceptables, la principale erreur porte sur le choix du graphème pour transcrire le phonème /e/ : *librerie* transcrit « e » avec ou sans accent (56 occurrences sur 80). Les autres erreurs portent sur la finale muette avec l'absence de « -e » muet : *librairi* (6) et une erreur sur le choix de la finale muette : *librairis*. 14 graphies cumulent absence de finale muette et erreur sur le choix du graphème pour transcrire le phonème /e/ : *libreri* avec ou sans accent.

Parmi les graphies phonographiquement non acceptables, nous n'observons pas de récurrence dans le type d'erreurs et nous observons des omissions (*liberie*) ou des ajouts de phonèmes (*libracrie*) ainsi que des erreurs dans le choix des graphèmes qui ne restituent pas le phonème attendu (*lipriri*).

Parmi les erreurs qui ne portent que sur un phonème :

- 9 enquêtés ont omis le phonème /R/ qui suit /b/ : *liberie* (6), *libairie* (2), *l'iberie* (1) ;
- 13 enquêtés ont choisi un graphème qui ne restitue pas le phonème /e/ : « i » dans *libriri* (2), *libirie* (4), *libirit* (1), « a » (ou omission du « i » dans la graphie de « ai » dans *librarie* (5) et « eu » dans *libreuri* (1).

Parmi les graphies qui comportent deux erreurs phonographiques ou plus, nous relevons des formes très éloignées de l'item : par exemple *mipari* (1), *liphagi* (1), *imprepie* (1), *colebé* (1) et qui marquent des difficultés très importantes de la part de certains enquêtés.

Nous relevons également des graphies qui se rapprochent du mot LIVRE : *livrairie* (2), *livreri* (1) et du mot LIRE : *lirerie* (1). Ces formes présentent une proximité phonologique avec l'item demandé et suggèrent une connaissance encyclopédique associée à ce mot.

Ainsi, nous ne décelons pas de récurrence dans le type d'erreurs parmi les graphies qui ne permettent pas de restituer la forme phonologique de l'item LIBRAIRIE ; celles-ci sont variées et semblent être le reflet de difficultés propres à chaque enquêté. En revanche, parmi les graphies qui fournissent une forme lisible de l'item, les erreurs semblent davantage refléter une ou plusieurs difficultés propres à l'item (en lien avec son degré de consistance), notamment :

- le choix du graphème « ai » qui n'est pas le graphème le plus fréquent parmi les différentes possibilités de transcription du phonème /e/ avec 56 occurrences sur 80, soit « e » ou « é » pour « ai » chez 70% des enquêtés « bas » qui ont fourni une graphie phonographiquement acceptable ;
- la finale muette¹³⁵ : sur ce point, nous observons néanmoins parmi toutes les réponses qui proposent le phonème /i/ en finale¹³⁶ (135), la présence de 99 mots qui présentent un « -e » muet final (et 2 graphies qui utilisent un « -s » ou un « -t » final).

Ce sont ainsi 75% des enquêtés parmi ceux qui n'ont pas fourni la forme orthographique attendue qui ont néanmoins marqué la finale muette, parfois même lorsque la graphie proposée comporte plusieurs erreurs phonographiques : ex. *libererie*, *l'imprimerie*, *lubierie*. Ces résultats suggèrent des connaissances de type morphographique avec l'enregistrement de régularités chez les enquêtés les plus en difficulté avec l'écriture.

2.3.2.3 FOCUS SUR LES SCRIPTEURS EN GRANDE DIFFICULTÉ EN PHONOGRAPHIE

Nous nous centrons ici sur les scripteurs qui semblent avoir de grandes difficultés en phonographie. Pour cette analyse, nous observons les graphies du mot LIBRAIRIE qui contiennent plus d'une erreur phonographique (colonne 3 du tableau 19) Nous continuons

¹³⁵ Une analyse qui porte spécifiquement sur le « -e » muet final de cet item est proposée section 2.5.1.

¹³⁶ Ce chiffre comprend les 80 graphies phonographiquement acceptables auxquelles on ajoute 55 graphies parmi les réponses phonographiquement non acceptables (ex. *lebrerie*, *l'imprimerie*, *lipriri*).

cette description des graphies non conformes sur cet item parce qu'il fait partie de ceux qui ont été particulièrement mal restitués d'un point de vue orthographique mais aussi phonographique. En outre, il comporte deux difficultés inhérentes au système orthographique du français : la présence d'un graphème qui entre en concurrence avec d'autres graphèmes plus fréquents permettant de transcrire le même phonème (ici « ai » pour transcrire /e/) et la présence d'un « -e » muet final.

Pour chaque réponse non conforme, nous observons 6 variables issues du module biographique :

- l'âge de l'enquêté au moment de l'enquête ;
- sa langue maternelle (parlée vers l'âge de 5 ans) ;
- la langue avec laquelle il a appris la lecture ;
- le niveau d'études (variable qui nous permet de repérer les personnes qui n'ont jamais été scolarisées) ;
- l'âge d'arrivée en France pour les enquêtés nés à l'étranger.

Le tableau 20 présente les variables ainsi que la graphie proposée pour transcrire l'item LIBRAIRIE. Les graphies sont restituées telles qu'observées dans les dictées manuscrites (majuscules ou non, accentuation).

Les données sont classées par ordre de numéro d'enquêté.

n° enquêté	âge	langue maternelle	langue apprent. lecture	âge arrivée en France	graphie proposée
1556	64 ans	Turc	Turc	23 ans	liphagi
1563	55 ans	Français Ardéchois ¹³⁷	Français		epréi
1567	44 ans	Portugais	Portugais	21 ans	lévarie
1575	30 ans	Arabe	Arabe	24 ans	librire
1606	35 ans	Turc	Non renseigné	22 ans	libregi
1607	43 ans	Dialecte Eyoundo ¹³⁸	Français	36 ans	l'abirerie
1608	56 ans	Arabe	Arabe	25 ans	lebriri

¹³⁷ Nous relevons ce qui est indiqué dans le fichier fourni par l'INSEE. L'enquêté fait peut-être référence à un sous-dialecte de la langue d'oc, qui prendrait la forme d'un patois Ardéchois. Nous retenons de cette précision que pour l'enquêté, sa langue maternelle n'est pas celle qu'il a pratiquée à l'école.

¹³⁸ L'enquêté fait probablement référence à l'Ewondo, qui est une langue parlée dans la partie centrale et dans la partie sud du Cameroun. Nous précisons que les deux langues officielles au Cameroun sont le français et l'anglais. Cependant, les différentes régions du Cameroun sont majoritairement francophones.

1610	38 ans	Arabe	Arabe	30 ans	les brire
1615	65 ans	Arabe	Arabe	34 ans	librille
1624	25 ans	Turc	Turc	20 ans	Liprerie
1636	50 ans	Arabe	Arabe	30 ans	Mipari
1639	54 ans	Français	Français		lubierie
1647	42 ans	Français	Français	7 ans	lébraire
1660	65 ans	Français	Français		imprepie
1666	48 ans	Français	Français		L'IMPRERIE
1746	58 ans	Arabe	Français	25 ans	liperi
1747	43 ans	Arabe	Arabe	28 ans	livrirer
1748	29 ans	Italien	Français		lirerie
1756	43 ans	Portugais	Portugais	24 ans	Librier
1760	41 ans	Arabe	Arabe	20 ans	LIPRIRI
1763	62 ans	Français	Français	12 ans	Colebé
1766	39 ans	Portugais	Portugais	35 ans	livrarir
1790	41 ans	Arabe	Arabe	34 ans	libererie
1799	65 ans	Français	Français		libourie
1802	44 ans	Français	Français		Luiprerie
1805	51 ans	Anglais	Anglais	27 ans	Livraraire
1810	48 ans	Arabe	Arabe	29 ans	les birire
1822	34 ans	Arabe	Arabe	13 ans	libriré
1849	60 ans	Français	Français		Libere
1887	42 ans	Italien	Français		lebriri
1904	57 ans	Français	Français		Libre
1905	22 ans	Français	Français	13 ans	elrairie
1906	46 ans	Arabe	Français	20 ans	epririe
1907	63 ans	Berbère	Arabe	30 ans	livreire
1923	24 ans	Arabe	Arabe	15 ans	libraiser
1960	29 ans	Arabe	Arabe	17 ans	LIBRIRE
1974	41 ans	Français	Français		libr?
1994	35 ans	Français	Arabe	25 ans	LIBRIRE
2015	30 ans	Arabe	Arabe	2 ans	LIBRRI
2046	59 ans	Français	Français		Libracrie
2047	41 ans	Français	Français		libeuri
2048	41 ans	Français	Arabe	16 ans	liberiryai

Tableau 20. Relevé des 40 graphies de LIBRAIRIE qui contiennent plus d'une erreur phonographique chez les 441 enquêtés « bas ».

Dans ce corpus, nous trouvons deux langues maternelles dominantes : le français (16 graphies soit 40% des scripteurs) et l'arabe (16 graphies soit 40%). Ainsi, pour 60% des enquêtés qui ont produit deux erreurs phonographiques ou plus sur LIBRAIRIE, le français n'est pas leur langue maternelle. Par ailleurs, seuls trois d'entre eux sont arrivés en France avant l'âge de 16 ans (enquêtés 1822, 1923, 2015 ; les deux premiers arrivés, à 13 ans et 15 ans, ayant peut-être bénéficié d'une scolarisation en France). Pour ces enquêtés, les difficultés à l'écrit semblent davantage relever d'une difficulté d'apprentissage du français qui n'est pas leur langue maternelle.

Parmi les enquêtés dont le français est la langue maternelle, trois indiquent être arrivés en France à l'âge de 7 ans (enquêté 1647), 12 ans (enquêté 1763) et 13 ans (1905). Les 11 autres enquêtés (soit 28% de ceux qui ont produit plus de deux erreurs phonographiques), nés et scolarisés en France, étaient âgés de 41 à 60 ans au moment de l'enquête. Nous retrouvons ici l'effet de génération décrit par Jonas (2012). Leurs difficultés orthographiques et surtout phonographiques sur cet item (ex. *libeuri*, *libracrie*, *libere*) prédisent des difficultés en production de l'écrit suffisamment importantes pour restreindre leur capacité à produire un message lisible, ce qu'il faudrait confirmer en observant le reste de la dictée.

Ces variables ne présentent qu'un aperçu des facteurs pouvant expliquer certaines graphies particulièrement éloignées de la forme attendue. Par ailleurs la taille du corpus, avec 40 sujets, ne permet pas non plus de rendre compte de profils. Nous soulignons seulement ici que les difficultés orthographiques mais également phonographiques sont multifactorielles et qu'elles sont aussi en adéquation avec la diversité des profils des scripteurs du module « bas ».

2.3.3. BILAN SUR LA PHONOGRAPHIE

À travers l'analyse de la phonographie des mots et des pseudo-mots de la dictée, nous cherchons à mieux distinguer les composantes de la maîtrise de l'orthographe. Nous valorisons ainsi la capacité des scripteurs à coder les phonèmes en proposant une forme graphique du pseudo-mot ou du mot, conforme au système phonographique du français. Cette analyse nous permet de distinguer la proportion d'enquêtés capables de restituer la forme phonologique des mots dictés, de ceux pour qui les connaissances phonographiques sont insuffisantes pour passer du canal oral au canal écrit. Les compétences phonographiques représentent en effet une étape intermédiaire et essentielle avant l'acquisition des conventions d'écriture (Fayol, 2013).

La prise en compte du critère phonographique pour analyser les graphies des enquêtés non conformes aux attentes orthographiques montrent que dans ce domaine, la majorité d'entre eux possède des connaissances suffisantes pour proposer une forme graphique lisible. La lisibilité de ces formes dépend notamment du respect des règles de position et de distribution du français et du choix de graphèmes fréquents pour pallier un manque de connaissance orthographique sur un item en particulier (par exemple, le graphème « e » pour transcrire /e/ dans LIBRAIRIE. Cette lisibilité qui permet de facilement rapprocher la forme graphique proposée (par exemple *librerie*) de la forme normée LIBRAIRIE est essentielle pour permettre au lecteur d'accéder au message facilement. À l'inverse, nous avons vu que certaines graphies considérées comme approximantes (parce qu'elles renvoient à une transcription phonétique des mots) posent quelques difficultés. En effet, certains problèmes de correspondances entre phonèmes et graphèmes peuvent faire naître un sentiment d'étrangeté au niveau visuel qui peut gêner le déchiffrement du mot comme la graphie *saile* pour transcrire SEL ou *solanaille* pour SOLENNEL.

Pour certains mots, il apparaît que la prise en compte du critère phonographique fait augmenter les résultats de manière très importante : ainsi sur SOLENNEL (qui est un item qui a particulièrement posé problème aux enquêtés « bas » avec seulement 19% de graphies correctes), la prise en compte du critère phonographique sur la réussite fait augmenter la restitution de 48 points, avec 67% de graphies acceptables. Sur l'ensemble des mots, le score moyen de réussite lorsque l'on prend en compte le critère phonographique passe de 72% à 85%. Si l'on compare ce taux de réussite avec celui observé sur les pseudo-mots (77% en moyenne, avec 4 pseudo-mots - GOBAR, JOC, DUX et PÉVANOR - sur 6 qui se situent en dessous de la moyenne et du score médian), il semble que le fait d'écrire des mots connus (vs des mots inventés) permette une meilleure restitution de la forme phonographique attendue, même lorsque les enquêtés n'en connaissent pas l'orthographe.

Le tableau 21 récapitule les compétences des enquêtés observées dans les mots de la dictée à partir de deux critères : le critère orthographique et le critère phonographique.

Le score est calculé d'après la restitution orthographique et/ou phonographique sur les 16 mots de la dictée¹³⁹ (nous avons retiré les pseudo-mots et n'avons pas pris en compte la morphographie). Nous répartissons les réponses en trois catégories :

¹³⁹ PHARMACIE - ANTI - RHUME - ALCOOL - ÉPICERIE - TOMATE - PAYS - SEL - CONFITURE - CERISE - FROMAGE(S) - SENTE(NT) - FORT - LIBRAIRIE - FIANCÉ - SOLENNEL.

- le score orthographique ;
- le score phonographique, qui comprend les graphies qui tiennent compte des règles de position et de distribution ;
- le score phonographique dit « approximant » pour les graphies qui proposent une transcription phonétique.

Le pourcentage de réussite a été calculé d'après trois niveaux de réussite :

- plus de 80% de réussite sur les 16 mots analysés (score compris entre 80 et 100% de réussite) ;
- 1 erreur relevée au maximum (score de 1 erreur ou 100% de réussite) ;
- aucune erreur relevée : 100% de réussite.

	score orthographique	score phonographie acceptable	score phonographie approximante
plus de 80% de réussite	33%	60%	65%
1 erreur	13%	19%	22%
100% réussite	8%	22%	28%

Tableau 21. Proportion de réussite observée sur 16 mots en fonction du type de critère : orthographique ou phonographique et du niveau de réussite attendu chez les 441 enquêtés « bas ».

D'après l'analyse des compétences en orthographe lexicale des enquêtés « bas » réalisée sur 16 mots de la dictée d'IVQ, il apparaît que 33% des enquêtés « bas » ont fait au maximum 3 erreurs (plus de 80% de réussite) et n'éprouvent pas de difficultés majeures lorsqu'il s'agit de retrouver l'orthographe de mots dictés. 13% n'ont fait qu'une seule erreur au maximum et 8% ont réussi à restituer l'orthographe des 16 mots analysés.

L'analyse phonographique des réponses erronées sur le plan orthographique montre que 60% des enquêtés ont produit une forme phonographiquement acceptable sur plus de 80% des mots. Lorsque l'on prend en compte le critère phonographique dans une acception large, ce résultat passe à 65% et il apparaît même que 50% des enquêtés ont réussi à transcrire phonétiquement les 16 items, en ne produisant qu'une seule erreur.

En revanche, les 35% d'enquêtés qui n'ont pas réussi à transcrire au moins phonétiquement les 80% des 16 mots dictés sont en grande difficulté face à l'écriture : leurs connaissances en phonographie sont insuffisantes pour passer du canal oral au canal écrit. Une analyse précise de leur profil permettrait peut-être de déterminer la nature de leurs difficultés (notamment si le français a fait l'objet d'un apprentissage à l'école, en France ou à l'étranger et pendant combien d'années).

Nous allons à présent nous intéresser aux lettres finales, qui représentent une difficulté emblématique du système orthographique du français.

2.4. COMPÉTENCES EN MORPHOGRAPHIE OBSERVÉES SUR LES MOTS

2.4.1. SCORES DE RÉUSSITE EN MORPHOGRAPHIE FLEXIONNELLE DU NOMBRE

Nous observons dans cette partie, les compétences en morphographie flexionnelle du nombre chez les enquêtés « bas » à partir des possibilités offertes dans la dictée d'IVQ : deux accords en nombre (le « -s » de FROMAGES et le « -nt » de SENTENT) et sur l'absence de marque sur l'adverbe FORT.

Les marques de nombre erronées sont également comptabilisées : parmi les réponses phonographiquement correctes, nous relevons 6 occurrences de *sentes* (pour SENTENT) qui montrent une confusion entre les différents graphèmes qui marquent le pluriel sur le verbe et sur le nom (« -nt » et « -s »). Cette erreur est également relevée sur les formes erronées sur le plan orthographique : *santes* (1 occ.) et sur les formes phonographiquement non acceptables : *sents* (1) et *sens* (1). Nous avons ainsi pris en compte les 9 marques de nombre erronées comme preuve de compétence en morphographie, car seule la présence d'un accord nous intéresse ici.

Les taux de réussite présentés dans la figure 14 (accord en nombre de FROMAGES et de SENTENT ainsi que l'absence de marque sur FORT) ne concernent donc que la morphographie. Ils ont été calculés en fonction du nombre de réponses recueillies¹⁴⁰ (et non en fonction du nombre d'enquêtés). Les items sont présentés du mieux réussi d'un point de vue morphologique (flexionnel) au moins bien réussi.

¹⁴⁰ Ce pourcentage est calculé d'après le nombre de réponses réelles. Sur ces items, le nombre de non-réponses est particulièrement élevé : 25 sur FROMAGES et 42 sur SENTENT (soit 33 non-réponses en moyenne sur les deux items) ainsi que 31 non-réponses sur FORT.

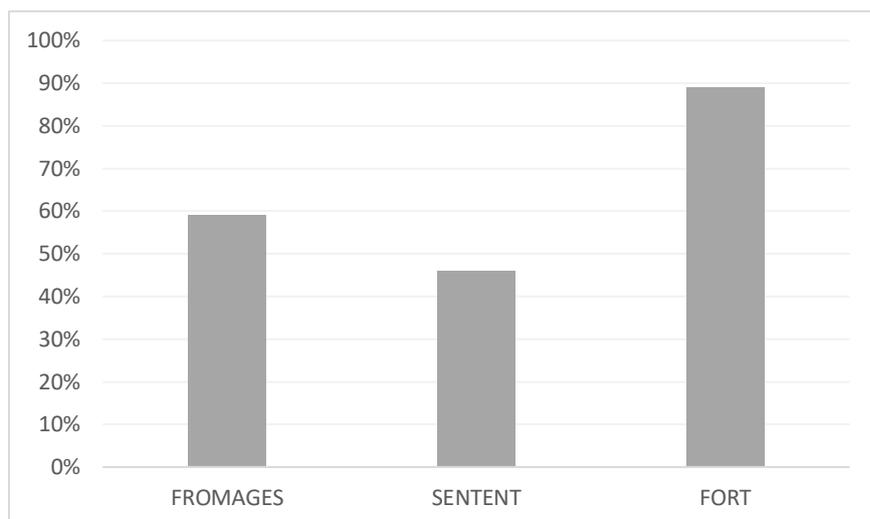


Figure 14. Réussite en morphographie flexionnelle du nombre observée sur trois items chez les 441 enquêtés « bas ».

Chez les enquêtés « bas », c'est l'absence de marque de nombre sur l'adverbe FORT (89%) qui a été la mieux respectée par les enquêtés « bas ». En revanche, les marques de nombre sur le nom FROMAGES (59%) et le verbe SENTENT (46%) ont été plutôt mal restituées : il semble plus facile de produire une forme non marquée qu'une forme marquée.

Pour FORT, c'est l'absence de marque qui correspond à une réussite. Nous ne pouvons donc déterminer si ceux qui ont réussi ont eu conscience de ne pas devoir marquer le nombre sur un adverbe. Cet exemple pourrait constituer un indicateur de compétence s'il pouvait être comparé avec l'accord en nombre de l'item utilisé comme adjectif : ex. TROIS FROMAGES QUI NE SENTENT PAS FORT (FORT = adverbe) vs TROIS FROMAGES FORTS (FORTS = adjectif).

L'analyse des réponses montre que 169 enquêtés « bas » (soit 38%) n'ont réussi aucune marque de nombre (sur le nom et le verbe) et 121 (27%) n'ont réussi qu'une seule des deux marques. Lorsque les enquêtés n'ont réussi qu'un seul accord, c'est surtout celui du nom FROMAGES qui est davantage maîtrisé (92 marques réussies contre 29 pour SENTENT). Enfin, seuls 157 enquêtés « bas » ont réussi les deux marques de nombre (36%).

Ainsi, dans le corpus « bas », il y a presque autant d'enquêtés qui ont réussi les deux marques de nombre (36%), que d'enquêtés qui n'en ont réussi aucune (38%). La morphographie semble ainsi être un problème pour plus de la moitié des enquêtés « bas ». Pour voir si la maîtrise de la morphographie prédit les compétences en orthographe lexicale, nous observons dans la figure 15, le score global obtenu sur la dictée par les enquêtés qui ont réussi les deux marques de nombre de la dictée d'IVQ : FROMAGES et SENTENT.

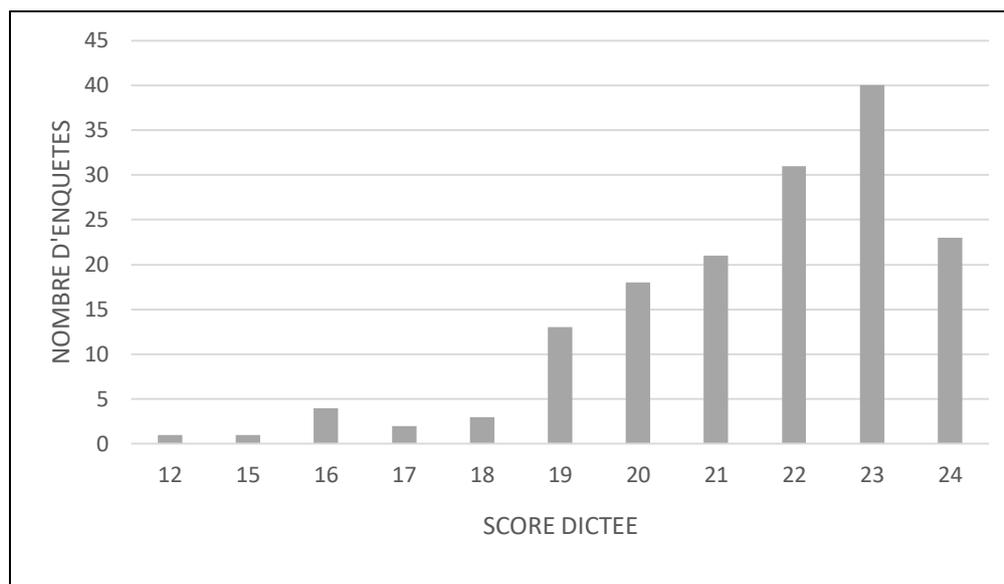


Figure 15. Répartition des scores obtenus sur la dictée d'IVQ par les 157 enquêtés « bas » qui ont marqué le nombre sur FROMAGES et sur SENTENT.

Dans le corpus « bas », 157 enquêtés (soit 36%) ont marqué le nombre sur les deux items qui sont utilisés au pluriel dans la dictée (FROMAGES et SENTENT).

Nous rappelons que l'analyse des dictées de notre échantillon montre que 59% des enquêtés « bas » ont marqué le nombre sur FROMAGES et 46% sur SENTENT.

Parmi ceux qui ont réussi les deux marques de nombre :

- 133 soit 85% (ce qui représente 32% des enquêtés « bas ») ont obtenu un score moyen de réussite sur l'intégralité de la dictée, supérieur à 80% (avec une note comprise entre 20 et 24 sur 24) ;
- 63 enquêtés soit 40% (ce qui représente 15% des enquêtés « bas ») ont fait au maximum une erreur sur la dictée (avec une note de 23 ou 24 sur 24) ;
- 39 soit 25% ont obtenu une note comprise entre 21 et 22 sur 24, et ont donc fait 3 ou 4 erreurs sur la dictée ;
- Enfin, seuls 11 enquêtés parmi ceux qui ont réussi les deux marques de nombre, ont obtenu un score global inférieur à 19 sur 24 et l'enquêté qui a obtenu la note la plus basse (12 sur 24) obtient néanmoins la moyenne.

L'accord en nombre du nom FROMAGES et celui du verbe SENTENT semble donc constituer un bon indicateur pour distinguer les personnes qui ont des difficultés en orthographe puisque ceux qui ont réussi la morphographie ont également mieux réussi la dictée. Ces deux accords permettent aussi de distinguer les enquêtés qui semblent avoir de grandes difficultés en morphographie mais ils ne suffisent pas, en revanche, pour cerner les compétences dans ce

domaine des personnes qui ont de meilleures connaissances orthographiques. Il aurait en effet été intéressant de davantage tester les compétences en morphographie des enquêtés « bas » qui ont réussi les deux accords en nombre.

2.4.2. LES AUTRES TRACES DE MORPHOGRAPHIE OBSERVÉES DANS LES DICTÉES

Pour davantage rendre compte du comportement des enquêtés « bas » par rapport à la morphographie, nous observons les traces de morphographie relevées sur les autres items de la dictée. Dans cette analyse, nous observons les marques de nombre présentes sur les mots qui ont été dictés sans déterminant. Dans cette analyse, nous prenons en compte toutes les réponses des enquêtés, y compris celles qui ne restituent pas la forme phonologique des items, seules les marques de morphologie nous intéressant.

Dans cet exercice, certains mots dictés sans déterminant appellent implicitement une marque de nombre : (DES) TOMATES – (DES) CERISES. D'autres en revanche ne demandent pas de pluriel parce qu'ils font référence au nom d'un rayon : (RAYON) PHARMACIE, (RAYON) ÉPICERIE, (RAYON) LIBRAIRIE ou à une quantité : (DE L') ANTI-RHUME, (DE L') ALCOOL, (DE LA) CONFITURE, (DU) SEL. Nous résumons cette observation des marques de nombre par « marques implicites ». Dans un deuxième temps, nous observons également les marques de nombre et de genre ajoutées de façon erronée dans l'énoncé LE FIANCÉ SOLENNEL.

Exemples de marque de morphographie relevées dans les dictées :

- enquêté 1972 : tomates, confiture, cerises, sels, trois fromages qui ne sentent pas fort ;
- enquêté 1978 : tomates, confiture, cerises, sel, trois fromage qui ne sent pas fort ;
- enquêté 2045 : tomate, confiture, cerise, sel.

2.4.3. LES MARQUES DE NOMBRE IMPLICITES

Sur les noms de rayon, nous observons peu de traces de pluriel dans les réponses, avec seulement 7 graphies qui ont été orthographiées au pluriel et que nous présentons ci-dessous :

- PHARMACIE : *pharmaces* (1 occ.) ;
- ÉPICERIE : *episseris* (1), *epesseries* (1), et *pusseris* (1) ;
- LIBRAIRIE : *librairies* (3).

Nous observons ensuite l'ajout d'une marque de nombre sur les items qui font référence à une quantité (ANTI-RHUME – ALCOOL – CONFITURE – SEL) :

ANTI-RHUME : nous n'avons relevé qu'une seule marque de nombre sur ANTI et 2 sur RHUME. L'enquêté qui a ajouté un « s » à ANTI n'a pas ajouté de « s » à RHUME : *antis-rhume* (1) et *anti-rhumes* (2). Ces trois enquêtés ont par ailleurs réussi les marques de nombre attendues sur le nom FROMAGES et le verbe SENTENT.

ALCOOL : nous ne trouvons qu'une seule forme *alcols* chez un enquêté qui n'a pas réalisé les accords sur FROMAGES et SENTENT.

SEL : nous observons 4 marques de nombre : *sels* (3), *cels* (1). Parmi les enquêtés qui ont produit ces graphies, 3 ont réalisé les accords attendus et un (qui a choisi la graphie *sels*) n'en a réalisé aucun.

CONFITURE : nous observons en revanche 28 marques de nombre sur cet item : *confitures* (26), *comfitures* (1) et *confitures* (1).

Nous observons ensuite le comportement des scripteurs sur deux mots, sur lesquelles la présence d'une marque de nombre n'est pas attendue, mais souhaitable.

CERISES : cet item est celui sur lequel il paraît le plus souhaitable d'ajouter une marque de nombre, que l'enquêté ait compris qu'il s'agisse de confiture de cerises (CERISES vient juste après CONFITURE dans la dictée) ou d'acheter des cerises (nous n'achetons pas une seule cerise). L'analyse des réponses montre que 144 enquêtés (soit 34%) ont choisi de mettre cet item au pluriel. La répartition des réponses qui comportent une marque de nombre est la suivante :

- formes phonographiquement acceptables : *cerises* (130), *se rises* (1), *ceurises* (1), *cerizes* (1) ;
- formes phonographiquement non acceptables : *cerices* (8), *cherries* (1), *cérisés* (1), *cerecies* (1).

TOMATES : si 143 enquêtés ont choisi la graphie *tomates*, nous n'observons que 4 graphies parmi les réponses non conformes à la norme orthographique qui portent une marque de nombre sur cet item : *tomats*, *tomâtas*, *tonates*, *tomatos*.

Si nous n'observons que peu de marques de nombre sur le nom des rayons et sur les noms qui indiquent une quantité (ANTI-RHUME, SEL), cette analyse montre qu'une proportion non négligeable d'enquêtés (jusqu'à 34% d'entre eux pour les items TOMATES et CERISES) a choisi de mettre certains items au pluriel alors que ceux-ci ont été dictés sans déterminant.

Il apparaît également que lorsque le nombre doit être marqué, un signal clair représenté par un déterminant semble représenter une aide pour les scripteurs : 34% de « -s » sur CERISES vs 59% sur (TROIS) FROMAGES. L'observation du comportement des scripteurs qui n'ont pas de difficulté particulière à l'écrit (corpus des enquêtés « haut ») sur ces items, nous permettra de voir si eux-aussi ont eu conscience des marques implicites de nombre attendues et si l'absence d'un signal clair pour l'indiquer représente, comme chez les enquêtés « bas », une difficulté.

Si la majorité des enquêtés « bas » n'a effectivement pas marqué le nombre lorsque sa présence était implicitement demandée (notamment sur CERISES), l'absence de marque sur les items faisant référence à une quantité ou sur les noms de rayon semble néanmoins montrer une conscience des règles de fonctionnement en matière de morphographie. Même si sur ces items, c'est l'absence de marque qui correspond à une réussite, la proportion d'absence de marques est bien plus importante (une seule marque de nombre relevée sur ALCOOL et 4 sur SEL) que sur les items sur lesquels le nombre était attendue (comme FROMAGES qui ne porte pas de marque dans 41% des graphies).

Par ailleurs, les exemples de graphies produites par quelques enquêtés (voir ci-dessus) montrent un comportement pluriel des enquêtés vis-à-vis de la présence ou de l'absence de marques de nombre. Les enquêtés qui ont réussi les deux marques attendues n'ont pas forcément marqué les pluriels implicitement attendus ou ont pu marquer le nombre sur des noms faisant référence à une quantité (ex. enquêté 1972). Enfin, les enquêtés qui n'ont réussi aucune des deux marques attendues ont parfois marqué les pluriels implicites (ex. enquêté 1978).

2.4.4. ERREURS MORPHOGRAPHIQUES RELEVÉES DANS « LE FIANCÉ SOLENNEL »

Nous observons ensuite dans le tableau 22, la morphographie présente dans l'énoncé « LE FIANCÉ SOLENNEL ». Pour FIANCÉ et SOLENNEL, nous présentons les graphies qui ont une marque morphographique, notamment de genre, ainsi que le nombre d'occurrences pour chacune.

FIANCÉ	fiancée (33)	fiancee (2)	fianceé (1)	fiancès (1)	fiancée (1)
	fiançee (1)	fiançaie (1)	fiançés (1)	fiencée (1)	phiancée (1)
	fillancée (1)	fillencée (1)	fianciée (1)	fiaceé (1)	
SOLENNEL	solennelle (6)	solanelle (42)	solannelle (5)	sollanelle (5)	solanèle (3)
	solanèle (2)	solennele (2)	solanelles (1)	sollenele (1)	solanéle (1)
	saulanelle (1)	solanaille (1)	colanelle (1)	solenelle (6)	solonelle (4)
	sonanelle (2)	selon elle (2)	sonaelle (2)	sous la nelle (2)	solanlle (1)
	solanielle (1)	sollenele (1)	sonnelle (1)	sou la nelle (1)	se la nèle (1)
	celenle (1)	ceulanele (1)	selon nelle (1)	solonaile (1)	selanniele (1)
	sonaelle (1)	selanalle (1)	sonalle (1)	sor la nelle (1)	soulanele (1)
	soulanelle (1)	sous lenmenle (1)	sulenelle (1)	celannelle (1)	celemanle (1)
	salennele (1)				

Tableau 22. Détail des graphies qui présentent une marque de genre ou de nombre sur FIANCÉ et SOLENNEL chez les 441 enquêtés « bas ».

Nous présentons d'abord les graphies de FIANCÉ qui présentent une marque de morphographie, puis nous observons les graphies de SOLENNEL.

ANALYSE DES GRAPHIES DE « FIANCÉ » QUI COMPORTENT UNE ERREUR MORPHOGRAPHIQUE

Parmi toutes les propositions de réponses observées sur l'item FIANCÉ, une seule graphie présente une marque de pluriel (*fiancés*). En revanche, nous avons observé 45 graphies qui présentent un « -e » final que l'on peut associer à la marque du genre féminin, même si parfois, l'ordre des graphèmes n'est pas respecté (*fianceé*).

Une grande partie des graphies qui présentent une marque de morphographie (68%), se trouve parmi les réponses conformes à la norme orthographique. En effet, 36 graphies ont un « -e » muet final qui marque le féminin (soit 13% des réponses orthographiquement correctes). Parmi ces 36 graphies, nous observons seulement 4 erreurs de déterminant : LA ayant été utilisé à la place du LE dicté. Nous notons que le corpus des enquêtés « bas » ne comporte que 5 erreurs de ce type et tous les enquêtés qui ont utilisé LA pour LE, ont accordé en genre l'item FIANCÉ (*la fiancée*, 5 occ.).

Parmi les réponses non conformes à la norme orthographique, nous observons seulement 9 marques de féminin dont une présente une erreur sur le déterminant (LA utilisé pour LE) : *la phiancée (1), le fillencée (1), le fiançaié (1), etc.*

ENQUÊTÉS QUI ONT RÉUSSI LES DEUX MARQUES DE NOMBRE ATTENDUES

Pour déterminer les difficultés des scripteurs dans le domaine de la morphographie, nous avons vu que nous ne disposons, dans la dictée d'IVQ, que de deux accords en nombre qui sont attendus et que 157 enquêtés ont réussi ces deux marques (figure 14, section 2.4.1). Nous vérifions alors si ceux qui ont réussi les deux marques de nombre attendues (et qui ne sont donc pas considérés en difficulté dans ce domaine), ont ajouté une marque de genre erronée sur FIANCÉ. Nous regardons également s'ils ont marqué l'accord (de manière erronée) sur l'adverbe FORT.

Nous relevons à présent les graphies qui comportent au moins une erreur morphographique sur FIANCÉ ou sur FORT. Nous soulignons les graphies qui contiennent une erreur en morphographie :

- enquêté 1753 : *Le fiancée solennelle / fort*
- enquêté 1768 : *La phiancée solanel / bon*
- enquêté 1807 : *La fiancée solannelle / forts*
- enquêté 1824 : *Le fiancée solannel / fort*
- enquêté 1827 : *La fiancée solannelle / forts*
- enquêté 2007 : *Le fiancée solennel / forts*

Seuls six scripteurs (soit 4%) parmi les 157 qui ont réussi les deux marques de nombre, ont marqué le genre féminin sur FIANCÉ et la moitié de ces scripteurs ont également fait l'accord avec SOLENNEL, deux sur les trois ayant écrit le déterminant LA à la place de LE. Nous observons également que trois d'entre eux ont accordé de manière erronée le nombre sur l'adverbe FORT¹⁴¹.

¹⁴¹ L'adverbe FORT dans la construction : TROIS FROMAGES QUI NE SENTENT PAS FORT, a pu être confondu avec un adjectif. Il a été accordé en nombre par 11% des enquêtés.

ENQUÊTÉS QUI N'ONT RÉUSSI AUCUNE DES DEUX MARQUES DE NOMBRE ATTENDUES

Nous relevons ensuite les graphies qui comportent au moins une erreur morphographique sur FIANCÉ ou sur FORT chez les 169 enquêtés qui n'ont réussi aucune des deux marques de nombre attendues (sur FROMAGES et SENTENT). Nous soulignons les graphies qui contiennent une erreur de type morphographique. Nous indiquons par un X une absence de mot :

- enquêté 901 : *Le fiancée solanelle / fort*
- enquêté 1584 : *Fiancée solanel / fort*
- enquêté 1609 : *Le fiancée solamaile / fort*
- enquêté 1615 : *Le fiancee sola laine / fort*
- enquêté 1650 : *X fiancee solanielle / forte*
- enquêté 1657 : *Le fiancée solanel / fort*
- enquêté 1665 : *Le fiancée X / fort*
- enquêté 1677 : *Le fiancée solonelle / for*
- enquêté 1756 : *X fianceé solanel / fours*
- enquêté 1788 : *X fianceé solanel / fort*
- enquêté 1805 : *Le fiancée solanel / fort*
- enquêté 1828 : *Le fillancée solanel / fort*
- enquêté 1949 : *Le fiancée solanel / fort*
- enquêté 1962 : *Le fiancée solanel / fort*
- enquêté 1982 : *Le fiancée solanelle / fort*
- enquêté 2002 : *Le fiancée solannel / fort*
- enquêté 2013 : *Le fiancée sona lel / fort*
- enquêté 2020 : *Le finacée solanelle / fort*
- enquêté 2028 : *X fiancée saulanel / fort*
- enquêté 2029 : *Le fiancée solennel / fort*

Parmi les enquêtés « bas » qui n'ont réussi aucune des deux marques de nombre attendues, 20 scripteurs (soit 12%) ont marqué le genre féminin sur FIANCÉ. Nous observons 5 finales « -le » sur SOLENNEL et aucune marque de nombre sur FORT (dans la graphie *fours*, nous ne pouvons interpréter le « -s » comme une marque de nombre).

Il apparaît que les enquêtés qui ont réussi les deux marques de nombre attendues sur le nom et le verbe possèdent une gestion plus efficace des règles d'accord en nombre (observé sur les marques attendues) et en genre puisque seuls six scripteurs (soit 4%) ont marqué le genre féminin sur FIANCÉ, contre 20 scripteurs (soit 12%) des enquêtés qui n'ont réussi aucune des

deux marques attendues. Concernant cet ajout d'un « -e » final qui a pour conséquence de marquer le genre sur FIANCÉ chez quelques enquêtés, nous nous interrogeons sur un éventuel effet de la fréquence de la forme fléchie, sur les productions des enquêtés, par récupération d'instance.

Par ailleurs, sur FORT, parmi les enquêtés qui n'ont réussi aucune des deux marques attendues, aucun n'a marqué l'accord en nombre sur FORT. Cela conforte l'hypothèse selon laquelle il est plus facile de produire une forme non marquée, notamment lorsqu'il s'agit de choisir entre le singulier et le pluriel.

LA GRAPHIE FIANCÉE : UN EFFET DE CONSISTANCE OU DE FRÉQUENCE ?

Pour expliquer l'ajout chez certains enquêtés, d'un « -e » final (qui marque le genre féminin) sur FIANCÉ notamment chez les scripteurs qui ont su restituer la forme orthographique de cet item peu consistant¹⁴² malgré la présence du déterminant LE, nous observons la fréquence et la consistance de la forme fléchie. D'après la base Manulex-Infra¹⁴³, la graphie FIANCÉE est plus fréquente (U = 2,10) que la graphie FIANCÉ (U = 0,46) et un peu plus consistante (62 vs 59).

L'ajout d'une marque de genre sur l'item FIANCÉ, est davantage observé chez les scripteurs qui ont su restituer l'orthographe de cet item, mais également davantage chez les scripteurs qui n'ont pas réussi les deux marques de nombre attendues sur FROMAGES et SENTENT. Il semble donc que le fait de marquer le genre féminin provienne davantage d'un effet de fréquence (FIANCÉE est en effet plus fréquent que FIANCÉ) que d'une volonté de marquer le genre féminin, notamment chez les scripteurs qui ont des difficultés dans le domaine de la morphographie.

ANALYSE DES GRAPHIES DE « SOLENNEL » QUI COMPORTENT UNE ERREUR MORPHOGRAPHIQUE

Sur l'item SOLENNEL, nous n'observons qu'une seule marque de nombre (*solanelles*). En revanche, 111 graphies (26% des réponses) présentent un « -e » final (dont seulement 6, soit 8%, parmi les graphies correctement orthographiées : *solennelle*). Il semble donc que les enquêtés qui ont réussi à orthographier cet item complexe, aient davantage conscience que

¹⁴² Voir chapitre 4 consacré à la méthodologie, section 1.2.1.1 tableau 3 pour une présentation des indices de consistance sur les mots de la dictée.

¹⁴³ Calculs effectués par token, statistiques du CP au CM2, sur le mot complet (consistance Phonologie vers Orthographe - PO totale et fréquence lexicale par million U).

l'ajout d'un « -e » final (voir même l'ajout d'un redoublement de consonne du « -l » suivi du « -e ») marque le genre féminin, qui dans ce cas n'était pas demandé. À ce sujet, nous observons dans EOLE (tableau 5), que si seulement 5% des élèves de CM2 ont réussi à orthographier SOLENNEL, la forme fléchie SOLENNELLE est quant à elle réussie par 19% des élèves (contre 5% des élèves de CM1). Ainsi, s'il est plus facile d'écrire l'adjectif au féminin qu'au masculin en CM2, c'est peut-être que les formes de féminin sont réussies par défaut, l'élève ressentant une incomplétude à écrire une terminaison en « -el ». Cette hypothèse pourrait être confirmée (ou infirmée) si on observe des formes *solennelle* produites quand le masculin est attendu.

À la différence de FIANCÉ, ce sont plutôt les enquêtés qui n'ont pas réussi à orthographier SOLENNEL, qui ont ajouté un « -e » final et nous ne pouvons être certaine que cet ajout provienne d'une volonté de marquer le féminin.

Nous regardons alors si la présence d'un « -e » muet final sur cet item pourrait provenir d'une volonté de marquer le genre féminin lorsque FIANCÉ a également été orthographié avec un « -e » final (la graphie *fiancée* provoquerait l'accord en genre : *solennelle*). Pour cela, nous avons sélectionné 103 graphies qui restituent phonographiquement (et orthographiquement) l'item /solanel/ (ou qui comportent une erreur phonographique qui ne porte pas sur les graphèmes en finale du mot (ex. /solonel/, /sœlanel/, /sɔrlanel/, etc.¹⁴⁴). Il apparaît que parmi ces 103 graphies, seules 15 présentent un « -e » en finale sur FIANCÉ et sur SOLENNEL : « *fiancée solanelle* », « *fiancée solannelle* ». L'observation des résultats obtenus par ces 15 enquêtés qui ont proposé une graphie de SOLENNEL qui se termine par un « -e » final, montre une diversité de profil orthographique :

- 3 enquêtés ont réussi les deux accords attendus (sur FROMAGES et SENTENT) ;
- 5 n'ont réalisé aucun des deux accords attendus ;
- leur score orthographique est réparti de 6 à 20 sur 24 (avec une moyenne de 16 et une médiane à 17).

Par ailleurs d'après la base Manulex-Infra, la graphie SOLENNELLE est moins fréquente (U = 0,94) que la graphie SOLENNEL (U = 2,53) et également moins consistante (67,22 vs 86,64).

Nous notons que parmi ces 111 marques de genre observées sur SOLENNEL, nous trouvons seulement 17 graphies qui n'ont pas doublé la consonne finale « -l » (*solanèle*, *sollenele*,

¹⁴⁴ Voir la liste des graphies non conformes de SOLENNEL dans l'annexe 3.

salennele) ; les autres graphies marquant de manière correcte le féminin sur cet adjectif (*solennelle, sollanelle, soulanelle*).

BILAN DES RÉPONSES QUI COMPORTENT UNE ERREUR MORPHOGRAPHIQUE SUR « FIANCÉ SOLENNEL »

L'observation de l'ajout d'un « -e » final sur les items FIANCÉ et SOLENNEL (dans la construction LE FIANCÉ SOLENNEL) semble montrer qu'il n'y a pas nécessairement de lien entre l'ajout d'un « -e » sur FIANCÉ et l'ajout d'un « -e » sur SOLENNEL.

En effet, la graphie *fiancée* semble davantage provenir d'un effet de la fréquence que de la volonté de marquer le genre féminin puisque :

- 1/ seules 4 graphies s'expliquent par la présence du déterminant LA ;
- 2/ ceux qui ont ajouté un « -e » final font plutôt parti des scripteurs qui ont montré une fragilité en morphographie.

La présence d'une finale à SOLENNEL est également davantage observée chez les enquêtés qui ont montré des difficultés dans le domaine de la morphographie avec peu d'accords réalisés entre FIANCÉ (lorsqu'il a été orthographié avec un « -e » final) et SOLENNEL. L'ajout d'un « -e » final sur cet item peu fréquent et dont l'orthographe a plus particulièrement posé problème aux enquêtés ne peut s'expliquer par la fréquence de sa forme fléchie puisque selon la base Manulex-Infra, la graphie SOLENNELLE est moins fréquente ($U = 0,94$) que SOLENNEL ($U = 2,53$). Dans le Robert électronique, les mots qui se terminent par « -el » (comme COLONNEL, BÉCHAMEL, APPEL) sont deux fois plus fréquents que ceux qui se terminent par « -elle » (VERMICELLE, VAISSELLE). Ainsi, l'hypothèse d'une sensibilité aux formes qui se terminent par « -elle » pour expliquer la présence d'une finale à SOLENNEL n'est étayée par aucun outil à notre disposition. Il semblerait donc que ces graphies ne correspondent ni à une volonté de marquer le genre féminin, ni à l'application d'une régularité du français. Nous avons en revanche relevé que dans EOLE, les élèves de CM2 ont mieux réussi écrire l'adjectif au féminin qu'au masculin. On peut alors émettre l'hypothèse que les formes de féminin sont réussies par défaut, parce que les élèves (et certains adultes), ressentent une incomplétude à écrire une terminaison en « -el ».

2.4.5. BILAN SUR LA MORPHOGRAPHIE

L'analyse du comportement des enquêtés « bas » sur les marques de nombre attendues (le « -s » de FROMAGES et le « -nt » de SENTENT), sur les marques de nombre implicites (ex. *des CERISES*), sur l'absence d'accord en nombre (FORT utilisé comme adverbe) et sur le respect du genre (ex. Le FIANCÉ SOLENNEL) montre que la morphographie semble être un problème pour plus de la moitié des enquêtés « bas ».

En effet, seuls 36% des enquêtés « bas » ont réussi les deux marques de nombre et 27% n'ont réussi qu'une seule des deux marques. Lorsque les enquêtés « bas » n'ont réussi qu'une seule des deux marques, c'est l'accord en nombre du nom qui est davantage maîtrisé (avec 92 marques réussies contre 29 pour SENTENT).

Il apparaît également que 85% des enquêtés qui ont réussi les deux marques de nombre attendues (ce qui représente 32% des enquêtés « bas ») ont obtenu un score moyen de réussite sur l'intégralité de la dictée, supérieur à 80% (avec une note comprise entre 20 et 24 sur 24).

L'accord en nombre du nom FROMAGES et celui du verbe SENTENT semblent donc constituer un bon indicateur pour distinguer les personnes qui ont des difficultés en orthographe lexicale puisque ceux qui ont réussi la morphographie sont également ceux qui réussissent le mieux la dictée (orthographe lexicale et pseudo-mots). Par contraste, la pertinence de l'item FORT est discutable dans la mesure où c'est l'absence de marque qui correspond à une réussite et qu'il n'est pas possible, sans interroger le scripteur sur ses choix, de déterminer s'il a conscience de ne pas marquer le nombre sur un adverbe.

Si ces deux accords permettent également de distinguer les enquêtés qui semblent avoir d'importantes difficultés en morphographie (41% n'ont pas réussi à accorder FROMAGES et 54% n'ont pas réussi l'accord de SENTENT), notamment les 169 enquêtés (38%) qui n'ont réussi aucune des deux marques, ils ne suffisent pas, en revanche, pour cerner les compétences dans ce domaine des personnes qui ont de meilleures connaissances orthographiques. Il aurait en effet été intéressant de davantage tester les compétences en morphographie des enquêtés « bas » de ceux qui ont les deux accords, pour déterminer leur niveau de compétence dans ce domaine.

Enfin, l'analyse des marques de nombre implicites sur les noms dictés sans déterminant (ex. *des CERISES*) montre que deux items (CERISES et TOMATES) ont été plus particulièrement mis au pluriel chez les enquêtés : 34% des enquêtés ont marqué le nombre sur CERISES ainsi que sur TOMATES. Nous remarquons également que l'ajout d'un « -s » est davantage observé sur

les graphies qui sont correctes d'un point de vue orthographique. En effet, pour CERISES, 90% de « -s » sont observés sur la graphie *cerises* (et 10% parmi les graphies phonographiquement acceptables ou non : *ceurises, cérises, etc.*). Pour TOMATES, l'ajout d'une marque de pluriel est observé à 97% sur la graphie *tomates* (et 3% parmi les graphies phonographiquement acceptables : *tomats* ou erronées : *tonates, etc.*).

Cela nous permet également de constater que lorsque le nombre doit être marqué, un signal clair représenté par un déterminant semble représenter une aide pour les scripteurs (34% de « -s » sur CERISES vs 59% sur TROIS FROMAGES).

L'absence de marque attendue sur les noms qui indiquent une quantité (ex. *DU SEL*) ainsi que sur les noms de rayon (ex. *LIBRAIRIE*) a été bien respectée même si cette compétence, qui correspond à une absence de marque (comme sur *FORT*), ne permet pas de déterminer si le scripteur a bien eu conscience de ne pas devoir marquer le nombre. Néanmoins, si la majorité des enquêtés « bas » n'a effectivement pas marqué le nombre lorsque sa présence était implicitement demandée (notamment sur CERISES), la quasi absence de pluriel sur les noms de rayon (2 sur *PHARMACIE* et *EPICERIE*, 4 sur *LIBRAIRIE*) indique une conscience des règles de fonctionnement en matière de morphographie.

L'observation du comportement des scripteurs qui n'ont pas de difficulté particulière à l'écrit (corpus des enquêtés « haut ») sur ces items, nous permettra de voir si eux-aussi ont eu conscience des marques implicites de nombre attendues et si l'absence d'un signal clair pour l'indiquer représente comme chez les enquêtés « bas », une difficulté.

2.5.ANALYSE DE QUELQUES MOTS CHOISIS POUR LEUR COMPLEXITÉ ORTHOGRAPHIQUE

Afin d'affiner la description des compétences des enquêtés, nous proposons maintenant de détecter les effets de la fréquence et de la consistance.

2.5.1. SIX MOTS CHOISIS PAR RAPPORT AUX CRITÈRES DE FRÉQUENCE ET DE CONSISTANCE

Nous avons vu que les résultats de la dictée d'IVQ montraient des écarts importants selon les mots concernés. Ces résultats vont dans le sens des études qui mettent en avant l'importance des critères de fréquence et de consistance dans la réussite orthographique (Lété, 2018 ; Pérez et Giraud, 2019). Pour étayer ce constat et mieux cerner ainsi les compétences en orthographe lexicale des scripteurs, nous proposons d'analyser la restitution orthographique de six mots choisis en fonction de ces deux critères.

Il s'agit :

- des trois mots les plus consistants du corpus : TOMATE, FROMAGE(S), SEL ;
- de deux mots fréquents mais peu consistants : PHARMACIE, PAYS ;
- d'un mot peu consistant et peu fréquent : FIANCÉ.

L'objectif étant l'analyse des compétences en orthographe lexicale, nous écartons la question de la morphographie : la présence, ou non, d'un « -s » final qui marque le pluriel sur FROMAGE(S) n'est pas prise en compte, de même que la présence d'un « -e » final à FIANCÉ. Dans le même sens, nous faisons le choix de nous centrer sur les enquêtés pour lesquels le score à la dictée d'IVQ est compris entre 35 et 80%. La barre de 80% délimite les scripteurs sans difficulté ; celle des 35% sépare arbitrairement les scripteurs qui ont de fortes difficultés de ceux qui sont en graves difficultés avec l'écrit. Nous ne prenons pas en compte les enquêtés qui ont moins de 35% de réussite car les graphies produites cumulent les difficultés et présentent pour beaucoup des problèmes phonographiques. En outre, certains n'ont produit aucune graphie.

Pour cette analyse, nous choisissons donc de nous centrer sur les 202 scripteurs dits « en difficulté » (45% des enquêtés « bas » à l'issue du module d'orientation, qui ont obtenu un score compris entre 9 et 19 à la dictée d'IVQ, soit entre 35 et 80% de réussite) et nous comparons leurs résultats à ceux obtenus par les 176 enquêtés qui n'ont que peu (ou n'ont pas) manifesté de difficultés dans la dictée d'IVQ (scores > à 80%).

Cette analyse est illustrée par la figure 16. Les mots sont classés du mieux réussi chez les enquêtés en difficulté au moins bien réussi :

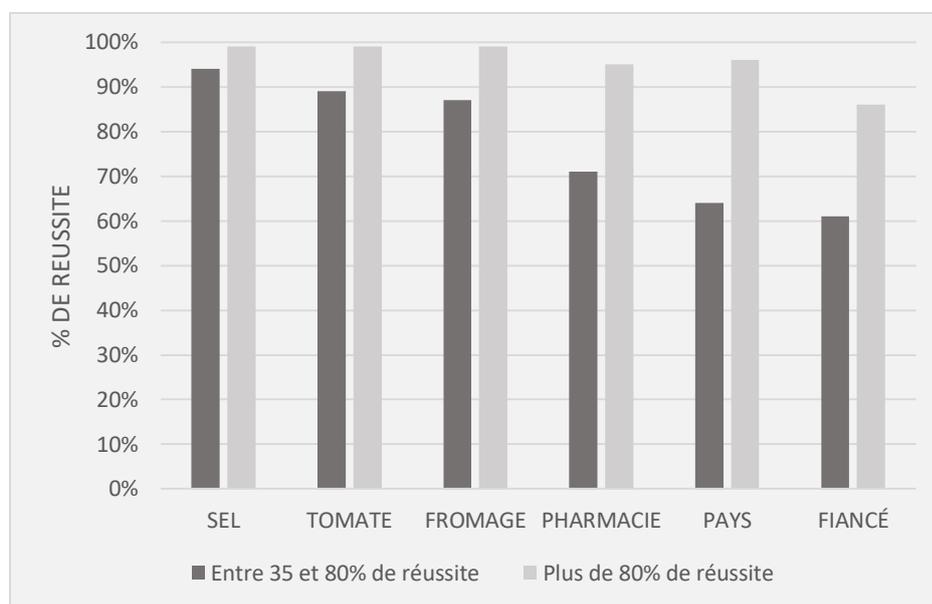


Figure 16. Comparaison du taux de réussite de 6 mots entre les 215 enquêtés « bas » en difficulté dans la dictée (score compris entre 35% et 80%) et les 176 enquêtés « bas » qui ne sont pas en difficulté (plus de 80% de réussite).

Les enquêtés qui ont le moins de difficultés (scores > 80%), obtiennent de très bons taux de réussite (96% en moyenne) sur cette sélection de mots. Les mots consistants (SEL, TOMATE, FROMAGE) approchent 100% de réussite. Lorsque les correspondances phonie-graphie sont moins régulières mais que les mots sont fréquents (PHARMACIE, PAYS), le taux de réussite orthographique est également élevé, notamment sur PAYS (96% de réussite) qui est particulièrement peu consistant. C'est l'item FIANCÉ, à la fois peu fréquent et peu consistant, qui obtient le taux de réussite le moins élevé avec 86%.

Les mêmes tendances s'observent chez les enquêtés en difficulté dans la dictée d'IVQ, avec les mots consistants qui sont les mieux réussis et les moins fréquents et/ou consistants qui sont les moins bien restitués. Ici aussi, c'est l'item FIANCÉ qui est le moins bien restitué avec 61% de réussite. D'autre part, on remarque que les écarts se creusent entre les deux populations lorsque les mots contiennent des graphèmes peu consistants (le « ph » de PHARMACIE, le « y » de PAYS, ou le « i » de FIANCÉ). Ainsi, on note entre les deux populations un écart de 32 points sur la réussite de PAYS (alors que cet écart n'est par exemple que de 5 points sur SEL). On peut donc conclure que des difficultés se manifestent pour tous les scripteurs lorsque les correspondances phonie-graphie sont moins consistantes et/ou que les mots sont moins fréquents. Ce sont des difficultés intrinsèques au système orthographique, qui impactent tous les enquêtés, et davantage encore les enquêtés en difficulté chez qui l'on observe des écarts de réussite plus marqués selon les mots.

Le corpus manque de contrastes plus nets pour pouvoir mesurer l'impact de l'effet de consistance et de fréquence : les mots les plus rares sont aussi les moins consistants. Les résultats montrent néanmoins que la réussite sur les items semble davantage provenir de l'effet de consistance et donc de la régularité des combinaisons de phonèmes et de graphèmes. La présence de correspondances phonie-graphie irrégulières fait en effet baisser les résultats chez tous les enquêtés « bas » y compris sur les mots fréquents comme PAYS. L'effet de fréquence joue cependant un rôle non négligeable sur la réussite orthographique puisque certains mots particulièrement irréguliers comme PAYS, montrent que même les scripteurs en difficulté ont mémorisé certaines formes graphiques complexes, par exemple sous l'effet de la scolarisation ou la fréquentation des mots au quotidien).

PRÉSENTATION DES GRAPHIES DE 53 SCRIPTEURS EN DIFFICULTÉ

Dans le tableau 23 qui suit, nous observons les erreurs produites sur les six mots chez les enquêtés en difficulté sur la dictée (avec un score compris entre 9 et 19, soit entre 35 et 80% de réussite) et qui ont fait au moins une erreur sur les trois mots les plus consistants du corpus : TOMATE(S), FROMAGE(S), SEL. Parmi les 202 enquêtés qui ont été en difficulté sur l'exercice, 53 enquêtés (26%) ont fait au moins une erreur sur ces trois mots. Dans le tableau 23, les graphies ont été reproduites telles qu'elles apparaissent dans les dictées (majuscules, diacritiques, etc.). Le chiffre 0 indique que l'enquêté n'a pas produit de graphie et la lettre X que la graphie était illisible. Les enquêtés sont classés par ordre décroissant en fonction du score obtenu.

n° enquêté	TOMATE	SEL	FROMAGE-S	PHARMACIE	PAYS	FIANCÉ	SCORES
1552	tommate	Sel	fromage	pharmacie	pays	fiance	19
1569	Thomate	Sel	fromages	Phamacie	pays	fiancé	19
1596	Tomâtas	sel	fromages	Pharmacie	Pays	Fiancié	19
1894	Tomate	0	fromages	Pharmacie	Pay	Fiancé	19
897	Tomade	sel	fromage	Pharmacie	Pays	faincé	18
1983	Tomote	sel	fromage	Pharmacie	pays	fiance	18
1990	Tomate	Sel	Frommage	Pharmacie	pays	Fiancé	18
1678	thomate	sel	fromages	pharmacie	Pays	fillansais	17
1862	tomates	sel	0	pharmacie	pays	fiancé	17
1868	Tomete	sel	fromage	pharmacie	pays	fiance	17
1902	tomates	cel	fromage	pharmacie	pays	fiancé	17
1929	Tomate	Sel	0	Pharmacie	Pays	fiancé	17
1568	Tomate	sel	Fomages	PHARMacie	Pay	Fiance	16
1605	tomate	sel	formages	pharmacie	pays	fiancé	16
1622	tomatte	sel	fromage	pharmaci	pays	fillancé	16
1657	Tomade	Sel	Fromage	Pharmacie	PAYS	fiancée	16
1749	tomate	Sel	framages	Pharmacie	pays	fioncé	16
1861	tomates	sel	0	PHarmacie	pays	fiançé	16
1967	tommate	sel	fromage	Pharmacie	pay	fiancé	16
898	Tomade	Sel	Fromage	pharmacie	Pay	fiancé	15
1562	tomate	Cél	fromages	pharmacie	pays	fiancé	15
1620	Tomato	Cels	fromages	PHarmacie	Peyer	fiancé	15
1870	tomate	sel	0	Pharmacie	Pays	fiençe	15
1883	Tomate	sel	fromaje	pharmacie	pays	fiance	15
1904	Tomates	Sel	Frome	Pharmacie	0	X	15
1961	tomates	sele	fromage	pharmacie	pais	fiancé	15
2030	tomatte	sel	fromges	pharmacie	pays	fiancé	14
1629	Tomate	cel	Fromages	pharmacie	peyi	fiancé	13
1769	Tomate	Cel	fromage	Fharmaçie	paye	fiencençais	13
1811	Tomatate	sel	Fromages	PHARMACIE	pays	fianciée	13
1855	Tomate	Sel	Fraumage	pharmacie	peilhe	Fiance	13
1876	tommatte	sel	frommage	pharmaci	pay	fiancé	13
1621	Tommate	sel	Fromage	pharmacie	payie	fiancer	12
1826	tomate	cel	fromages	pharmacie	payé	fiancé	12
1567	tomate	sel	Fomage	Farmacie	pais	fiance	11
1617	tomate	sel	fromges	pharmaçi	payie	fionçé	11

1804	TOMA	SEL	FROMAGE	PHARMACI	PEY	FIANCER	11
1816	TOMATE	SEL	FORMAGE	FARMACIE	PEI	FIANCE	11
1821	tomate	sal	fromage	Pharmacy	peuy	fiance	11
1873	tonates	sel	Fronage	PHARMaci	poye	Fiancer	11
1992	tomates	Se	Fomage	pharmacice	pays	fuiancer	11
2026	tomate	sel	0	phramacie	paye	fiense	11
1610	tomate	Selle	fromage	pharmacie	pye	fionsé	10
1615	tomate	SEL	Formaj	FARMACIE	pei	fiancee	10
1663	tomate	sel	Formages	Farmaçie	pays	fiance	10
1764	Tomatte	cel	Fromage	FARMACIE	pey	Fiance	10
1841	TOMATE	Sel	Framje	PHARMACIE	PIYER	fiance	10
2012	Tomat	Sel	fromage	Farmacie	pais	fiance	10
1752	Tomates	Sel	0	0	pays	Fiance	9
1766	Tomate	Sel	fromage	Phamaci	Pays	fiace	9
1790	tomatte	Sel	fromage	Farmacille	payer	fiancé	9
1815	tomato	Sel	Formage	Pharmacy	payee	fiancé	9
1881	tomate	Sel	frommge	pharmacie	pays	fincie	9

Tableau 23. Détail des erreurs produites chez les 53 scripteurs en difficulté sur la dictée (score compris entre 9 et 19 sur 24) et ayant fait au moins une erreur sur la sélection des 3 mots les plus consistants

Chez ces 53 enquêtés (26% des 202 enquêtés qui ont obtenu entre 40% et 80% de réussite à la dictée) qui ont fait au moins une erreur sur les trois mots les plus consistants, SEL, TOMATE(S) et FROMAGE(S), c'est l'item TOMATE (l'item le plus consistant du corpus selon MANULEX-INFRA) qui contient le plus d'erreurs (21) suivi de FROMAGE(S) (14) puis SEL (8).

ANALYSE DES PROPOSITIONS GRAPHIQUES DE TOMATE, FROMAGE-S ET SEL

TOMATE

Pour **TOMATE** : sur les 21 erreurs faites par les 53 enquêtés, nous relevons 11 graphies qui ne sont pas phonographiquement acceptables (on ne retrouve pas une unité lexicale du français oral en les déchiffrant).

- erreurs par ajout de phonèmes : *tomato* (3) - *tomâtas* (1) - *tomatate* (1) ;
- confusion entre deux phonèmes : /d/ à la place de /t/ : *tomade* (3) ; /o/ et /ə/ à la place de /a/ : *tomote* (1) - *tomete* (1) ; /n/ à la place de /m/ : *tonates* (1) ;

À l'inverse, 11 erreurs n'impactent pas la forme phonologique de l'item :

- 7 erreurs qui concernent l'ajout d'une consonne double : *tomatte* (4), *tommatte* (1), *tommate* (2) ;
- 2 erreurs qui portent sur l'ajout d'une lettre muette, ici un « h » muet : *thomate* (2) ;
- 1 erreur concerne l'omission de la finale « -e » dans *tomat* (1).

Ce « -e » final est décisif par rapport au critère de consistance, dans cet item qui est celui qui est le plus consistant du corpus. En effet, au niveau phonographique, cette finale « -e » présente le mode de fonctionnement suivant : il s'agit d'une lettre muette qui rend sonore le « -t » précédent (on peut parler de valeur auxiliaire, Blanche-Benveniste, 1969). La finale « -at » se lit /a/ par exemple dans AVOCAT, dans ABAT (et dans les finales verbales au subjonctif imparfait - ex. AIMÂT) et c'est le cas de figure le plus fréquent. Cependant, MAT dans ÉCHEC ET MAT ou OUEST montrent que l'oralisation du « t » sans le « -e » final est possible.

FROMAGE-S

Pour **FROMAGE-S**, les erreurs qui impactent la forme phonologique et celles qui ne l'impactent pas ne sont pas autant équilibrées. La grande majorité des graphies erronées n'est pas phonographiquement acceptable :

- confusion dans l'ordre des graphèmes : *formage(s)* (5), *formaj* (1) ;
- omission d'un graphème : *fomage(s)* (3), *fromges* (2), *frommge* (1)
- confusion entre deux phonèmes : /a/ pour /o/ : *framages* (1) ; /n/ pour /m/ : *fronage* (1).

Quelques graphies erronées d'un point de vue orthographique, sont phonographiquement acceptables :

- ajout d'une consonne double : *frommage* (2) ;
- choix de la graphie d'un archigraphème : « au » pour « o » : *fraumage* (1) ; « g » pour « j » : *fromaje* (1).

Ces exemples de graphies montrent que le principe de phonographie est essentiel (le mot reste oralisable, donc le message communicable),

- enfin, nous relevons une graphie qui cumule deux types d'erreurs : *framje* (1).

SEL

L'item **SEL** compte seulement 11 graphies erronées et seulement deux graphies produisent une forme erronée au niveau phonologique :

- - substitution d'un phonème : *sal* ;
- - omission d'un graphème : *se*.

La majorité des graphies erronées est phonographiquement acceptable. La principale erreur concerne le choix du premier graphème avec le recours à « c » dans *cel(s)* (7). Nous relevons également l'homophone *selle* (1) et l'ajout d'un « e » muet final dans *sele* (1)

Ainsi, les erreurs relevées sur les mots très consistants de la dictée chez les enquêtés en difficulté sont généralement révélatrices des difficultés inhérentes à l'écriture du français, comme la variété de graphèmes possibles pour un même phonème, la possibilité de doubler certaines consonnes ou encore l'ajout ou non de lettres muettes, notamment en fin de mot. La variété des erreurs est elle-même révélatrice du système orthographique du français : dès lors que l'on ne se situe pas dans une écriture exclusivement phonographique et parfaitement consistante à cet égard, tout est possible, en quelque sorte, ou pour le dire plus précisément, un très grand nombre de possibilités existe ; on a vu par exemple qu'on obtient dans notre corpus de 53 dictées, 7 graphies, en plus de la graphie normée TOMATE, pour la même chaîne sonore /tomat/. Malgré la consistance de cet item qui est le plus consistant des items de notre dictée, le problème de la consistance phonographique dans l'orthographe du français est tel que si on ne connaît pas l'orthographe du mot, elle est peu prédictible : on pourrait envisager plusieurs choix possibles pour le phonème /o/, on peut imaginer un dédoublement du « m » ou du second « t » comme dans TOMME ou NATTE, la présence ou l'absence d'un « -e » final. Avec TOMATE, on ne trouve pas dans ce corpus de cas de substitution de « o » par « au » par exemple (mais nous verrons que cette substitution est présente dans certaines graphies du corpus quart sud-est) ; il faudrait d'autres études plus ciblées, mais ce fait suggère que les scripteurs même en difficulté ont sans doute des intuitions sur ce qui est plus probable en termes de régularités dans les correspondances phonèmes-graphèmes. Ces notions relèveraient de la compétence phonographique : encoder de manière à produire une forme phonologiquement correcte et phonographiquement probable.

Enfin, les graphies qui ne permettent pas de restituer la forme phonologique de l'item révèlent chez l'enquêté, des difficultés avec l'écrit d'un autre ordre, que l'on ne pourrait expliquer sans observer leur profil. Outre l'illettrisme, elles révèlent des connaissances partielles du lexique

chez les personnes dont le français n'est pas la langue première ou encore la méconnaissance d'un item en particulier¹⁴⁵.

Nous allons observer à présent la manière dont les 53 enquêtés en difficulté (score de 35 à 80% de réussite) et qui ont fait au moins une erreur sur les trois mots très consistants de la dictée, ont orthographié les items FIANCÉ, PAYS (peu consistants mais fréquents) et PHARMACIE (peu consistant et peu fréquent)¹⁴⁶.

ANALYSE DES PROPOSITIONS GRAPHIQUES DE FIANCÉ, PAYS ET PHARMACIE

FIANCÉ

Chez les 53 enquêtés qui ont produit au moins une erreur sur les mots consistants et dont les graphies sont listées dans le tableau 23 ci-dessus, nous observons pour l'item **FIANCÉ**, 29 graphies non conformes et une graphie illisible. Dans Manulex-Infra, cet item est le moins consistant de la dictée et il n'est pas non plus considéré comme fréquent. En revanche il est plutôt fréquent selon la base Lexique Films.

La majorité des graphies erronées ne restituent pas la forme phonologique de FIANCÉ et on observe de multiples cas de figure :

- inversion de graphèmes dans *faincé* (1) ;
- ajout d'un ou plusieurs phonèmes : /i/ dans *fiancié(e)* (2), *fuiancer* (1), *fincie* (1) ; /s/ et /ã/ dans *fiencençais* (1) ; /j/ dans *fillansais* (1) ;
- erreur dans la transcription d'un phonème :
- /a/ pour /ã/ : *fiace* (1) ;
- /õ/ pour /a/ : *fionsé* (1), *fiomcé* (1) ;
- /z/ pour /s/ dans *fionsé* (1), *fillansais* (1), *fiense* (1) qui témoigne d'une méconnaissance des règles de distribution spécifiques au français.
- omission d'un graphème dans *fincie* (1) qui (outre le problème de l'ajout du « i » traité au-dessus) ne contient qu'un seul des deux graphèmes qui permettent de transcrire le phonème /ã/.

¹⁴⁵ Suite aux résultats de l'enquête de 2004, l'item AUBERGINE a été retiré de l'enquête car l'INSEE a constaté qu'une proportion non négligeable d'enquêtés ne connaissait pas ce mot, ce qui constituait un biais.

¹⁴⁶ Voir la section 1 du chapitre 3 pour une présentation des outils Manulex et Lexique et de et la section 1.2. du chapitre 4 pour une présentation des indices de consistance des mots de la dictée calculés dans MANULEX et de fréquence calculés dans LEXIQUE.

Quant aux 2 graphies qui permettent de restituer la forme phonologique de l'item, une graphie ne contient comme erreur que l'ajout d'une cédille dans *fiançé* (1). Nous relevons également une substitution de « ia » par « illa », autrement dit un ajout de graphèmes pour marquer le phonème /j/ dans *fillancé* (1).

Les autres erreurs portent sur le choix d'un graphème :

- graphie de l'archigraphème qui permet de transcrire le dernier phonème /e/ : « er » *fiancer* (2), « ais » *fillansais* (1), *fiencençais* (1) ;
- choix du graphème pour transcrire le phonème /ã/ : *fiençé* (1).

Il est intéressant de voir qu'on obtient très peu de « en » à la place de « an » (*fiencençais, fiença, fiense*) alors même que la distribution entre les deux est jugée être quasiment aléatoire. Ici la régularité dans le système ne pouvait pas jouer sur les graphies produites, et pourtant ces graphies erronées attestent d'une certaine régularité, bien qu'elle soit à nuancer car à « an » se substitue plusieurs fois « on » notamment (*fiomcé, fionçé, fionsé*).

Sur cet item, deux graphèmes vocaliques ont plus particulièrement posé problème aux enquêtés : la transcription graphique de /ã/ (9 erreurs) et surtout le choix de la graphie de l'archigraphème « é » (24 erreurs).

PHARMACIE

Chez les enquêtés qui ont produit au moins une graphie erronée sur les trois mots les plus consistants, 19 (36%) ont également produit une graphie non conforme de **PHARMACIE**, soit autant que sur l'item FROMAGE-S, alors que les correspondances phonie-graphie sont moins régulières.

Les graphies ne restituant pas la forme phonologique du mot sont les suivantes :

- omission d'un graphème : *phamacie* (1), *phamaci* (1) ;
- inversion de deux graphèmes : *phramacie* (1), *pharmacei* (1) ainsi que dans *framacie* (1) qui contient également une erreur dans le choix de l'archigraphème /F/ ;
- ajout de graphèmes : *pharmacice* (1), *pharmacille* (1).

Parmi les réponses phonographiquement acceptables on observe principalement :

- l'omission du « e » muet final¹⁴⁷ : *pharmaci* (4), *pharmaçi* (1), *farmaci* (2) ;

¹⁴⁷ Une analyse sur la restitution du « -e » final sur cet item est proposée dans la section 2.5.2.1

- les graphies *pharmacy* (2) ont été produites par deux scripteurs dont le pays de première scolarisation n'est pas la France (Europe de l'Est et reste du monde) ;
- une confusion dans le choix de l'archigraphème /F/ : *farmacie* (5), *farmaçie* (1), *farmaci* (2), *fharmaçie* (1).

Sur cet item, les choix graphiques montrent que certains enquêtés possèdent des compétences orthographiques partielles : sur *pharmacille* (1), l'enquêté a enregistré la graphie « ph » pourtant moins fréquente que « f », tout en produisant une erreur phonographique sur la fin du mot.

PAYS

La restitution de l'item **PAYS** a davantage posé problème, avec 26 graphies erronées (49%) dans ce corpus de 53 dictées. L'analyse des propositions erronées montre que 20 graphies présentent des traces de la structure complexes de ce mot et contiennent à minima le graphème « y » : *pay* (5), *paye* (2), *payee* (1), *payis* (1), *payie* (1), *payé* (1), *payer* (1), *pey* (2), *peyi* (1), *peyer* (1), *pye* (1), *peuy* (1), *poye* (1).

9 graphies semblent provenir d'une volonté de restituer la forme phonographique de la première syllabe : *pei* (2), *pais* (3), *peilhe* (1).

Finalement, malgré les propositions parfois éloignées de la forme orthographique attendue (par exemple la graphie *peilhe* qui hors contexte ne pourrait pas permettre au lecteur de la rapprocher de PAYS) aucune n'est totalement éloignée de la forme phonologique ou de la structure particulière de cet item et les réponses des enquêtés montrent des traces de ce qui est attendu.

Ainsi, les erreurs relevées sur les mots les moins consistants de la dictée font apparaître comme celles des mots les plus consistants différents aspects des difficultés de l'écriture du français. Nous avons relevé la grande diversité des graphies possibles ou envisageables au niveau phonographique et les influences possibles des profils singuliers des scripteurs (ex. y en finale de *pharmacy*). Si l'on prend en compte dans son ensemble, cette sélection de six mots issus de la dictée chez les enquêtés en difficulté, on observe qu'elles reflètent notamment la variété de graphèmes disponibles pour transcrire un même phonème, la possibilité de doubler certaines consonnes ou d'ajouter une lettre muette. Même lorsque les graphies sont régulières et ne sont composées que des graphèmes les plus fréquents (comme on l'a vu à propos de TOMATE), la complexité orthographique de l'écriture du français semble venir perturber les scripteurs lorsqu'ils doivent orthographier des mots simples (*tommate*, *thomate*, *tomat*). Ainsi, les 29 graphies non conformes de FIANCÉ présentent 16 formes

graphiques différentes. Or, la variation graphique observée, nous le rappelons, diminue avec le degré d'expertise orthographique (Blondel, Conseil, 2015). Comme le suggère le corpus de graphies analysé ici, même si l'orthographe est peu prédictible et que de nombreuses formes pourraient être acceptables au niveau phonographique, certains graphèmes sont plus probables que d'autres, comme « o » plutôt que « au » dans TOMATE.

Cette analyse des réponses sur trois mots consistants (SEL, TOMATE et FROMAGE-S) et trois mots peu consistants (PHARMACIE, PAYS et FIANCÉ) n'a porté que sur 53 sujets, mais il en ressort que la consistance des mots impacte leurs compétences orthographiques. On le voit très clairement avec l'exemple de PAYS, retenu pour sa faible consistance et sa fréquence importante, et pour lequel on a relevé 26 graphies erronées sur 53 dictées. Cependant, ce facteur ne suffit pas pour expliquer les difficultés. Si on trouve en effet davantage d'erreurs d'ordre phonographique dans les mots moins consistants, qui présentent des correspondances phonie-graphie moins régulières (PHARMACIE, PAYS), l'item PHARMACIE est plutôt bien réussi chez les scripteurs en difficulté alors que ce mot cumule les difficultés (longueur, choix de graphème pour « ph » et « c », finale muette). On obtient 19 erreurs, soit autant que sur FROMAGE-S qui est plus consistant. Nous avons également observé que certains enquêtés en difficulté sur des items consistants comme FROMAGE-S ou TOMATE sont néanmoins parvenus à restituer complètement (ou partiellement) certaines structures de mots plus complexes (comme PAYS ou PHARMACIE).

La fréquence d'exposition semble ainsi avoir un effet sur la réussite orthographique de mots qui font partie du quotidien. Pour mesurer cet effet de fréquence se pose un problème d'outil.

Dans la base Lexique, comme nous l'avons dit, le mot PHARMACIE apparaît comme peu fréquent. Il est difficile de rendre compte de la fréquence d'exposition de mots que nous rencontrons au quotidien, dans la rue par exemple et de croiser la connaissance du lexique et la connaissance de la forme graphique des mots. Il faudrait aussi pouvoir évaluer ce qu'il en est pour chaque individu. Les compétences des scripteurs en difficulté sont inhérentes à chacun, par exemple pour la connaissance que certains ont de mots complexes et la méconnaissance de mots simples.

Nous avons également constaté sur cet échantillon de six items produits par 53 scripteurs que si certaines graphies semblent provenir d'une transcription phonétique de l'item dicté (*pei* pour PAYS), souvent les choix graphiques proposés sont plausibles et respectent les grands principes de l'écriture du français. Nous trouvons par exemple des redoublements de consonnes erronés (*tommate, frommage*) mais jamais une succession de trois consonnes dont une géminée, non attestée en français (par exemple *pharmmacie*). Enfin, les graphies qui ne

permettent pas de restituer la forme phonologique de l'item révèlent chez l'enquêté, des difficultés avec l'écrit qui semblent d'un autre ordre et que l'on pourrait vraisemblablement expliquer par l'observation de leur profil, par exemple pour la méconnaissance d'un item en particulier ou l'usage de graphèmes surprenants en français (« fh »), qui pourraient être influencés par d'autres systèmes linguistiques (« -y » en finale de *pharmacy*).

2.5.2. LE CAS DU « -E » FINAL

Pour compléter l'analyse sur les compétences en orthographe lexicale des enquêtés « bas », nous observons parmi les réponses non conformes à la norme orthographique, la restitution des finales muettes. Pour chaque analyse, les pourcentages de réussite sont calculés en fonction du nombre réel de réponses (et non selon le nombre d'enquêtés). Nous opposons à chaque fois les graphies phonographiquement acceptables aux autres graphies (approximantes et non acceptables).

Nous menons cette analyse en trois temps : d'abord sur les mots qui possèdent une marque flexionnelle : PHARMACIE, ÉPICERIE, LIBRAIRIE puis sur les mots dont la présence d'un « -e » final permet la prononciation de la consonne finale : TOMATE, CERISE, FROMAGE, RHUME et enfin sur CONFITURE, qui ne nécessite pas la présence d'un « -e » final pour que sa forme phonographique soit respectée. Cette analyse nous permettra de rendre compte de la sensibilité des enquêtés aux finales des mots et à certaines régularités du français, même s'ils ne maîtrisent pas l'orthographe des mots observés.

Pour calculer la proportion de restitution de « -e final », nous ne conservons que les graphies qui proposent un phonème final identique à celui attendu : pour la première analyse (PHARMACIE, ÉPICERIE et LIBRAIRIE), il s'agit des graphies qui se terminent par /i/. Cela nous permet de compléter nos analyses en confrontant les réponses des scripteurs avec la fréquence des finales sur les mots qui se terminent par le même phonème (à partir du Robert électronique de 2011). En effet, une graphie comme *nhamacia* proposée pour PHARMACIE, ne peut pas (si l'on respecte les règles de distribution spécifiques au français) se terminer par un « -e » muet final. À l'inverse, la graphie *formachie*, malgré les erreurs d'ordre phonographique qu'elle présente, se termine par le phonème /i/ et il est intéressant de constater que cette proposition présente un « -e » final.

2.5.2.1 LA RESTITUTION DU « -E » FINAL DANS LE CAS D'UNE MARQUE DE DÉRIVATION

Nous présentons dans le tableau 24, la proportion de restitution du « -e » final sur les items PHARMACIE, ÉPICERIE, LIBRAIRIE, lorsque les propositions des enquêtés se terminent par /i/. Nous distinguons les graphies que nous avons considérées comme acceptables d'un point de

vue phonographique des autres. Nous indiquons dans la première colonne, le pourcentage de réussite en orthographe lexicale relevé en moyenne sur ces mots. Les mots sont présentés dans le tableau en fonction du taux de restitution du « -e » final parmi les réponses des enquêtés qui s'écartent de la norme orthographique : du plus fort taux de restitution en moyenne à la plus faible restitution en moyenne.

mots	forme ortho conforme	phono acceptable avec « -e » final	phono non acceptable/appro avec « -e » final	moyenne « -e » final (propositions non conformes) ¹⁴⁸
PHARMACIE	75%	56%	64%	60%
ÉPICERIE	73%	48%	49%	49%
LIBRAIRIE	62%	74%	69%	72%

Tableau 24. Proportion de graphies qui contiennent un « -e » final parmi les graphies phonographiquement acceptables et les graphies phonographiquement non acceptables/approximantes sur 3 mots qui contiennent une marque de dérivation chez les 441 enquêtés « bas ».

D'après le *Petit Robert électronique* (2011), 2665 mots se terminent par une finale en « -ie » (BIOLOGIE, BATTERIE, GARANTIE), contre 770 en « -i » (dont 56 en « -oi ») parmi lesquels il y a beaucoup d'adjectifs (JOLI, POLI) ou SOUCL pour la catégorie des noms. Ainsi, dans le lexique français, les mots qui se terminent par /i/ se terminent plus fréquemment par un « -e » final. L'observation de la restitution moyenne du « -e » final dans notre échantillon montre que les graphies proposées ont majoritairement un « -e » final, sauf ÉPICERIE qui possède presque autant de graphies sans « -e » final, qu'avec « -e » final. Nous constatons également que la restitution du « -e » final ne dépend pas nécessairement du type de réponse (phonographiquement acceptable ou phonographiquement non acceptable et approximantes. Si pour LIBRAIRIE il est vrai que les réponses phonographiquement acceptables présentent un taux supérieur de « -e » final à celui des autres réponses (74% contre 69%), cela ne se vérifie pas pour les autres mots : sur PHARMACIE, c'est même le phénomène inverse qui est observé.

Si l'on observe la restitution moyenne du « -e » final dans notre échantillon (tout type de graphies confondues), c'est l'item LIBRAIRIE qui contient le plus fort taux de « -e » final. Nous observons également un écart important entre la restitution moyenne du « -e » final sur les items ÉPICERIE et LIBRAIRIE. Lorsque l'on distingue les graphies phonographiquement acceptables (ex. *librairi*, *l'hibrerie* pour LIBRAIRIE) et les graphies phonographiquement non

¹⁴⁸ Moyenne calculée parmi les propositions non conformes à la norme orthographique : graphies phonographiquement acceptables, approximantes et non acceptables.

acceptables et approximantes (ex. *liberie, lebriri* pour LIBRAIRIE), on constate peu d'écart : le taux de restitution du « -e » final est presque identique, notamment sur ÉPICERIE.

Nous détaillons ci-dessous les résultats pour chaque item :

Le « -e » final de **PHARMACIE**

Parmi les graphies qui ne restituent pas la forme orthographique attendue mais dont le dernier phonème est un /i/ (ex. *farmassi, phramaci*), la proportion de « -e » final est de 60% en moyenne avec :

- 56% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable (ex. *farmacie*) ;
- 64% qui ont produit une forme phonographiquement non acceptable (ex. *phomarcie*).

Le « -e » final de **ÉPICERIE**

Parmi les graphies qui ne restituent pas la forme orthographique attendue mais dont le dernier phonème est un /i/ (ex. *hépicerie, ibsouri*), la proportion de « -e » final est de 49% en moyenne, répartie de la manière suivante :

- 48% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable ou approximante (ex. *episserie, épicrie*) ;
- 49% de ceux qui ont produit une forme phonographiquement non acceptable (ex. *epecerie*).

Parmi les réponses non conformes à la norme orthographique, c'est sur cet item que le « -e » final a été le moins bien restitué, par rapport aux deux autres items analysés. Comme pour PHARMACIE, parmi les formes phonographiquement acceptables, il y a presque autant d'enquêtés qui ont mis un « -e » final, que d'enquêtés qui n'en ont pas mis.

ÉPICERIE est également l'item le moins fréquent des trois selon Manulex-Infra (U = 7,14 vs 16,84 pour PHARMACIE et 11,13 pour LIBRAIRIE).

Le « -e » final de **LIBRAIRIE**

Parmi les graphies qui ne restituent pas la forme orthographique attendue mais dont le dernier phonème est un /i/ (ex. *librerie, elrairie*), la proportion de « -e » final est de 72% en moyenne, répartie de la manière suivante :

- 74% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable (ex. *librérie*) ;

- 69% de ceux qui ont produit une forme phonographiquement non acceptable (ex. *libérie, lubierie*).

Ainsi, sur les items féminins en « -ie » (dont la règle de formation en français est relativement régulière, avec peu d'exceptions), la restitution du « -e » final sur les formes graphiques phonographiquement acceptables (l'item entier, comme *farmacie*) ou sur celles qui restituent à minima le dernier phonème (comme *framassi*) est plutôt importante (entre 56 et 74% des réponses selon les mots). La moyenne de restitution du « -e » final parmi les graphies non conformes à la norme orthographique est comprise entre 49 et 72%. Sauf pour ÉPICERIE (qui nous le rappelons est le moins fréquent des trois selon Manulex-Indra), une majorité des enquêtés « bas » semblent néanmoins posséder une sensibilité aux finales des mots. La distinction entre les enquêtés qui sont parvenus à restituer la forme phonographique des mots et les autres, davantage en difficulté, montre que cette sensibilité aux finales muettes ne semble pas nécessairement liée aux compétences phonographiques : en effet, la restitution du « -e » final est presque identique sur ÉPICERIE et on observe seulement 5 points d'écart sur LIBRAIRIE. Sur PHARMACIE, ce sont même les enquêtés qui ont des difficultés d'ordre phonographique qui ont le taux le plus élevé de restitution du « -e » final, même si cet écart n'est que de 8 points.

2.5.2.2 LA RESTITUTION DU « -E » FINAL SUR LES AUTRES MOTS

Nous observons ensuite la proportion de restitution correcte du « -e » final sur les mots dont la présence de la finale permet la prononciation de la consonne finale : CERISE, FROMAGE, RHUME. Nous complétons cette analyse par l'observation de TOMATE et de CONFITURE, pour qui le « -e » final n'a pas pour fonction de permettre la prononciation de la dernière consonne.

Le tableau 25 présente la proportion de restitution du « -e » final sur les items TOMATE, CERISE, FROMAGE, RHUME et CONFITURE lorsque les propositions des enquêtés respectent le dernier phonème attendu ou le dernier graphème¹⁴⁹ (ex. le /t/ de TOMATE comme dans *temate* = /təmat/ ou le « t » comme dans *tomat* = /tomat/). Nous distinguons les graphies que nous avons considérées comme acceptables d'un point de vue phonographique des autres. Nous indiquons dans la première colonne, le pourcentage de réussite en orthographe lexicale relevé en moyenne sur ces mots. Les mots sont présentés dans le tableau en fonction du taux

¹⁴⁹ Le choix d'ajouter aussi la prise en compte du dernier graphème provient des réponses recueillies pour l'item RHUME. Pour cet item ; nous n'avons conservé que les graphies qui se terminent par /ym/ (ex. *reume*) ou les graphèmes « -um » (ex. *rhûm*) avec ou sans « -e » final. Cela nous a permis d'exclure les graphies telles *roume* ou *rim* qui ne posent pas les mêmes difficultés phonographiques que *rhume* vs *rhum*, pour lesquels le « -e » final conditionne la prononciation de la voyelle.

de restitution du « -e » final parmi les réponses des enquêtés qui s'écartent de la norme orthographique : du plus fort taux de restitution en moyenne à la plus faible restitution en moyenne.

mots	forme ortho conforme	phono acceptable avec « -e » final	phono non acceptable/appro avec « -e » final	moyenne ¹⁵⁰ « -e » final (propositions non conformes)
CERISE	74%	100%	80%	90%
FROMAGE	88%	75%	94%	85%
TOMATE	88%	54%	75%	65%
CONFITURE	83%	57%	72%	65%
RHUME	54%	83%	8%	46%

Tableau 25. Proportion de graphies qui contiennent un « -e » final parmi les graphies phonographiquement acceptables et les graphies phonographiquement non acceptables/approximantes sur 5 mots qui contiennent un « -e » final chez les 441 enquêtés « bas ».

Sur 4 mots (CERISE, FROMAGE, TOMATE et CONFITURE), on observe un taux de restitution du « -e » final parmi les graphies non conformes à la norme orthographique relativement important (entre 90% et 65% des réponses), notamment sur CERISE et FROMAGE. Pour rappel, FROMAGE, TOMATE et CONFITURE font partie des mots les plus consistants de la dictée selon Manulex-Infra ainsi que les plus fréquents. Ils sont aussi ceux qui ont obtenu le plus fort taux de réussite orthographique. Mais au-delà de la réussite orthographique, on observe que CERISE, FROMAGE en particulier, ont un taux de restitution du « -e » final particulièrement élevé, parmi les réponses phonographiquement acceptables ou non. Par ailleurs, sur FROMAGE, comme observé précédemment, ce sont les graphies qui ne fournissent pas une forme qui permettent de restituer la forme phonologique du mot qui présentent le plus de « -e » final.

Seul RHUME présente un fort décalage entre le taux de restitution de la finale selon le type de graphies. Nous verrons plus bas que cela s'explique par la graphie RHUM (/Rɔm/) concurrente, qui ne se distingue que par la lettre finale mais dont la forme phonologique est différente (RHUME = /rym/ vs RHUM = /rɔm/).

¹⁵⁰ Moyenne calculée parmi les propositions non conformes à la norme orthographique : graphies phonographiquement acceptables, approximantes et non acceptables.

Nous détaillons ci-dessous les résultats item par item :

Le « -e » final de **CERISE**

Parmi les graphies qui ne restituent pas la forme orthographique attendue mais dont le dernier phonème est /z/ ou /s/¹⁵¹ (ex. *ceurises*, *courize*, *cerisse*), la proportion de « -e » final est de 90% en moyenne, répartie de la manière suivante :

- 100% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable (ex. *cerrise*). Ce pourcentage est normal puisque le « -e » final conditionne la prononciation de la dernière consonne. Nous avons en effet considéré que les graphies telles *ceriz* ne permettaient pas de restituer la forme phonologique du mot ;
- 80% de ceux qui ont produit une forme phonographiquement non acceptable (ex. *cerisse*, *sourise*).

Cet item n'a pas été particulièrement bien réussi d'un point de vue orthographique par les enquêtés « bas » (classé 8^{ème} sur 16¹⁵²), par rapport à d'autres mots. En revanche, même lorsqu'ils ne connaissent pas l'orthographe exacte de ce mot (qui contient plusieurs difficultés phonographiques), les propositions des enquêtés tiennent compte, en majorité, de la nécessité d'un « -e » final associé à la dernière consonne. Ce phénomène s'observe également lorsque les graphies ne restituent pas le dernier phonème /z/ attendu : par exemple *cerisse*. Cela provient probablement de la fréquence d'un « -e » final lorsque les mots se terminent par une consonne audible (CRISE, PAGE, PATE, etc.) et qui est donc une configuration stable en français.

Le « -e » final de **FROMAGE**

Pour cet item, la présence d'un « -e » final permet la prononciation de la consonne finale /ʒ/. En effet selon le *Petit Robert électronique*, 1786 graphies en français se terminent par un « -g » mais ce graphème ne se prononce pas /ʒ/ dans cette position : COING, GAG. En revanche, le « -ge » en finale est très stable (1851 graphies) tout comme la configuration « -age » = /aʒ/ (PAGE, MONTAGE).

¹⁵¹ Pour cet item, le nombre important de graphies phonographiquement non acceptables parce que le dernier phonème proposé était /s/ et non /z/ (ex. *cerisse*) nous amène à prendre également en compte les graphies qui proposent comme dernier phonème /s/.

¹⁵² Se référer à la figure 10 (section 2.2.1 de ce chapitre), qui classe les mots selon leur réussite en orthographe lexicale chez les enquêtés « bas ».

Parmi les graphies qui ne restituent pas la forme orthographique attendue mais qui se terminent par /z/ ou /g/ (ex. *fromaj*, *fromag*), la proportion de « -e » final est de 85% en moyenne, ce qui va dans le sens de la stabilité de la forme « -age » indiquée dans le *Petit Robert électronique*. Ces graphies sont réparties de la manière suivante :

- 75% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable (ex. *fraumage*, *fromaj*¹⁵³) ;
- 94% de ceux qui ont produit une forme phonographiquement non acceptable (ex. *farmage*).

Ce mot apparaît à la fois comme consistant et fréquent dans Manulex-Infra. Il est le 3^{ème} mot le plus fréquent et le plus consistant parmi les 15 mots de la dictée présents dans la base (CPO¹⁵⁴ : 81,72 et U¹⁵⁵ : 188,16).

Le « -e » final de **TOMATE**

Parmi les graphies qui ne restituent pas la forme orthographique attendue mais dont le dernier phonème est un /t/ (ex. *taumatte*, *tomatate*) ou dont le graphème final est « t » (ex. *tomat*), la proportion de « -e » final est de 65% en moyenne avec :

- 54% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable (ex. *taumate*). Comme dans CERISE, ce pourcentage est normal puisque le « -e » final conditionne la prononciation de la dernière consonne « t ». Nous avons en effet considéré que les graphies telles *tomat* ne permettaient pas de restituer la forme phonologique du mot ;
- 75% qui ont produit une forme phonographiquement non acceptable (ex. *tomote*).

Pour cet item, nous observons une restitution plus faible du « -e » final lorsque les graphies permettent pourtant de restituer la forme phonologique de TOMATE. Cela est dû à une certaine récurrence de la finale « -at » que nous n’observons pas dans les autres mots. En effet, les graphies : *thomat*, *tomat* et *tomats* ont été produites par 18 scripteurs (parmi les 26 graphies phonographiquement acceptables). Malgré tout, comme pour CERISE et FROMAGE présentés ci-dessus, il semble que les scripteurs aient conscience de la régularité du « -e » final dans les mots du français, notamment si l’on exclut cette forme récurrente *tomat*). En

¹⁵³ Nous avons considéré que la graphie *fromaj* produite par 2 enquêtés était acceptable d’un point de vue phonographique.

¹⁵⁴ CPO : consistance dans le sens Phonologie-Orthographe (production écrite).

¹⁵⁵ U : fréquence orthographique du lemme.

effet, très peu de mots qui se terminent par « -at » ont une consonne audible /t/ ; le graphème « -t » sans « -e » final, devient plus souvent une finale muette comme dans MAT, RAT (auxquels s'opposent quelques mots comme MAT = /mat/).

Le « -e » final de **CONFITURE**

Parmi les graphies qui se terminent par le phonème /R/, (ex. *confitour, caunfieture*), la proportion de « -e » final est de 65% en moyenne, répartie de la manière suivante :

- 57% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable (ex. *comfiture*) ;
- 72% de ceux qui ont produit une forme phonographiquement non acceptable (ex. *confuture, kouifetoure*).

Dans le *Petit Robert électronique*, les mots qui se terminent par « -ure » (AGRICULTURE, PEINTURE) sont plus fréquents que ceux qui ont un « -e » (MUR, FÉMUR). En revanche, de nombreux mots en français se terminent par « -eur » (VAPEUR, MOTEUR). Parmi les graphies considérées comme phonographiquement acceptables, c'est un des deux mots (avec TOMATE) parmi notre sélection, qui contient le moins de « -e » final (57%) : *comfitur, confitur* qui sont deux réponses récurrentes.

Il semble que pour cet item, les enquêtés qui n'ont pas restitué la forme orthographique mais qui ont néanmoins proposé une graphie phonographiquement acceptable aient repéré que le « -e » final n'était pas obligatoire pour que la graphie restitue la forme phonologique de l'item.

Le « -e » final de **RHUME**

Cet item est un des mots qui a posé le plus de difficultés aux enquêtés « bas » (classé 15^{ème} sur 16). La difficulté phonographique de cet item se porte sur la valeur phonologique de /y/ qui varie selon la présence ou non du « -e » final : « -um » = /ɔm/ (ex. *maximum*) vs « -ume » = /ym/ (ex. *plume*). Aussi pour cette analyse, nous n'avons conservé que les graphies qui se terminent par /ym/ (ex. *reume*) ou les graphèmes « -um » (ex. *rhûm*) avec ou sans « -e » final. Nous avons ainsi exclu les graphies telles *roume, rim*.

Parmi les graphies qui ne restituent pas la forme orthographique attendue mais qui se terminent par /ym/ (ex. *reume*) ou « -um » (ex. *rum*), la proportion de « -e » final est de 46% en moyenne. Elle est répartie de la manière suivante :

- 79% des enquêtés « bas » qui ont produit une forme graphique phonographiquement acceptable (ex. *ruhme*) ;
- 8% de ceux qui ont produit une forme phonographiquement non acceptable (ex. *rueme*).

2.5.2.3 BILAN SUR LA RESTITUTION DU « -E » FINAL

Nous avons proposé dans les deux sections précédentes, une analyse de deux sélections de mots : PHARMACIE, ÉPICERIE, LIBRAIRIE pour qui le « -e » final est une marque flexionnelle et TOMATE, FROMAGE, CERISE et RHUME pour qui en revanche, le « -e » final permet la réalisation phonologique de la dernière consonne (/t/, /ʒ/ et /z/) pour les trois premiers et la réalisation phonologique de la dernière voyelle dans le cas de RHUME (qui sinon fait /rɔm/). Dans cette sélection, nous avons ajouté TOMATE et CONFITURE pour qui la présence ou non d'un « -e » final n'a pas de conséquence sur la réalisation phonologique de la dernière consonne.

Il apparaît en premier lieu que le « -e » final a été moins bien restitué sur les mots lorsqu'il s'agit d'une marque flexionnelle, notamment ÉPICERIE sur qui on observe 49% de « -e » final en moyenne parmi les graphies qui ne sont pas conformes à la norme orthographique. Pour les mots sur lesquels la présence d'un « -e » final permet la réalisation phonologique de la dernière consonne, la restitution du « -e » final est particulièrement importante sur FROMAGE et CERISE (respectivement 85% et 90%). En revanche, sur l'item CONFITURE, le « e » final n'a été proposé que par 57% des enquêtés qui ont proposé une forme graphique plausible. En revanche une proportion importante d'enquêtés (72%) a proposé un « -e » en finale même lorsque la graphie proposée ne permettait pas de restituer la forme phonographique attendue : ex. *confutire*, *coinfuture*. Ce phénomène est également observé sur les graphies orthographiquement non conformes de TOMATE (54% des propositions qui ont pour dernier phonème /t/), avec la récurrence de la graphie *tomat*. Cela confirme que même les enquêtés les plus en difficulté ont enregistré des régularités de la langue française et/ou sont capables de restituer partiellement l'orthographe d'un mot.

Enfin, sur RHUME, l'absence de « -e » final produit la graphie RHUM (/Rɔm/) qui n'est pas acceptable d'un point de vue phonographique. La faible proportion de graphies qui contiennent un « -e » final malgré les erreurs phonographiques s'explique par la graphie *rhum* produite par 51 scripteurs. La présence du « h » semble indiquer que les scripteurs ont confondu la graphie de la boisson (RHUM) avec celle de l'item attendu, ce qui semble avoir perturbé la restitution orthographique de ce mot.

La proportion de restitution correcte du « -e » final est particulièrement élevée sur les items FROMAGE et CERISE, pour qui sa présence est indispensable à la restitution phonologique.

Il apparaît néanmoins que les enquêtés « bas » (y compris chez ceux qui ont le plus de difficultés), possèdent une bonne maîtrise des régularités du français sur les finales muettes, que l'on observe notamment dans la restitution des marques flexionnelles. Le « -e » final a été moins bien restitué que celui des autres mots de la sélection mais il faut préciser que les mots PHARMACIE, LIBRAIRIE et ÉPICERIE font partie des items considérés comme les moins fréquents dans la base Manulex-Infra (respectivement rang 10, 11 et 13 sur 15) à la différence de FROMAGE et TOMATE (rang 3 et 6).

2.6. CONCLUSION GÉNÉRALE SUR LES 447 ENQUÊTÉS « BAS »

2.6.1. LES DIFFÉRENTS SCORES CHEZ LES ENQUÊTÉS « BAS »

Nous récapitulons enfin dans la figure 17, le score moyen obtenu par les 441 enquêtés « bas » dans les différentes composantes de l'écrit que nous avons évaluées à partir de la dictée d'IVQ : l'orthographe lexicale sur 16 mots, la morphographie à partir des deux accords en nombre attendus, la phonographie sur 6 pseudo-mots. Nous indiquons également le score total sur 24 points obtenu par l'ensemble des enquêtés « bas » (447 enquêtés), y compris ceux qui ont obtenu 0 sur 24 parce qu'ils n'ont pas produit d'écrit. Nous rappelons à ce sujet que les enquêteurs ont indiqué qu'ils n'avaient pas pu produire de dictée en raison de difficultés trop importantes à l'écrit.

Nous observons la moyenne des différents scores dans chaque catégorie (ou profil de compétences) établie par l'INSEE :

- en grave difficulté avec moins de 40% de réussite ;
- en forte difficulté avec un score compris entre 40 à 60% ;
- en difficulté, mais ni grave ni forte avec un score compris entre 60 et 80% ;
- pas de difficulté avec plus de 80% de réussite ;

Nous ajoutons à ces quatre catégories :

- aucune difficulté, avec 100% de réussite.

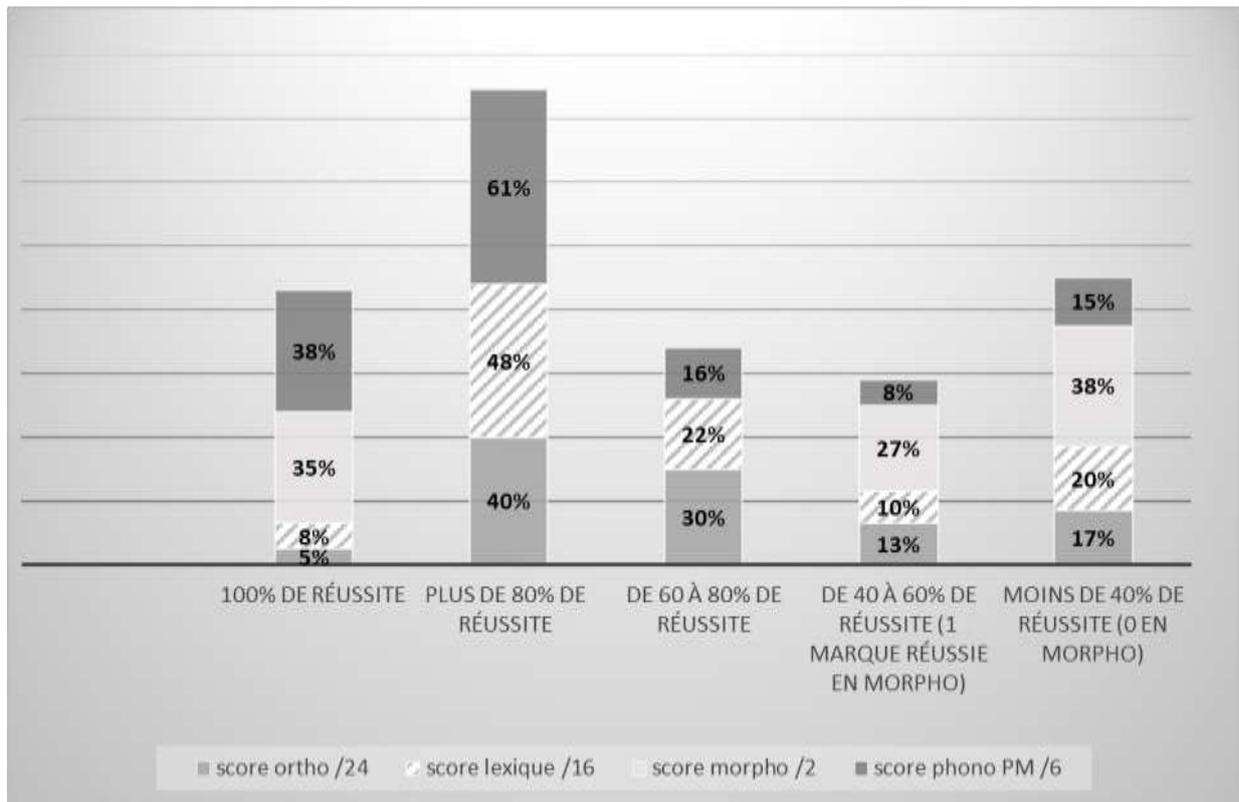


Figure 17. Scores moyens des 441 enquêtés « bas » obtenus en orthographe, en phonographie (sur les pseudo-mots) et en morphographie, répartis selon 5 profils de compétences.

Nous commençons cette synthèse par le score orthographique, sur 24 points, qui regroupe toutes les composantes de l'écriture évaluées dans cette dictée par l'INSEE : les compétences en orthographe lexicale observées sur les mots, les compétences en orthographe grammaticale d'après l'accord en nombre d'un nom et d'un verbe, les compétences en phonographie d'après la graphie de pseudo-mots. L'évaluation de l'ensemble de ces compétences montre que 5% des enquêtés « bas » de notre échantillon (23 enquêtés) ont réussi toutes les difficultés analysées (lexicales, morphographiques et phonographiques) et ont obtenu un score de 24 sur 24. Lorsque l'on sélectionne les dictées qui contiennent entre 80% et 100% de réussite sur les 24 critères évalués, ce sont 40% des enquêtés « bas » à qui cet exercice de production écrite n'a pas posé de problème particulier. Inversement, pour 17% des enquêtés « bas », l'exercice de production écrite révèle d'importantes difficultés à l'écrit, avec des scores inférieurs à 9 sur 24.

Nous observons à présent les différents scores.

LA MORPHOGRAPHIE ET L'ORTHOGRAPHE LEXICALE

L'orthographe grammaticale apparaît comme une compétence particulièrement discriminante dans cette dictée. La majorité des enquêtés « bas » (38%) n'a réussi aucune des deux marques de nombre et 27% n'ont réussi à produire qu'une marque (généralement le

nom FROMAGES). Chez les enquêtés qui ont obtenu les meilleurs résultats à la dictée (plus de 80% de réussite sur l'exercice complet), le score de réussite - lorsque l'on ne prend en compte que les compétences lexicales - passe de 40% (tous scores confondus) à 48%.

Ce sont en revanche les enquêtés qui possèdent les meilleures compétences lexicales, qui ont le mieux réussi la morphographie :

- 33 des 36 enquêtés (soit 92%) qui ont réussi à orthographier correctement les 16 mots analysés ont également réussi les deux marques de nombre ;
- 78 des 95 des enquêtés (soit 82%) qui ont fait 1 erreur orthographique au maximum sur le lexique ont également réussi les deux marques de nombre ;
- à l'inverse, sur les 235 enquêtés qui ont réussi moins de 80% du lexique, seuls 21 enquêtés (soit 9%) ont réussi les deux marques de nombre.

LES COMPÉTENCES EN PHONOGRAPHIE SUR LES PSEUDO-MOTS

Les pseudo-mots, en revanche ont été plutôt bien réussis par les enquêtés « bas » puisque plus de la moitié (61%) ont réussi à transcrire plus de 80% des 6 pseudo-mots.

Ainsi, alors que ce public a été sélectionné dans l'enquête IVQ pour ses faibles compétences à l'écrit, d'après les tests qui composent le module d'orientation (compréhension de l'oral, lecture et numératie), presque la moitié d'entre eux (48%) possèdent des compétences lexicales suffisantes pour restituer l'orthographe de plus de 80% des mots de la dictée. Pour la morphographie, ce sont 35% des enquêtés « bas » qui ont réussi les deux marques de nombre de cette dictée et 27% une seule des deux marques. La morphographie semble donc être un problème pour plus de la moitié des enquêtés « bas ». Enfin, leurs compétences en phonographie sont plutôt bonnes puisque 61% sont parvenus à proposer une graphie plausible sur plus de 80% des pseudo-mots. Leurs compétences dans ce domaine observées sur les mots (phonographie acceptable)¹⁵⁶ lorsque ceux-ci contiennent au moins une erreur orthographique, montre un taux moyen de réussite de 71% sur les mots. Ce sont ainsi 22% des enquêtés « bas » qui ont à transcrire de manière phonographique, l'intégralité des mots de la dictée analysés et 68% qui ont proposé une graphie acceptable sur plus de 80% des mots.

¹⁵⁶ Voir section 2.3.2 pour une analyse des compétences en phonographie observées sur les mots, chez les 441 enquêtés « bas ».

2.6.2. DES DIFFÉRENCES INTER-RÉGIONALES DANS LES COMPÉTENCES ORTHOGRAPHIQUES ?

Nous nous questionnons enfin sur l'existence de différences, entre les régions enquêtées, dans les compétences d'écriture des scripteurs en difficulté.

Une comparaison des dictées d'IVQ entre les régions Haute-Normandie et Rhône-Alpes a fait l'objet d'une publication (Blondel, Conseil, 2015¹⁵⁷). Cette analyse avait pour objectif de déterminer le niveau de maîtrise de l'écriture d'enquêtés ayant répondu au module « bas » avec un calcul du score de réussite orthographique et phonographique sur 16 mots et définir leur profil de scripteur. Ce travail a permis la comparaison des résultats entre les deux régions afin de voir s'il existait des différences régionales dans les réponses des enquêtés. Une analyse de la variation présente sur chaque mot est proposée, ainsi que la mise en relief des formes graphiques récurrentes produites sur les items qui ont posé le plus de difficulté aux enquêtés avec moins de 60 % de réussite orthographique : ANTI, RHUME et SOLENNEL.

En Rhône-Alpes, 822 personnes ont été interrogées dans le cadre d'IVQ et parmi elles, 108 enquêtés ont été considérés en difficulté avec l'écrit suite au module « d'orientation » et ont passé le module ANLCI, ce qui représente 13 % de l'échantillon total. La Haute-Normandie a bénéficié d'une extension de l'enquête IVQ, ce qui a permis d'interroger un plus grand nombre d'enquêtés. Ce sont ainsi 1622 personnes qui ont passé l'enquête IVQ en 2011 et parmi elles, 262 personnes ont été dirigées vers le module ANLCI, soit 16 % des enquêtés. Les analyses portent donc sur un corpus total de 370 dictées.

Selon les scores établis par l'INSEE, les régions Haute-Normandie (Brendler, 2013) et Rhône-Alpes présentent des résultats similaires. En effet, au terme de l'enquête, une proportion semblable de personnes (81 % en HN et 83 % en RA) sont considérées comme n'ayant pas de difficulté particulière avec l'écrit. Le classement des enquêtés au sein du module ANLCI est très proche, les Hauts-Normands sont toutefois un peu plus nombreux à présenter des difficultés graves ou fortes dans au moins un des domaines fondamentaux de l'écrit (60 % contre 57 % pour RA).

L'analyse des performances lexicales sur les mots a montré des résultats similaires dans les deux régions. Les mots les mieux réussis (SEL, TOMATE, CERISE et FROMAGE-S) sont les mêmes et les deux mots les mieux restitués (SEL et TOMATE) approchent même 90 % de

¹⁵⁷ Cette partie est issue de l'article de Blondel et Conseil (2015) paru dans Glottopol : http://glottopol.univ-rouen.fr/numero_26.html

réussite dans les deux régions. Deux mots ont été particulièrement mal restitués dans les deux régions : RHUME et SOLENNEL qui présentent des taux de réussite moyens inférieurs à 50 %.

L'analyse de la réussite phonographique observée parmi les réponses dont l'orthographe est erronée (sur 16 mots de la dictée) montre également un taux de réussite phonographique semblable dans les deux régions. Dans la région Rhône-Alpes, 42 % des réponses qui ne sont pas conformes à la norme orthographique respectent néanmoins le principe phonographique, tout comme 47 % des enquêtés en Haute-Normandie.

Les scripteurs des deux régions ont donc globalement les mêmes difficultés à proposer une forme graphique acceptable au moins d'un point de vue phonographique pour certains mots de la dictée qui sont complexes, car leurs correspondances entre phonèmes et graphèmes sont moins régulières. En revanche, il n'a pas été observé de manière systématique, de lien entre la réussite orthographique et la réussite phonographique sur les formes non conformes. En d'autres termes, ce ne sont pas nécessairement les mots qui ont été les moins bien réussis sur le plan orthographique qui sont également les moins bien réussis sur le plan phonographique. Certains mots particulièrement mal restitués d'un point de vue orthographique ont été bien restitués de manière phonographique, comme SOLENNEL). À l'inverse, certains mots sur lesquels il a été observé une bonne réussite orthographique (comme FROMAGE-S) présentent des graphies non conformes à la norme qui ne sont pas non plus acceptables sur le plan phonographique. Dans notre travail, nous avons développé cet aspect des difficultés orthographiques à partir des travaux en psycholinguistique qui traitent de la consistance et de la fréquence des mots.

La comparaison des résultats dans les deux régions n'a pas permis de différences significatives. Si quelques écarts de réussite ont été observés (Blondel, Conseil, 2015), différences qui peuvent être imputées à la taille de l'échantillon (la Haute-Normandie ayant bénéficié d'une extension de l'enquête IVQ), la tendance générale reste similaire. Ainsi, bien qu'ayant recours à une orthographe éloignée de la norme, une majorité de scripteurs dans les deux régions, possède de réelles compétences phonographiques, non négligeables, qui leur permettent de transcrire l'ensemble des mots. Et malgré le calcul d'un score phonographique qui valorise les compétences à l'écrit des scripteurs les plus en difficulté avec l'écrit, une proportion similaire d'enquêtés en Rhône-Alpes et en Normandie (respectivement 15 % et 14 %) présentent toujours des difficultés « graves ou fortes » dans la tâche de production écrite. Ces scripteurs, dont les compétences en orthographe et en phonographie sont faibles, se trouvent dans une situation particulièrement préoccupante.

Ce travail de comparaison entre deux régions, constitue une première approche de ce corpus de dictées produites par des scripteurs en difficultés dans l'enquête IVQ. Le travail sur ce corpus de dictées a été approfondi dans le cadre des deux recherches doctorales. La thèse menée en Haute-Normandie au sein du laboratoire *Dylis* de l'Université de Rouen a été financée dans le cadre du projet PEON¹⁵⁸. Dirigée par Marie-Claude Penloup et co-encadrée par Clara Mortamet, elle a été présentée et soutenue par Jeanne Conseil en octobre 2017 sous le titre « *Situations d'illettrisme et difficultés à l'écrit en Haute-Normandie : le cas de l'orthographe* ».

Dans le chapitre suivant, nous complétons les analyses proposées sur les 441 dictées du corpus quart Sud-Est par l'apport de 87 dictées issues d'une enquête complémentaire à IVQ 2011, menée dans la région Rhône-Alpes en 2014

¹⁵⁸ Pratiques et enseignement de l'orthographe en Normandie.

CHAPITRE 6. COMPÉTENCES
ORTHOGRAPHIQUES ET PRATIQUES
DE L'ÉCRIT CHEZ LES ADULTES :
APPORT DE DEUX ENQUÊTES

1. COMPARAISON DES DICTÉES PRODUITES PAR LES ENQUÊTÉS « BAS » ET « HAUT ».....	330
1.1. RÉPARTITION ET COMPARAISON DES SCORES.....	330
1.2. COMPARAISON DES COMPÉTENCES EN ORTHOGRAPHE LEXICALE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT ».....	332
1.3. COMPARAISON DES COMPÉTENCES EN PHONOGRAPHIE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT ».....	334
1.4. COMPARAISON DES COMPÉTENCES EN MORPHOGRAPHIE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT ».....	340
1.5. PRÉSENTATION DE LA VARIATION SUR LES MOTS ET LES PSEUDO-MOTS CHEZ LES ENQUÊTÉS « BAS » et « HAUT ».....	344
2. COMPÉTENCES ORTHOGRAPHIQUES, PRATIQUES D'ÉCRITURE ET REPRÉSENTATIONS DE L'ÉCRIT DE 87 SCRIPTEURS.....	356
2.1. RÉSULTATS GÉNÉRAUX DU « RAYON BRICOLAGE ».....	356
2.2. PRATIQUES D'ÉCRITURE ET REPRÉSENTATIONS DE L'ÉCRIT : UNE HÉTÉROGÉNÉITÉ DE PROFILS ET DE PRATIQUES.....	364

Dans ce chapitre, nous présentons les résultats de l'enquête complémentaire menée auprès de 87 scripteurs « haut »¹⁵⁹. Cette enquête a été conçue pour permettre de mieux interpréter les productions des enquêtés « bas », en disposant d'une comparaison avec des écrits de scripteurs dont les résultats suite au module « d'orientation » de l'enquête IVQ ont montré qu'ils n'avaient pas de difficulté particulière avec l'écrit (enquêtés du module « haut » dans IVQ 2011).

Ce corpus complémentaire a été recueilli grâce au financement de la Délégation générale à la langue française et aux langues de France¹⁶⁰ (DGLFLF). C'est Jean-Pierre Jeantheau (chargé de mission à l'ANLCI) qui a été à l'initiative de ce projet. Il s'était rapproché d'universitaires pour leur proposer un partenariat avec l'université afin de procéder à une analyse plus fine des dictées issues de l'enquête IVQ (INSEE, 2011), auquel nous avons répondu.

Cet échantillon d'enquêtés « sans difficulté » avec l'écrit n'est pas représentatif de la population française, puisqu'il s'agit d'une partie seulement des enquêtés de l'enquête IVQ de 2011 : ceux qui ont accepté d'être réinterrogés (et non l'ensemble des enquêtés considérés sans difficulté avec l'écrit dans l'enquête IVQ). Sélectionnés sur la base du critère géographique (et avec un taux de perte relativement conséquent, trois ans après l'enquête réalisée par l'INSEE en 2011), les variables sociologiques (pourtant récupérables dans le fichier de 2011) ne peuvent pas être exploitées de manière quantitative. Nous les utiliserons à des fins qualitatives, pour étayer certains résultats

Nous commencerons par quelques éléments de présentation de cet échantillon à titre qualitatif, à partir des données de l'INSEE recueillies lors de l'enquête de 2011 (dans le module biographique).

PRÉSENTATION DE L'ÉCHANTILLON

L'échantillon comprend 87 personnes dont 54% d'hommes et 46% de femmes. Toutes les tranches d'âge sont représentées (de 20 ans à plus de 60 ans), avec une proportion plus importante d'enquêtés âgés de plus de 40 ans, ce que nous expliquons par la difficulté à retrouver les enquêtés les plus jeunes trois ans après l'enquête de l'INSEE. On peut notamment supposer que pour cette tranche d'âge, les déménagements sont plus fréquents.

¹⁵⁹ Pour une présentation de l'enquête complémentaire, se référer au chapitre 4, section 4.

¹⁶⁰ La DGLFLF a été sollicitée au titre de sa mission de soutien des actions visant « à développer la maîtrise partagée du français et à lutter contre l'exclusion linguistique ».

La majorité des enquêtés est née en France, avec seulement 9 personnes nées à l'étranger (Europe du nord : 2, Europe du sud : 3, Maghreb : 3, reste du monde : 1). Les données sociologiques de l'INSEE ne nous permettent malheureusement pas de connaître précisément le pays d'origine de ces enquêtés. Par ailleurs, nous savons également que seuls 5 enquêtés n'ont pas débuté leur scolarité en France. Nous sommes donc en présence d'un échantillon en grande majorité scolarisé dans un contexte francophone.

Tous les enquêtés ont été scolarisés (les plus âgés sont nés en 1946) et seulement 10 personnes (dont 9 sont nées entre 1946 et 1959) déclarent ne pas avoir de diplôme. 60% ont obtenu un baccalauréat (général, professionnel ou technique), 45% ont un bac+2 ou plus.

Nous disposons donc d'un échantillon mixte, âgé de 20 à 68 ans au moment de l'enquête avec une part plus importante de personnes âgées de plus de 40 ans. La majorité a obtenu un diplôme et près de la moitié a suivi des études supérieures.

Par contraste avec les enquêtés « bas » qui ont passé le module « bas », nous appellerons désormais ces enquêtés : enquêtés « haut ».

La première partie de ce chapitre présente les résultats obtenus par les 87 enquêtés « haut » sur la partie commune de la dictée (dictée d'IVQ) que nous comparons avec ceux obtenus par les 441 enquêtés « bas » (repérés en difficulté avec l'écrit en 2011). Pour éviter des redondances dans la présentation des résultats obtenus par deux populations sur une même dictée (score global, score sur chaque item, etc.), nous présentons les résultats des enquêtés « haut » en les comparant directement avec ceux obtenus par les enquêtés « bas ». Les données chiffrées concernant les enquêtés « bas » sont donc reprises du chapitre précédent.

Dans une deuxième partie, nous présentons les compétences orthographiques des enquêtés « haut » observées sur la partie complémentaire de la dictée (et qui ne concerne que les enquêtés « haut ») avec les items du rayon « bricolage ». Nous terminerons par une présentation de la partie « entretien » de cette enquête¹⁶¹.

¹⁶¹ Une partie de ces résultats ont fait l'objet d'une communication avec actes au Congrès Mondial de Linguistique Française (CMLF, Blondel, Brissaud, Rinck, 2016) : <http://dx.doi.org/10.1051/shsconf/20162707003> ainsi que d'une publication à paraître à la Délégation générale à la langue française et aux langues de France (Blondel, Brissaud, Rinck, à paraître).

1. COMPARAISON DES DICTÉES PRODUITES PAR LES ENQUÊTÉS « BAS » ET « HAUT »

1.1. RÉPARTITION ET COMPARAISON DES SCORES

Nous comparons dans la figure 18, le score global obtenu par les enquêtés « bas » et les enquêtés « haut » sur la dictée d'IVQ, calculé selon nos critères (voir chapitre 4 consacré à la méthodologie). Nous rappelons que pour notre analyse, les items ont été codés selon le respect de la forme orthographique pour les mots et le respect de la forme phonographique pour les pseudo-mots.

Pour rappel, la dictée a été notée sur 24 points, avec 1 point attribué par item réussi (sur les 16 mots et 6 pseudo-mots) et 2 points pour la morphographie (1 point pour le « -s » de FROMAGES et 1 point pour le « -nt » de SENTENT).

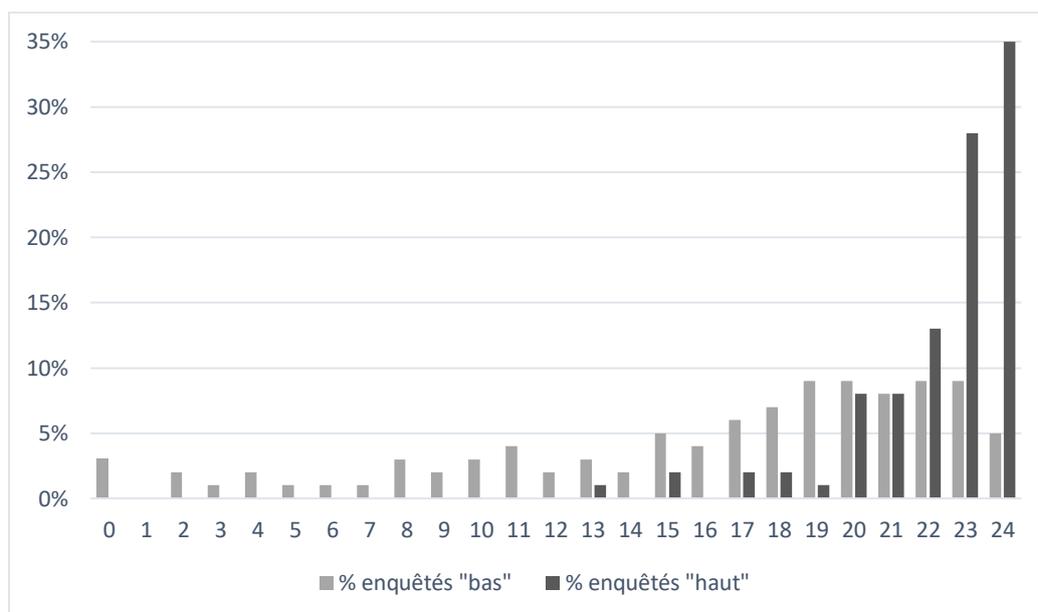


Figure 18. Répartition des scores (sur 24) obtenus à la dictée d'IVQ par les 447¹⁶² enquêtés « bas » et les 87 enquêtés « haut ».

La dispersion maximale des scores que nous observons chez les enquêtés « bas » ne s'observe pas chez les enquêtés « haut ». Chez ces derniers, le score le plus faible (et qui ne concerne qu'un seul enquêté) est de 13 sur 24. La majorité a en effet très bien réussi l'exercice : les scores sont concentrés entre 20 et 24 sur 24 avec une moyenne de 22 sur 24 (contre 16 sur 24 chez les enquêtés « bas ») et la médiane se trouve à 23 sur 24 (contre 18 sur 24 chez les enquêtés « bas »).

¹⁶² Pour rappel, dans le calcul du score total, l'intégralité des scores de l'échantillon (447 enquêtés « bas ») a été pris en compte (les 6 enquêtés qui n'ont pas produit de texte ont obtenu un score de 0 sur 24).

Ce sont ainsi 92% des enquêtés « haut » qui ont réussi plus de 80%¹⁶³ de la dictée (contre 40% chez les enquêtés « bas »). En se basant sur la dictée d'IVQ, 92% des enquêtés « haut » et 40% des enquêtés « bas » semblent avoir un profil orthographique similaire.

Ce résultat montre qu'un bon niveau de compétences en lecture ou en compréhension tels que révélés par le module « d'orientation » pour les enquêtés « haut », n'est pas nécessairement lié aux compétences orthographiques évaluées dans l'exercice de dictée. Chez les enquêtés « haut », le module « d'orientation » a permis de repérer les personnes les moins en difficulté avec l'écrit puisqu'aucun enquêté « haut » n'a obtenu de score inférieur à 13, mais il se révèle insuffisant pour affirmer qu'il existe un lien entre les compétences orthographiques et le reste des compétences évaluées (lecture et compréhension écrite). La dictée d'IVQ (qui ne comporte pas de difficultés orthographiques majeures) a ainsi posé quelques difficultés à certains enquêtés « haut ».

Il apparaît également que 40% des enquêtés repérés en difficulté avec l'écrit (enquêtés « bas ») ont obtenu plus de 80% de réussite sur la dictée, tout comme 91% des enquêtés « haut ». Inversement, 9% des enquêtés « haut » ont obtenu moins de 80% de réussite (tout comme 60% des enquêtés « bas ») sur un exercice de production écrite ne contenant pas de difficultés orthographiques majeures. Le profil orthographique de ces 8 enquêtés « haut » sera observé attentivement pour voir si les difficultés rencontrées sur cette partie de la dictée prédisent des difficultés particulières en orthographe (lexique ou morphographie) que nous pourrions davantage observer grâce à la partie complémentaire de la dictée (rayon bricolage) qui contient du lexique et des structures syntaxiques plus complexes. Par ailleurs, il conviendra d'observer le comportement des 92% d'enquêtés « haut » qui n'ont pas été en difficulté sur la dictée d'IVQ, lorsque l'exercice se complexifie.

Ainsi, il apparaît nécessaire de proposer un test d'écriture si l'on souhaite évaluer les compétences à l'écrit des enquêtés. En effet, une compétence en particulier (compréhension écrite et orale, lecture, écriture) ne permet pas nécessairement de prédire le niveau de compétences dans les différentes composantes de l'écrit ; celles-ci étant appelées à être évaluées indépendamment les unes des autres.

¹⁶³ Les enquêtés qui ont réussi plus de 80% de la dictée ont obtenu un score compris entre 20 et 24 sur 24.

1.2.COMPARAISON DES COMPÉTENCES EN ORTHOGRAPHE LEXICALE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT »

La figure 19 nous permet de comparer les compétences en orthographe lexicale des deux populations, sur 16 mots de la dictée classés selon leur pourcentage de réussite orthographique (du mot le mieux réussi chez les enquêtés « bas » au moins bien réussi). Nous rappelons que dans cette analyse, les marques de morphographie ne sont pas prises en considération.

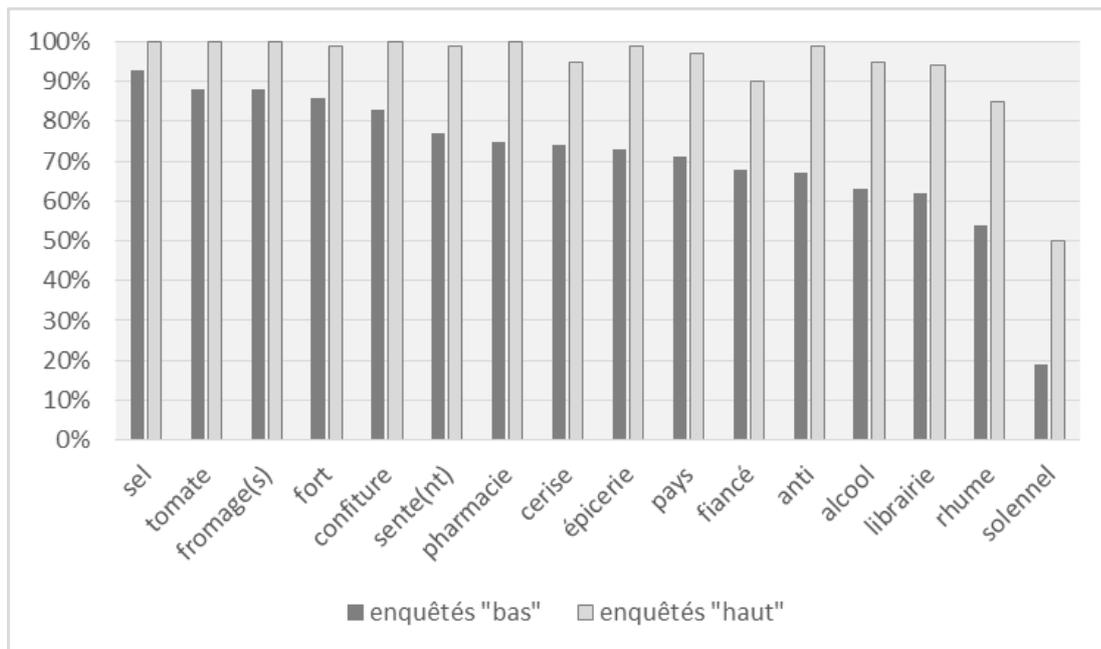


Figure 19. Compétences en orthographe lexicale des 441 enquêtés « bas » et des 87 enquêtés « haut » sur 16 mots de la dictée d'IVQ.

Pour les deux populations, il apparaît que les difficultés portent sur les mêmes mots, même si les taux de réussite diffèrent.

Dans les deux corpus, les mots particulièrement consistants (indice > à 80), qui, nous l'avons vu, sont également les plus fréquents du corpus (tels SEL, TOMATE, FROMAGE-S) ont été particulièrement bien réussis par les deux populations. Les trois mots les mieux réussis sont identiques chez les enquêtés « bas » comme chez les enquêtés « haut » : CONFITURE, SEL et TOMATE.

De même, le lexique qui a posé problème aux personnes les plus en difficulté à l'écrit a également davantage posé problème aux enquêtés « haut ». Les mots dont le degré de consistance est moindre (CERISE, PAYS) ou dont l'indice de fréquence est particulièrement faible (FIANCÉ, SOLENNEL) sont en effet les moins bien restitués chez les enquêtés « bas » comme chez les enquêtés « haut ». En revanche, les proportions d'erreurs sont bien

inférieures pour ces derniers. Chez les enquêtés « haut », seuls trois mots ont des scores de réussite inférieurs ou égal à 90% : FIANCÉ, RHUME et SOLENNEL alors que chez les enquêtés « bas », seul SEL obtient un score de réussite moyen supérieur à 90%.

Ce classement montre que les différences entre les deux populations portent plutôt sur les mots qui sont moins consistants ou moins fréquents (comme RHUME, FIANCÉ, SOLENNEL). Les mots plus fréquents (comme FORT ou PHARMACIE) ont été plutôt bien restitués par les enquêtés « bas » ainsi que les mots dont les correspondances phonie-graphie sont les plus régulières du corpus (CONFITURE, SEL, TOMATE) qui sont les mieux réussis par les deux populations (et particulièrement chez les enquêtés « haut » avec 100% de réussite).

Chez les enquêtés « bas », la réussite sur les items semble particulièrement liée à la régularité des combinaisons de phonèmes et de graphèmes et donc à l'effet de consistance. La présence de correspondances phonie-graphie irrégulières fait en effet baisser leurs résultats y compris sur les mots fréquents (ex. TOMATE réussi à 88% vs PHARMACIE réussi à 75%).

La fréquence des mots impacte également les résultats, chez les enquêtés « bas » et « haut » (ex. FIANCÉ, RHUME, SOLENNEL) et cet effet de fréquence est particulièrement présent chez les enquêtés « haut » lorsqu'il s'agit de mots très rares. Nous observons en effet un écart de réussite orthographique important entre les 16 mots de la dictée d'IVQ (réussis en moyenne à 97%) et SOLENNEL que seule la moitié des enquêtés « haut » a correctement restitué.

Inversement, l'item PAYS qui apparaît comme le plus fréquent du corpus tout en possédant une structure phonographique rare a été bien réussi par les enquêtés « haut » (97%), davantage que l'item FIANCÉ, plus consistant mais peu fréquent d'après Manulex-Infra.

Nous observons donc les mêmes fragilités face à l'orthographe parmi les enquêtés des deux populations, même si les difficultés ressortent de manière plus forte chez les enquêtés « bas », qui semblent davantage impactés par l'effet de consistance et de fréquence des mots.

L'analyse des graphies produites par les enquêtés « bas »¹⁶⁴ nous a permis d'observer que certains d'entre eux semblent avoir de véritables difficultés d'ordre phonographique. Ils confondent certains phonèmes comme /i/ vs /e/, /ã/ vs /õ/, /p/ vs /b/, ils inversent l'ordre des phonèmes ou ils ne restituent pas tous les phonèmes. La plupart possède néanmoins de bonnes compétences dans ce domaine, le problème à l'écrit se situant davantage sur la connaissance de la norme orthographique.

¹⁶⁴ Se référer au chapitre 5, section 2.3.2 pour une présentation des compétences en phonographie des enquêtés « bas » observées sur les mots, parmi les réponses non conformes aux attentes orthographiques.

L'utilisation de pseudo-mots peut permettre d'observer les comportements de scripteurs vis-à-vis des régularités du français, notamment l'utilisation des lettres finales muettes, analyse que nous proposons dans la deuxième partie dédiée aux pseudo-mots de la dictée d'IVQ.

1.3. COMPARAISON DES COMPÉTENCES EN PHONOGRAPHIE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT »

Dans cette section, nous comparons les compétences en phonographie des enquêtés « bas » et des enquêtés « haut » qui ont été observées sur les pseudo-mots et sur les mots.

1.3.1. COMPÉTENCES EN PHONOGRAPHIE SUR LES PSEUDO-MOTS

Nous comparons dans la figure 20, le pourcentage de réussite phonographique sur chaque pseudo-mot, dans les deux populations.

Les scores ont été calculés à partir des réalisations graphiques que nous avons considérées comme acceptables et approximantes. Pour faciliter la comparaison, les pseudo-mots sont classés du mieux réussi au moins bien réussi par les enquêtés « bas ».

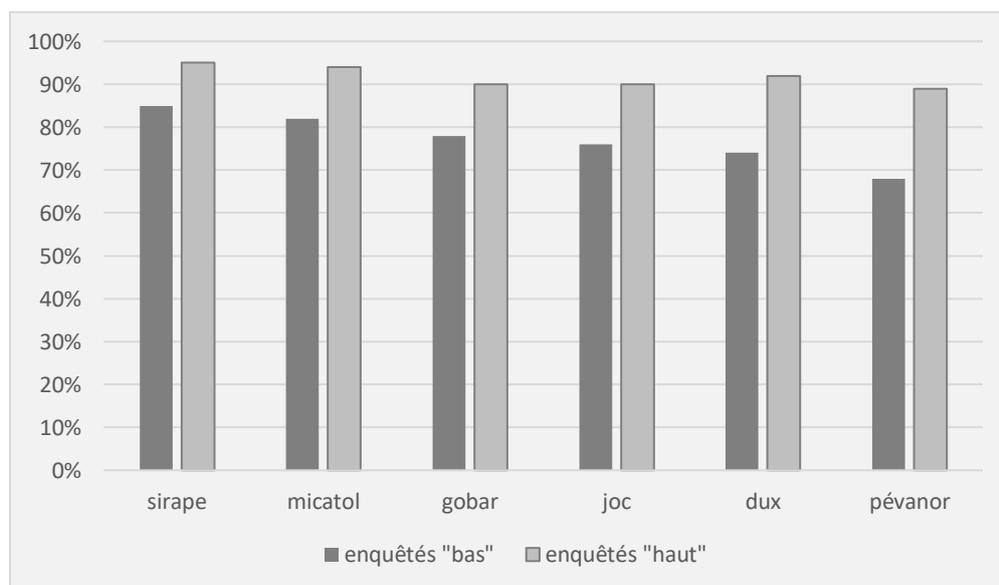


Figure 20. Comparaison de la proportion de réponses acceptables et approximantes obtenues par les 441 enquêtés « bas » et les 87 enquêtés « haut » pour chaque pseudo-mot.

Dans les deux populations, le classement des pseudo-mots par ordre de réussite est presque identique, sauf DUX que les enquêtés « haut » ont mieux restitué que GOBAR et JOC. Le pseudo-mot le mieux réussi (SIRAPE) et le pseudo-mot le moins bien restitué (PÉVANOR) sont les mêmes.

Chez les enquêtés « haut », les taux de réussite sur les pseudo-mots sont compris entre 89% et 95% de réussite et aucun n'atteint les 100% de restitution correcte observés sur certains mots¹⁶⁵. Les erreurs produites sur les mots chez les enquêtés « haut » sont majoritairement d'ordre orthographique¹⁶⁶ avec très peu d'erreurs d'origine phonographique. En d'autres termes, lorsque les enquêtés « haut » produisent des graphies non conformes aux attentes orthographiques, elles respectent néanmoins la forme phonographique attendue (exemple *alcol* pour ALCOOL ou *ruhme* pour RHUME), ce qui nous interpelle sur les résultats obtenus sur les pseudo-mots. En effet, sur l'item PÉVANOR, 11% des formes proposées ne sont pas plausibles phonographiquement (ex. *bivabor*, *Evanord*, *pevarnor*).

Dans le corpus « haut », nous observons également un taux de réussite plus bas que la moyenne sur les pseudo-mots, chez les enquêtés interrogés par un des enquêteurs (alors que 6 enquêtés sur les 11 interrogés ont pourtant obtenu de très bons scores à la dictée), et qui peut s'expliquer par la manière dont il a prononcé ces pseudo-mots.

Les résultats observés sur les pseudo-mots chez les enquêtés « haut » sont donc à nuancer, tout comme ceux des enquêtés « bas » car certaines erreurs pourraient provenir d'un biais du protocole de l'enquête lié à la prononciation de l'enquêteur¹⁶⁷.

Pour expliquer les différences de réussite entre les pseudo-mots (SIRAPE dans les deux corpus est mieux restitué que PÉVANOR) nous avons évoqué chez les enquêtés « bas », la composition phonémique de ces pseudo-mots, avec des correspondances phonie-graphie plus faciles à restituer que d'autres, peut-être parce qu'elles sont plus faciles à saisir à l'oral ou parce qu'elles relèvent de schémas plus fréquents à l'écrit. Cette explication peut être avancée chez les enquêtés « haut » également, même si comme sur les mots, le taux d'erreurs est moins important chez ces derniers.

Ces résultats nous questionnent sur la nécessité d'intégrer des pseudo-mots dans une dictée pour repérer des difficultés d'ordre phonographique chez les scripteurs. Les compétences en phonographie des enquêtés « bas » et « haut » peuvent en effet être saisies dans l'analyse des graphies erronées des mots de la dictée et nous avons observé que les difficultés des enquêtés « bas » dans ce domaine sont liées à des difficultés importantes avec l'écrit.

¹⁶⁵ Se référer à la figure 19 (section 1.2 du chapitre 6) pour voir les pourcentages de réussite obtenus en moyenne sur chaque mot, par les enquêtés « haut ».

¹⁶⁶ Se référer à l'annexe 3 et 3bis pour observer les graphies recueillies dans les dictées des enquêtés « haut », pour chaque mot.

¹⁶⁷ Se référer au chapitre 5, section 2.3.1, qui traite des hypothèses qui pourraient expliquer les écarts de réussite entre les pseudo-mots.

L'analyse des erreurs chez les enquêtés « haut » montre en revanche que ces derniers semblent être concernés par d'autres problématiques, d'origine lexicale et/ou morphographique, qui sont davantage observables sur la partie complémentaire de la dictée¹⁶⁸.

L'utilisation de pseudo-mots, en revanche, peut permettre d'observer les comportements à l'écrit des scripteurs à l'égard des finales muettes (utilisation ou non de lettres finales muettes et choix de la finale). Cette analyse est proposée dans la section 1.5.2.3 qui présente la variation observée sur les pseudo-mots chez les enquêtés « bas » et « haut ».

Dans la partie suivante, nous observons les compétences en phonographie des enquêtés « haut » d'après les réponses non conformes à la norme orthographique.

1.3.2. COMPÉTENCES PHONOGRAPHIQUES SUR LES MOTS CHEZ LES ENQUÊTÉS « HAUT »

L'analyse des compétences en orthographe lexicale des enquêtés « haut » sur la partie IVQ de la dictée (figure 19, section 1.2) a montré que ces derniers possèdent un bon niveau de connaissances dans ce domaine, avec 9 mots sur les 16 analysés qui obtiennent entre 98% et 100% de réussite orthographique.

Quatre mots obtiennent 100% de réussite : TOMATE, CONFITURE, FROMAGE-S, SEL et cinq mots contiennent très peu d'erreurs orthographiques, que nous détaillons ci-dessous :

- PHARMACIE : 1 graphie non conforme à la norme orthographique et également phonographiquement non acceptable (*farmaice*) sur laquelle on peut observer une inversion de deux graphèmes « a » et « i » qui sans cela, aurait conduit à une forme phonographiquement acceptable avec une erreur centrée sur le choix du graphème pour transcrire /f/ ;
- FORT : 1 graphie considérée comme phonographiquement acceptable (*ford*) ;
- ANTI : 2 graphies non conformes à la norme mais phonographiquement acceptables (*enti*), qui portent sur le choix du graphème pour transcrire /ã/ ;
- ÉPICERIE : 1 graphie phonographiquement acceptable (*épiceri*) sur laquelle il manque le « -e » final muet et 1 graphie approximante (*episerie*) dont l'erreur porte sur le choix

¹⁶⁸ Se référer chapitre 6, section 2.1.2 pour une analyse des compétences lexicales des enquêtés « haut » sur les mots du « rayon bricolage » (dictée complémentaire).

du graphème pour transcrire /s/ avec une proposition ici qui ne tient pas compte de la règle de position du graphème « s » ;

- SENTENT : 1 graphie approximante (*sent*) et une absence de réponse.

Nous observons dans la figure 21, le type d'erreurs (phonographie acceptable, approximante ou non acceptable) relevées sur les 7 mots qui ont obtenu entre 52 % et 93% de réussite orthographique : RHUME, ALCOOL, PAYS, CERISE, LIBRAIRIE, FIANCÉ, SOLENNEL. Les mots sont classés du mieux réussi orthographiquement au moins bien réussi.

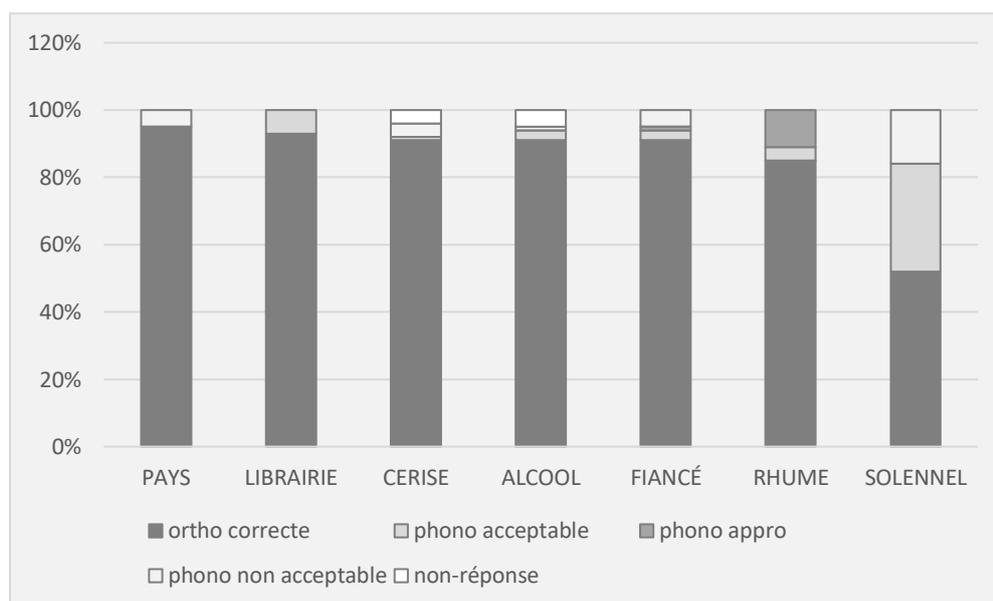


Figure 21. Répartition des réponses sur les 7 mots qui ont obtenus entre 52 % et 93% de réussite orthographique chez les 87 enquêtés « haut ».

Chez les 87 enquêtés « haut », la majorité des erreurs est d'origine orthographique avec peu de graphies erronées qui ne restituent pas à minima la forme phonographique des mots.

Nous présentons les résultats pour chaque mot.

PAYS : ce mot fait partie de ceux qui ont posé problème aux enquêtés « bas ». Parmi les réponses non conformes relevées chez les enquêtés « bas », c'est la réalisation graphique particulière des phonèmes /e/ et /i/ par la combinaison de graphèmes « ays » qui a posé problème à 4 enquêtés : *pay* (3), *paye* (1). Ici aussi, nous retrouvons la structure complexe de ce mot avec « pay » présent dans ces deux graphies.

LIBRAIRIE : sur cet item, les 6 graphies non conformes aux attentes orthographiques (*librerie* et *librerie*) contiennent pour seule erreur le choix du graphème pour transcrire le phonème /e/.

CERISE : sur cet item, les erreurs phonographiques portent sur le choix du graphème pour transcrire le /z/ final : *cerice* (1), *cerices* (1), *serice* (1).

ALCOOL : deux propositions non conformes respectent la forme phonologique du mot et portent sur le choix du graphème rare « oo » : *alcol* (2). Par ailleurs, nous relevons une graphie non acceptable phonographiquement : *accol*, qui cumule le type d'erreurs (omission d'un graphème, redoublement de consonnes non conforme, substitution du graphème attendu par un autre graphème (« o » à la place de pour « oo »).

FIANCÉ : sur cet item, les difficultés phonographiques qui ont été relevées portent sur le dernier phonème : /e/ : *fiance* (1), *fiancè* (1)¹⁶⁹. Deux graphies cumulent le type d'erreurs : *fiancier* (1) et *financier* (1).

RHUME : c'est sur ce item que se concentrent le peu de graphies qui ont été considérées comme approximantes dans le corpus « haut » avec *rhum* (8) qui rappelle les nombreuses confusions faites par les enquêtés « bas » sur ce mot qui désigne une boisson. On trouve également deux formes acceptables d'un point de vue phonographique mais assez étonnantes dans ce corpus « haut » : *rum* (2).

SOLENNEL a été particulièrement mal restitué chez les enquêtés « bas » et chez les enquêtés « haut ». Même si le taux de graphies erronées est moins important chez ces derniers, c'est le mot qui a néanmoins posé le plus de difficultés. Parmi les propositions phonographiquement non acceptables, c'est principalement le choix du graphème pour transcrire /a/ (/solanel/) qui pose problème : *solenel* (10), *solenelle* (1), *sollenel* (1) et *sollenelle* (1). Toutes ces propositions contiennent bien le graphème « e » attendu, mais les enquêtés n'ont pas doublé le « n », ce qui empêche la réalisation phonologique particulière de cette combinaison de graphèmes. Enfin, nous relevons une graphie que l'on ne s'attend pas à trouver chez un scripteur qui n'a pas été considéré en difficulté avec l'écrit dans IVQ : la graphie *sonalell* (1) qui cumule le type d'erreurs.

En résumé, chez les enquêtés « haut », les écarts à la norme orthographique portent généralement sur les mêmes difficultés : la transcription de /a/ par « en » dans SOLENNEL, le graphème « ai » de LIBRAIRIE, la structure complexe « ays » de PAYS. La grande majorité des enquêtés « haut » maîtrisent donc les règles de fonctionnement et de distribution relative au français et parviennent à proposer une forme graphique plausible lorsqu'ils ne connaissent

¹⁶⁹ Nous avons initialement considéré les graphies *fiance* et *fiancè* phonographiquement non conformes chez les enquêtés « haut ». Nous avons finalement neutralisé l'accent sur le dernier graphème, comme nous l'avons fait pour les enquêtés « bas ». Dans le corpus « haut », cela ne concerne que 2 enquêtés.

pas l'orthographe d'un mot : par exemple pour SOLENNEL : *solanel* (18), *solannel* (5), *solanelle* (2), *solannelle* (2), *sollanel* (1). Les enquêtés « haut » maîtrisent ces règles davantage que les enquêtés « bas » si l'on regarde les réponses recueillies sur cet item avec 84% de réponses conformes à la norme orthographique et au critère phonographique contre 63% chez les enquêtés « bas ». Sur un item dont la forme orthographique est moins complexe, comme FIANCÉ, cet écart demeure important avec 18 points d'écart (94% de réponses conformes à la norme orthographique et au critère phonographique chez les enquêtés « haut » contre 76% chez les enquêtés « bas »).

Nous relevons néanmoins quelques graphies qui comportent une ou plusieurs erreurs phonographiques : *farmaice* (PHARMACIE), *accol* (ALCOOL), *paye* (PAYS), *serice* (CERISE), *financier*, *fiancier* (FIANCÉ), *sonalell* (SOLENNEL). Nous pourrions penser que ces graphies se concentrent sur un nombre restreint d'enquêtés, peut-être même un seul qui serait en difficulté plus importante dans le domaine de la phonographie. Mais l'analyse des dictées montre que ces graphies sont réparties dans différentes dictées.

À titre d'exemple, nous observons les réponses relevées dans la dictée de l'enquêté n°137, qui a obtenu chez les enquêtés « haut » le score en orthographe lexicale le plus faible (9 sur 16) et le score phonographique calculé sur les mots le plus faible également (11 sur 16) sur la dictée d'IVQ. Nous soulignons les graphies qui ne sont pas conformes à la norme orthographique pour les mots et aux attentes phonographiques pour les pseudo-mots :

pharmacie, anti, rhume, mikatol, sirap, épiceri, tomate de paye, confiture, cerice, sel, gobare, 3 fromage qui ne sent pas fort, librairie, le fiancée sonalell de jok P.VNOR.

Dans cette dictée, nous constatons qu'un syntagme n'a pas été transcrit : ALCOOL À 90° DUX, ce qui fait chuter le score orthographique, ainsi que le score sur 16 qui ne concerne que l'orthographe lexicale. Trois mots sur les 15 mots qui ont été transcrits comportent des erreurs phonographiques : *paye, cerice, sonalell* mais d'autres mots ont une orthographe conforme : *pharmacie, anti, rhume, fiancée* (si l'on ne prend en compte l'erreur d'ordre morphographique). Par ailleurs, 4 pseudo-mots sur les 5 qui ont été transcrits ont une forme phonographique acceptable.

Cet enquêté se trouve ainsi en difficulté sur certains mots dont il ne connaît pas l'orthographe : PAYS (*paye*) et SOLENNEL (*sonalell*) en particulier, mais également ÉPICERIE (*épiceri*), CERISE (*cerice*) et LIBRAIRIE (*librairie*). Il réussit néanmoins d'autres mots, consistants comme TOMATE et plus complexes comme PHARMACIE.

1.4.COMPARAISON DES COMPÉTENCES EN MORPHOGRAPHIE CHEZ LES ENQUÊTÉS « BAS » ET « HAUT »

1.4.1. SCORES DE RÉUSSITE

Nous comparons ensuite les compétences en morphographie des 441 enquêtés « bas » et des 87 enquêtés « haut » d'après l'accord en nombre du nom FROMAGES, du verbe SENTENT et l'absence d'accord de FORT.

Dans la figure 22, les taux de réussite (accord en nombre ou respect du singulier) ne concernent que la morphographie et les mots sont classés de l'accord le mieux réussi à l'accord le moins bien réussi par les enquêtés « bas ».

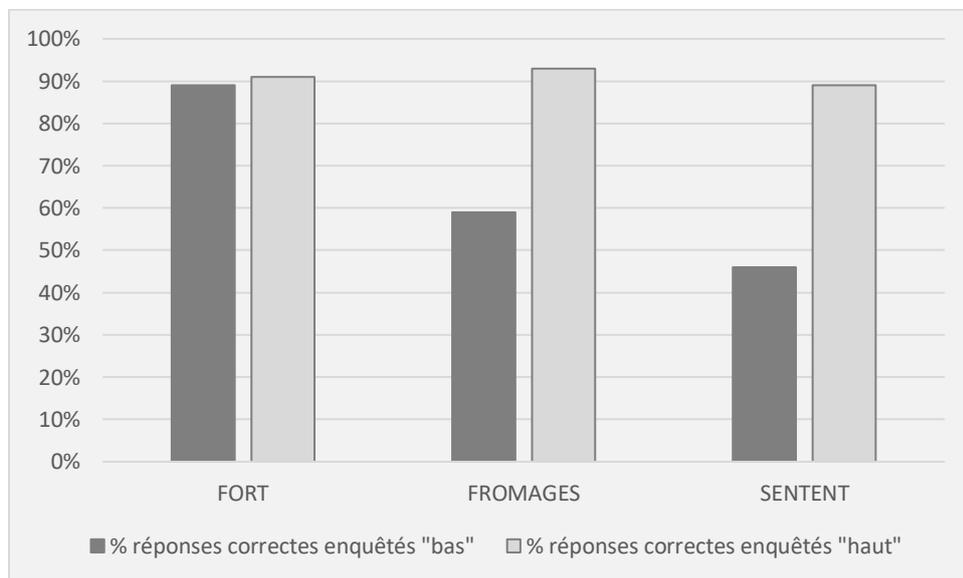


Figure 22. Comparaison du taux de réussite en morphographie chez les 441 enquêtés « bas » et les 87 « haut ».

Si l'on compare les performances en morphographie des deux populations sur les trois items qui permettent cette comparaison, on remarque que c'est sur l'adverbe FORT qui correspond à une absence de marque, que les résultats sont les plus proches (91% pour les « haut » vs 89% pour les « bas »).

Pour les enquêtés « bas », il semble donc plus facile de produire une forme non marquée, même si comme nous l'avons évoqué dans l'analyse des dictées du corpus « bas », nous ne pouvons savoir si l'absence de marque de nombre provient bien d'une volonté de respecter l'absence d'accord sur cet adverbe.

Concernant l'accord en nombre du nom FROMAGES et du verbe SENTENT, la différence de performance entre les enquêtés « bas » et « haut » est en revanche très marquée. Cette différence est davantage marquée sur l'accord du verbe, avec 43 points d'écart entre les

enquêtés « haut » (89% de réussite) et les enquêtés « bas » (46% de réussite) mais elle également importante sur celui du nom FROMAGES avec 34 points de différence (93% vs 59%).

Chez les enquêtés « bas », la différence de réussite sur les trois items est particulièrement marquée entre les deux marques de nombre (réussies à 59 et 46%) et l'absence de marque (réussie à 89%). Sur les deux marques de nombre, c'est l'accord en nombre du verbe qui a posé le plus de difficultés aux enquêtés « bas ».

En revanche chez les enquêtés « haut », cette différence de réussite selon le type de difficulté (respect de l'absence de marque sur un adverbe et accord en nombre d'un nom et d'un verbe) est très peu marquée, avec un écart compris entre 2 et 4 points. La catégorie grammaticale des deux accords en nombre n'affecte quasiment pas les performances des enquêtés « haut », même si l'accord en nombre du nom FROMAGES a été légèrement mieux réussi que celle du verbe SENTENT (tout comme chez les enquêtés « bas » pour qui en revanche la différence de réussite entre les deux catégories est bien plus marquée).

Il aurait cependant été intéressant de tester les compétences en morphographie, sur le rayon bricolage, des enquêtés « bas » qui ont obtenu, comme les enquêtés « haut », de très bons résultats à la dictée d'IVQ (plus de 80% de réussite). En effet, 33%¹⁷⁰ des enquêtés « bas » ont très bien réussi la morphographie de la dictée d'IVQ en obtenant 89% de réussite sur FROMAGES, 95% de réussite sur SENTENT et 89% de réussite sur FORT.

1.4.2. LES AUTRES TRACES DE MORPHOGRAPHIE OBSERVÉES DANS LES DICTÉES

Pour rendre compte du comportement des enquêtés « bas » par rapport à la morphographie, nous avons observé (section 2.4.2 du chapitre 5) les traces de morphographie relevées sur les autres items de la dictée d'IVQ :

Pour rappel, les analyses portent sur :

- des mots dictés sans déterminant qui appellent implicitement une marque de nombre : (DES) TOMATES – (DES) CERISES ;
- des mots qui en revanche ne demandent pas de pluriel parce qu'ils font référence au nom d'un rayon : (RAYON) PHARMACIE, (RAYON) ÉPICERIE, (RAYON) LIBRAIRIE ou à une quantité : (DE L') ANTI-RHUME, (DE L') ALCOOL, (DE LA) CONFITURE, (DU) SEL. Dans

¹⁷⁰ Se référer chapitre 5, section 2.4.1 pour une présentation des compétences en morphographie des enquêtés « bas ».

cette analyse, nous observons également les marques de nombre et de genre ajoutées de façon erronée dans l'énoncé LE FIANCÉ SOLENNEL.

LE COMPORTEMENT DES SCRIPTEURS SUR LES NOMS DE RAYON ET SUR LES ITEMS QUI FONT RÉFÉRENCE À UNE QUANTITÉ

Dans le corpus « bas », les noms de rayon PHARMACIE, ÉPICERIE et LIBRAIRIE ont été très peu mis au pluriel (seulement 7 graphies, avec 3 occurrences relevées sur ÉPICERIE et LIBRAIRIE). Chez les enquêtés « haut », aucun nom de rayon ne comporte de marque de pluriel.

Concernant l'ajout d'une marque de nombre sur les items qui font référence à une quantité (ANTI-RHUME, ALCOOL, CONFITURE, SEL), très peu d'items au pluriel ont été relevés chez les enquêtés « bas » : 4 marques ont été relevées sur SEL et 28 sur CONFITURE. Chez les enquêtés « haut », aucune marque de nombre n'a été ajoutée sur SEL, ANTI et une seule marque est observée sur RHUME (*anti-rhumes*). En revanche, 14 graphies *confitures* ont été relevées. Ce sont ainsi 16% des enquêtés « haut » qui ont choisi de marquer le nombre sur CONFITURE (contre 6% des enquêtés « bas »). Sur cet item, les enquêtés « haut » ont plutôt tendance à surinvestir la possibilité de marquer le nombre.

Pour les items sur lesquels une marque de nombre est souhaitable, TOMATES et surtout CERISES (nous n'achetons pas une seule cerise), nous avons observé la présence d'une marque de nombre dans 34% des graphies des enquêtés « bas », sur chacun de ces deux mots. Chez les enquêtés « haut », ce phénomène est observé de manière encore plus importante. Sur ces deux mots, une majorité a privilégié une graphie avec une marque de nombre :

- 55% de graphies *tomates* vs 45% de graphies *tomate* ;
- 56% de graphies *cerises* (46 occ.) et de graphies *cerices* (1 occ.) contre 44% de graphies sans marque de nombre.

Néanmoins, même si la majorité des enquêtés « haut » ont eu conscience des marques implicites de nombre « attendues », il apparaît que l'absence d'un signal clair représenté par un déterminant pour indiquer le nombre représente chez les enquêtés « bas » comme chez les enquêtés « haut », une difficulté. Chez les enquêtés « haut » nous observons en effet 56% de graphies au pluriel sur CERISES contre 93% sur FROMAGES.

Par ailleurs, le comportement pluriel des enquêtés vis-à-vis de la présence ou de l'absence de marques de nombre que nous avons observé chez les enquêtés « bas » est également observé chez les enquêtés « haut ». Les enquêtés « bas » qui ont réussi les deux marques attendues n'ont pas forcément marqué les pluriels implicitement attendus et ceux qui n'ont réussi aucune des deux marques attendues ont parfois marqué les pluriels implicites. Chez les

enquêtés « haut », ceux qui ont marqué les pluriels implicites ont tous réussi les deux marques de nombre attendues de la partie IVQ. En revanche, leur score en morphographie sur l'intégralité de la dictée (calculé en ajoutant le « rayon bricolage » qui contient davantage de difficultés dans ce domaine) est compris entre 8 et 14 sur 14. Ainsi, chez certains enquêtés « haut », le fait d'avoir conscience de devoir marquer le nombre même sur les items sur lesquels il n'était pas explicitement demandé de le faire ne signifie pas nécessairement qu'ils aient de bonnes compétences en morphographie (que l'on observe à partir des marques attendues). Ces enquêtés ont néanmoins tous obtenu la moyenne dans ce domaine (sur la partie IVQ et le rayon bricolage).

1.4.3. ERREURS MORPHOGRAPHIQUES RELEVÉES DANS « LE FIANCÉ SOLENNEL »

Chez les enquêtés « bas », nous avons également observé les erreurs morphographiques relevées sur l'énoncé LE FIANCÉ SOLENNEL (voir section 2.4.4, chapitre 5).

Dans le corpus « bas », parmi toutes les propositions de réponses observées sur l'item FIANCÉ, une seule graphie porte la marque nominale du pluriel (*fiancés*) mais 45 graphies présentent un « -e » final. Chez les enquêtés « haut » seuls 3 enquêtés ont ajouté de manière erronée, un « -e » final sur cet item.

Sur l'item SOLENNEL, si nous n'avons observé, chez les enquêtés « bas », qu'une seule marque de nombre (*solanelles*), 111 graphies (26% des réponses) présentent un « -e » final. Chez les enquêtés « haut », ce sont seulement 6 graphies (soit 7% des réponses) qui présentent un « -e » final : *solennelle* (2 occ.), *solanelle* (2), *solannelle* (2).

Ainsi chez les enquêtés « haut », nous n'observons que peu de scripteurs qui ajoutent une marque de nombre ou de genre de manière erronée (sur les noms de rayon, les items qui indiquent une quantité et ceux qui sont dictés avec un déterminant au singulier). Enfin, la majorité d'entre eux a conscience de devoir marquer le nombre sur certains items dictés sans déterminant (TOMATES et CERISES).

1.5. PRÉSENTATION DE LA VARIATION SUR LES MOTS ET LES PSEUDO-MOTS CHEZ LES ENQUÊTÉS « BAS » et « HAUT »¹⁷¹

Le corpus de dictées recueilli chez les enquêtés « bas » et « haut » nous a permis d'établir des scores de réussite dans les différentes composantes de l'écrit : score orthographique avec une distinction faite sur l'orthographe lexicale et sur la morphographie, score de réussite phonographique établi sur les pseudo-mots et sur les mots. Mais au-delà de ces différents scores, nous analysons plus en détail les réponses non conformes à la norme orthographique des scripteurs. Dans un premier temps, nous avons relevé pour chaque population, le nombre de variantes produites pour chaque mot et pour chaque pseudo-mot. Par cette analyse, nous essayons de voir s'il existe des différences dans les choix graphiques, entre les enquêtés « bas » et les enquêtés « haut » lorsque les scripteurs ne connaissent pas l'orthographe d'un mot ou ne parviennent pas à restituer par écrit, la forme phonologique d'un pseudo-mot.

Pour les mots, le nombre de variantes représente le nombre de formes graphiques différentes proposées qui ne respectent pas la norme orthographique : par exemple les graphies *épiceri*, *ipisiri* représentent deux variantes de ÉPICERIE. Nous avons neutralisé les marques de morphologie, en considérant comme une seule variante les formes qui ne diffèrent que d'une marque de nombre (ex. *épiceri*, *épiceris* = une variante de *épiceri*) ou d'une marque de genre (ex. *fiancé*, *fiancée* = une variante de *fiancé*).

Pour les pseudo-mots, le nombre de variantes représente le nombre de formes proposées qui respectent la forme phonographique attendue : par exemple les graphies *micatol*, *micatole* représentent deux variantes de /mikatɔl/.

Nous indiquons pour chaque mot et pseudo-mot, le nombre de réponses non conformes aux attentes orthographiques pour les mots et aux attentes phonographiques pour les pseudo-mots. Cette information permet d'interpréter les résultats et de se rendre compte de la récurrence des formes proposées. En effet, l'indication du nombre de variantes (sans l'indication du nombre de réponses erronées) ne permet pas de rendre compte de la récurrence des réponses. Par exemple, les 9 enquêtés « bas » qui ne connaissaient pas l'orthographe de SEL ont proposé 8 variantes de ce mot (*cel*, *sell*, *sl*, etc.). Les propositions graphiques pour SEL ne sont donc pas récurrentes. Inversement, pour PHARMACIE, 19

¹⁷¹ Cette analyse est issue de l'article rédigé pour la Délégation Générale à la langue Française et aux Langues de Française (Ministère de la culture), dont la parution est prévue fin 2019.

variantes ont été produites par 25 scripteurs, les variantes sont donc davantage récurrentes que celles relevées pour SEL, avec probablement une forme qui a été privilégiée.

Pour pouvoir comparer la variation dans les deux corpus, le nombre de réponses erronées est ramené à un pourcentage calculé en fonction du nombre total de réponses.

1.5.1. VARIATION SUR LES MOTS CHEZ LES ENQUÊTÉS « BAS » et « HAUT »

Nous présentons dans le tableau 26, la variation observée sur chaque mot de la dictée d'IVQ en distinguant les réponses des enquêtés « bas » et « haut ». Les items sont présentés selon leur ordre d'apparition dans la dictée.

mots	variantes enquêtés « bas »	réponses erronées enquêtés « bas »	variantes enquêtés « haut »	réponses erronées enquêtés « haut »
PHARMACIE	19	25 (24%)	1	1 (1%)
ANTI	19	45 (43%)	1	2 (2%)
RHUME	15	50 (48%)	4	13 (15%)
ALCOOL	22	38 (36%)	2	4 (5%)
ÉPICERIE	29	35 (35%)	2	2 (2%)
TOMATES	8	10 (9%)	0	0
PAYS	10	27 (29%)	2	4 (5%)
CONFITURE	10	10 (9%)	0	0
CERISES	20	33 (32%)	4	4 (5%)
SEL	8	9 (8%)	0	0
FROMAGE(S)	9	12 (11%)	0	0
SENTE(NT)	13	22 (22%)	1	1 (1%)
FORT	9	20 (19%)	1	1 (1%)
LIBRAIRIE	27	44 (43%)	2	6 (7%)
FIANCÉ	29	48 (46%)	8	8 (9%)
SOLENNEL	37	87 (82%)	10	42 (48%)

Tableau 26. Relevé de la variation présente sur les mots, relevé d'après les réponses phonographiquement acceptables, approximantes et phonographiquement non acceptables chez les 441 enquêtés « bas » et les 87 enquêtés « haut ».

L'analyse de la variation graphique montre que lorsque les enquêtés « bas » ne connaissent pas l'orthographe d'un mot, la variation peut être très importante, notamment :

- sur les mots de la dictée qui sont les moins fréquents (ex. SOLENNEL) ;
- sur les mots qui sont les moins réguliers (ex. RHUME, FIANCÉ).

L'analyse des réponses non conformes à la norme orthographique¹⁷² montre que ces dernières peuvent porter sur toutes les correspondances phonographiques, ce qui multiplie les possibilités de graphies (ex. *fillancé, fioncé, fiencé, fiensé* pour FIANCÉ). À l'inverse, même lorsque les erreurs sont importantes chez les enquêtés « haut », la variation reste peu importante. Par exemple, sur SOLENNEL, nous n'observons que 10 variantes chez les enquêtés « haut » (alors que le taux de graphies non conformes à la norme approche les 50%). Chez les enquêtés « bas », nous relevons 37 variantes, même si la comparaison est à nuancer vu la différence du nombre de sujets dans chaque échantillon.

Néanmoins, chez les enquêtés « haut », les erreurs sont principalement orthographiques et portent sur une difficulté précise, ce qui limite la variation. Sur LIBRAIRIE par exemple, c'est principalement la transcription du phonème /e/ qui pose problème (*librerie, librerie*) aux enquêtés « haut » (le « -e » muet final est systématiquement noté). Chez les enquêtés « bas »¹⁷³ en revanche, nous relevons en plus de l'omission du « -e » muet final du suffixe « ie » (*librairi, libreri*), des confusions sur le phonème /b/ (*liprerie, livreri*), des omissions de phonèmes (*libairie, librere, etc.*) ainsi que des formes qui cumulent les types d'erreurs (*liphagi, mipari, epréi, etc.*).

1.5.2. VARIATION SUR LES PSEUDO-MOTS CHEZ LES ENQUÊTÉS « BAS » et « HAUT »

Nous observons ensuite la variation graphique observée sur les pseudo-mots chez les enquêtés « bas » et « haut ».

Pour les pseudo-mots, il était attendu du scripteur qu'il transcrive la forme sonore entendue en choisissant parmi les correspondances phonie-graphie du français. Certains phonèmes acceptaient donc plusieurs graphèmes (ex. /i/ : « i », « y ») alors que d'autres ne pouvaient être transcrits que d'une seule manière (ex. /m/ à l'initiale : « m »).

Nous présentons dans un premier temps, la variation observée parmi les formes graphiques que nous avons considérées comme acceptables puis nous observons les propositions graphiques les plus récurrentes chez les enquêtés « bas » et les enquêtés « haut », afin de voir si les deux populations ont fait les mêmes choix graphiques.

¹⁷² Se référer au chapitre 5, section 2.3.2 pour une analyse détaillée des réponses orthographiquement non conformes relevées sur les mots de la dictée, chez les enquêtés « bas ».

¹⁷³ Voir l'analyse des graphies orthographiquement non conformes relevées sur l'item LIBRAIRIE, chapitre 5, section 2.3.2.2.

1.5.2.1 VARIATION SUR LES PROPOSITIONS GRAPHIQUES ACCEPTABLES

Dans le tableau 27, nous présentons la variation sur les formes considérées comme phonographiquement acceptables pour les pseudo-mots¹⁷⁴.

pseudo-mots	variantes acceptées enquêtés « bas »	réponses acceptées enquêtés « bas »	variantes acceptées enquêtés « haut »	réponses acceptées enquêtés « haut »
MICATOL	16	87 (82%)	8	81 (94%)
SIRAPE	13	90 (86%)	9	82 (95%)
DUX	5	71 (68%)	4	76 (92%)
GOBAR	9	79 (75%)	11	78 (90%)
JOC	15	79 (73%)	10	78 (90%)
PÉVANOR	12	64 (60%)	9	77 (89%)

Tableau 27. Relevé de la variation sur les pseudo-mots dont la forme est considérée comme phonographiquement acceptable chez les 441 enquêtés « bas » et les 87 enquêtés « haut ».

Pour les pseudo-mots, les possibilités de variantes sont importantes (choix différents de graphèmes, redoublement de consonnes, ajout de lettres muettes, segmentation). Excepté pour GOBAR, on trouve davantage de formes graphiques différentes chez les enquêtés « bas ». MICATOL par exemple a suscité deux fois plus de propositions chez les enquêtés « bas » (16 graphies différentes contre 8 chez les enquêtés « haut », même si ce résultat est à nuancer au vue de la taille de l'échantillon). Cela étant, on relève peu d'écarts dans le nombre de réponses acceptées sur le plan phonographique dans les deux populations et les taux de réussite sur les pseudo-mots chez les enquêtés « bas » sont plus proches de ceux des enquêtés « haut » que pour les mots de la dictée d'IVQ¹⁷⁵. Nous avons montré (Blondel, Brissaud, Rinck, 2016) que la prise en compte du critère phonographique sur les mots de la dictée (ex. la graphie *épissérie* acceptée du point de vue phonographique) faisait augmenter de manière nette les taux de réussite, notamment sur les mots les moins bien restitués (car peu fréquents et moins réguliers). Par exemple, le taux de réussite sur SOLENNEL passe ainsi de 19% (avec la prise en compte du critère orthographique) à 65% (quand on ajoute la prise en compte du critère phonographique).

¹⁷⁴ Se référer au tableau 8 du chapitre 4, section 3.2 pour la liste des formes retenues comme phonographiquement acceptables dans le corpus « bas ».

¹⁷⁵ Pour une comparaison des scores de réussite entre les enquêtés « bas » et « haut », se référer à la figure 19, chapitre 6 section 1.2 pour les mots et à la figure 20, chapitre 6 section 1.3.1 pour les pseudo-mots

En revanche, la polygraphie observée lorsque le principe phonographique est respecté, permet de rendre compte des représentations des scripteurs en matière d'orthographe, lorsqu'ils ont le choix : application d'une morphographie minimale et d'une orthographe épurée pour que la graphie remplisse une fonction de communication ou prise en compte de régularités du français (ex. ajout d'un « -e » final muet ou d'un morphogramme lexical, marque de famille ou d'affixe).

1.5.2.2 RÉCURRENCE DANS LES PROPOSITIONS GRAPHIQUES

Nous présentons dans le tableau 28, les formes graphiques des pseudo-mots les plus récurrentes chez les enquêtés « bas » et les enquêtés « haut ».

Dans ce tableau, nous recensons les deux ou trois formes graphiques les plus fréquentes, les autres n'ayant été utilisées que par très peu de scripteurs au sein de chaque population.

Pour le calcul du nombre d'occurrences, nous avons neutralisé les accents pour ne compter qu'une seule variante (*pévanor*, *pevanor*, *PÉVANOR*, *PEVANOR* = *pevanor*) et les pourcentages indiqués entre parenthèses ont été calculés selon le nombre de réponses correctes.

	enquêtés « bas »			enquêtés « haut »		
MICATOL	micatol (52%)	micatole (13%)	mikatol (11%)	micatol (57%)	micatole (14%)	mycatol (11%)
SIRAPE	sirap (38%)	sirape (26%)	cirape (12%)	sirap (50%)	sirape (21%)	cirap (10%)
DUX	dux (62%)	duxe (34%)	<i>pas d'autre forme récurrente</i>	dux (84%)	duxe (13%)	<i>pas d'autre forme récurrente</i>
GOBAR	gobar (38%)	gobard (29%)	gobare (14%)	gobard (53%)	gobart (19%)	gobar (13%)
JOC	jock (25%)	joc (22%)	jok (19%)	jock (33%)	joc (27%)	jok (13%)
PÉVANOR	pevanor (55%)	pevanord (23%)	pevanore (9%)	pevanor (52%)	pevanord (29%)	pevanore (12%)

Tableau 28. Relevé des formes les plus récurrentes des pseudo-mots parmi les variantes phonographiquement acceptables chez les 441 enquêtés « bas » et les 87 enquêtés « haut ».

Au sein de chaque population, les enquêtés ont proposé des formes graphiques similaires. Sauf pour GOBAR, les deux formes graphiques les plus utilisées pour chaque pseudo-mot sont les mêmes dans les deux corpus. Ainsi, lorsqu'ils doivent écrire des items qui n'ont pas d'orthographe, les deux populations ont fait les mêmes choix en utilisant des combinaisons de graphèmes similaires, qui sont sans doute les plus caractéristiques du lexique du français.

On remarque en effet que pour MICATOL, SIRAP, DUX et PÉVANOR, les enquêtés ont en majorité opté pour une graphie « simple », minimale, sans ajout de lettres muettes, en utilisant des suites de graphèmes relativement stables et fréquentes (Catach, 1980).

Pour JOC, on peut penser que les enquêtés ont choisi la graphie *jock* (33% des enquêtés « bas » et 25% des enquêtés « haut ») pour se rapprocher d'un prénom existant, JACK.

Pour GOBAR, on observe néanmoins une différence dans le choix ou non d'ajouter une lettre finale muette. Les enquêtés « haut » ont préféré l'écrire avec une lettre finale muette « -d » (53%) ou « -t » (19%), contre 29% de formes en « -d » pour les enquêtés « bas ». Le « -d » a également été choisi en priorité comme lettre finale par les deux populations avec la graphie *pevanord*.

Pour SIRAPE en revanche, l'ajout d'un « -e » muet final est davantage privilégié par les enquêtés « bas » (38% contre 21% pour les enquêtés « haut »).

1.5.2.3 LA GRAPHIE DES PSEUDO-MOTS POUR OBSERVER LES COMPORTEMENTS DE SCRIPTEURS

Dans l'analyse suivante, nous utilisons les pseudo-mots pour observer les comportements à l'écrit des scripteurs vis-à-vis de la graphie des lettres finales (choix d'ajouter ou non une finale muette et choix de la finale muette). Nous avons en effet recensé de nombreuses variantes dans la graphie des pseudo-mots, tant chez les enquêtés « bas » que chez les enquêtés « haut ».

Pour cette analyse, nous nous intéressons à 4 pseudo-mots :

- MICATOL et SIRAPE : avec ou sans « -e » muet final ;
- GOBAR et PÉVANOR avec une finale muette (« -e », « -d », « -t », « -s ») ou sans finale muette.

Pour tenter de mettre en évidence des comportements de scripteurs et pour déterminer s'il existe un lien entre l'ajout ou non d'une finale muette (et sur le choix de cette finale) avec les compétences à l'écrit des scripteurs, nous avons défini 4 groupes de scripteurs, à partir de leur résultat à la dictée.

L'analyse des résultats chez les enquêtés des deux populations a en effet montré que bien qu'ils aient été sélectionnés pour leurs compétences à l'écrit, ils ne forment pas un groupe de scripteurs homogène que l'on pourrait catégoriser de la manière suivante : enquêtés « bas » = enquêtés en difficulté avec l'écrit vs enquêtés « haut » = enquêtés sans difficulté avec l'écrit.

Cette hétérogénéité nous incite à proposer une nouvelle catégorisation, différente de celle proposée par l'INSEE, qui n'est plus uniquement basée sur leur score suite au module « d'orientation » (« bas » vs « haut »). Cette nouvelle catégorisation prend en compte leurs compétences à l'écrit dans le domaine de la morphographie (compétence observée dans la dictée), les compétences en morphographie se révélant en effet plutôt discriminantes dans les deux corpus.

La prise en compte de ce critère nous amène à diviser les enquêtés en quatre catégories (ou groupes) :

- les enquêtés « haut » sont divisés en deux catégories, selon leur score de réussite en morphographie mesuré sur l'intégralité de la dictée (morphographie de la dictée d'IVQ et du rayon bricolage) :
 - groupe 1 : ceux qui ont obtenu 100% de réussite dans ce domaine ;
 - groupe 2 : ceux qui n'ont pas obtenu 100% de réussite ;
- les enquêtés « bas » sont divisés en trois catégories, selon leur score de réussite en morphographie mesuré sur les deux marques de nombre attendues et selon leur score orthographique (score global) :
 - groupe 2 : les enquêtés « bas » qui ont réussi les deux marques de nombre de la dictée d'IVQ ;
 - groupe 3 : les enquêtés « bas » qui n'ont pas réussi les deux marques de nombre et qui ont obtenu un score global compris entre 13 et 22 ;
 - groupe 4 : les enquêtés « bas » qui n'ont pas réussi les deux marques de nombre et qui ont obtenu un score global inférieur à 13.

Nous utilisons le score de 12 comme frontière pour départager les enquêtés des catégories 3 et 4, parce qu'aucun enquêté « haut » n'a obtenu de score $<$ à 13 sur 24 sur la dictée d'IVQ.

La catégorie 1 est réservée aux enquêtés « haut » car ne pouvons pas fusionner le résultat obtenu sur les deux dictées : obtenir 100% de réussite en morphographie sur la dictée complémentaire demande en effet davantage de compétences qu'obtenir 100% de réussite sur la dictée d'IVQ, la première contenant davantage de marques à restituer et de difficultés auxquelles les enquêtés « bas » n'ont pas été confrontés.

Nous récapitulons les groupes de scripteurs :

- groupe 1 : enquêtés « haut » qui ont obtenu 100% de réussite en morphographie (dictée d'IVQ et rayon bricolage) ;

- groupe 2 : enquêtés « haut » qui n'ont pas obtenu 100% de réussite en morphographie (dictée d'IVQ et rayon bricolage) et enquêtés « bas » qui ont réussi les deux marques de nombre de la dictée d'IVQ ;
- groupe 3 : enquêtés « bas » qui ont obtenu un score orthographique compris entre 13 et 22 ;
- groupe 4 : enquêtés « bas » qui ont obtenu un score inférieur à 13.

Dans les analyses suivantes, nous cherchons à voir si le comportement des quatre groupes de scripteurs face aux finales constitue un marqueur de réussite et s'il montrerait une sensibilité à une régularité du français. Pour cela, nous comparons les finales choisies avec la fréquence des finales sur les mots du français.

PROPORTION DE FINALES MUETTES SUR MICATOL ET SIRAPE POUR CHAQUE GROUPE DE SCRIPTEURS

Pour effectuer cette analyse sur la présence ou non d'un « -e » final muet sur les pseudo-mots MICATOL et SIRAPE, nous ne sélectionnons que les graphies pour qui les deux derniers phonèmes sont identiques aux deux derniers phonèmes sur les deux pseudo-mots : /ɔl/ ou /ol/ (nous neutralisons la prononciation du « o ») pour MICATOL et /ap/ pour SIRAP.

Sur MICATOL :

- corpus « bas » : 112 graphies sur les 389 réponses qui se terminent par /ɔl/ ou /ol/ ont un « -e » final muet, ce qui représente 29% des graphies qui se terminent par /ɔl/ ou /ol/ relevées chez les enquêtés « bas » ;
- corpus « haut » : 14 graphies sur les 89 réponses obtenues sur l'item ont un « -e » final muet, ce qui représente 16% des graphies relevées chez les enquêtés « haut ».

Sur SIRAPE :

- corpus « bas » : 209 graphies sur les 370 réponses qui se terminent par /ap/ ont un « -e » final muet, ce qui représente 56% des graphies qui se terminent par /ap/ relevées chez les enquêtés « bas » ;
- corpus « haut » : 26 graphies sur les 85 réponses qui se terminent par /ap/ ont un « -e » final muet, ce qui représente 31% des graphies relevées chez les enquêtés « haut ».

Dans le tableau 29, nous classons le comportement des scripteurs concernant l'ajout ou non d'un « -e » final en distinguant les 4 groupes de scripteurs.

	MICATOL		SIRAPE	
	« -ole » : « -e » final muet	« -ol » : pas de finale muette	« -ape » : « -e » final muet	« -ap » : pas de finale muette
groupe 1	12%	88%	31%	69%
groupe 2	9%	91%	40%	60%
groupe 3	33%	67%	59%	41%
groupe 4	56%	44%	74%	26%

Tableau 29. Proportion de « -e » final muet sur MICATOL et SIRAPE, calculée pour chacun des 4 groupes de scripteurs.

Pour MICATOL, il apparaît très clairement que les deux groupes les plus compétents privilégient une graphie minimale de ce pseudo-mot, sans ajout de lettre muette. Par contraste, le groupe le moins compétent ajoute un « -e » final plus d'une fois sur deux. Le groupe 3 se situe dans une position intermédiaire, qui penche néanmoins plutôt vers le choix d'une graphie minimale.

Pour SIRAPE également, l'ajout d'un « -e » final caractérise les scripteurs les moins compétents (groupes 3 et 4) et ce phénomène est particulièrement marqué pour le groupe 4 avec 74% des graphies qui comportent une lettre finale.

Pour interpréter ces résultats, nous observons la fréquence des finales¹⁷⁶ sur les mots se terminant par « -ap » (pour SIRAPE) et par « -ol » (pour MICATOL).

Il apparaît que pour les mots en « -ap », la tendance est plutôt à la prononciation de la consonne finale : *rap, cap, gap* vs *drap, sparadrap*. Le « -e » final n'est donc généralement pas obligatoire pour que la consonne finale soit prononcée et il semble ainsi les enquêtés les plus compétents aient davantage enregistré ce phénomène.

Pour MICATOL, l'observation de la fréquence des finales sur les mots se terminant par « l » montre en revanche qu'il existe davantage de mots se terminant par « -ole » (182 mots) que par « -ol » (141 mots). Cependant, la présence ou non d'un « -e » final n'influe jamais sur la prononciation du « l » (à l'inverse du « -p » en finale).

Ainsi, tout se passe comme si les enquêtés les plus compétents privilégiaient une graphie minimale, sans finale muette. Dans cette logique, il semble également qu'ils aient davantage enregistré que le « -e » muet n'est pas obligatoire pour que le « -p » en finale se prononce.

¹⁷⁶ Les fréquences des finales sur les mots ont été calculées avec la version électronique du *Petit Robert*, 2011.

PROPORTION DE FINALES MUETTES SUR GOBAR ET PÉVANOR POUR CHAQUE GROUPE DE SCRIPTEURS

Dans l'analyse suivante, nous observons le comportement des scripteurs sur la graphie de la finale des pseudo-mots GOBAR et PÉVANOR. Sur ces deux pseudo-mots, nous regardons si les groupes de scripteurs ont opté ou non pour une finale muette et nous détaillons la lettre qui a été privilégiée (« -d », « -e », « -t », « -s »). Comme dans l'analyse précédente, nous ne sélectionnons que les graphies pour lesquelles les deux derniers phonèmes sont identiques aux deux derniers phonèmes sur les deux pseudo-mots : /aR/ pour GOBAR et /ɔR/ ou /oR/ (nous neutralisons la prononciation du « o ») pour PÉVANOR.

Sur GOBAR :

- corpus « bas » : 214 graphies parmi les 401 réponses qui se terminent par /aR/ ont une lettre finale muette, ce qui représente 53% des graphies qui se terminent par /aR/ relevées chez les enquêtés « bas » ;
- corpus « haut » : 16 graphies sur les 87 réponses obtenues sur l'item ont un « -e » final, ce qui représente 18% des graphies relevées chez les enquêtés « haut ».

Sur PÉVANOR :

- corpus « bas » : 179 graphies sur les 366 réponses qui se terminent par /ɔR/ ou /oR/ ont une lettre finale muette, ce qui représente 49% des graphies qui se terminent par /ɔR/ ou /oR/ relevées chez les enquêtés « bas » ;
- corpus « haut » : 46 graphies sur les 85 réponses qui se terminent par /ɔR/ ou /oR/ ont une lettre finale muette, ce qui représente 54% des graphies relevées chez les enquêtés « haut ».

Dans le tableau 30, nous observons le comportement de chacun des 4 groupes de scripteurs sur la graphie des finales des pseudo-mots GOBAR et PÉVANOR. Sur ces deux pseudo-mots, nous regardons si les groupes de scripteurs ont opté ou non pour une finale muette et nous détaillons la lettre qui a été privilégiée (« -d », « -e », « -t », « -s »).

	GOBAR				PÉVANOR			
	« -rd »	« -re »	« -rt »	« -r »	« -rd »	« -re »	« -rt »	« -r »
groupe 1	58%	3%	21%	18%	39%	13%	3%	45%
groupe 2	40%	4%	12%	44%	19%	21%	6%	55%
groupe 3	34%	15%	11%	41%	16%	24%	6%	53%
groupe 4	6%	34%	6%	54%	10%	36%	12%	41%

Tableau 30. Proportion de finale muette sur GOBAR et PÉVANOR, calculée pour chacun des 4 groupes de scripteurs.

À l'inverse de ce l'on a pu observer sur les pseudo-mots précédents, il apparaît qu'en ce qui concerne le pseudo-mot GOBAR, plus les enquêtés sont compétents et plus ils privilégient une finale muette. Le choix de la finale se porte alors majoritairement sur un « -d », voire un « -t », le « -e » étant très peu représenté. Le groupe le moins compétent en revanche, privilégie une finale minimale dans 54% des cas et à l'inverse du groupe de scripteurs le plus compétent, ceux qui choisissent d'ajouter une finale muette privilégient le « -e » (comme dans les pseudo-mots précédents : MICATOL et SIRAPE).

Pour PÉVANOR, les différences entre les groupes sont moins nettes : entre 45% et 55% des graphies ne présentent pas de lettre finale muette ; le choix d'ajouter ou non une finale étant donc plus homogène (environ 1 cas sur 2). En revanche, lorsque les scripteurs ont opté pour une finale muette, la tendance est la même que pour GOBAR : il s'agit très majoritairement du « -d » pour le groupe 1 et plutôt du « -e » pour le groupe 4. Par ailleurs, le choix de la finale chez les deux groupes les plus compétents correspond à la fréquence des finales observée sur les mots se terminant par /ar/ en français (*Petit Robert électronique*, 2011) : « -arre » (21 mots), « -ard » (288 mots), « -ar » (117 mots), « -are » (79 mots) et enfin « -art » (51 mots).

1.5.3. BILAN : APPORT DE L'ENQUÊTE COMPLÉMENTAIRE POUR MIEUX CERNER LES COMPÉTENCES ORTHOGRAPHIQUES DE SCRIPTEURS EN DIFFICULTÉ

Nous observons donc les mêmes fragilités face à l'orthographe parmi les enquêtés des deux populations, même si les difficultés ressortent de manière plus forte chez les enquêtés « bas », qui semblent davantage impactés par l'effet de consistance et de fréquence des mots.

Chez les enquêtés « haut », les écarts à la norme orthographique portent généralement sur les mêmes difficultés : la transcription de /a/ avec les graphèmes « en » dans SOLENNEL, le graphème « ai » de LIBRAIRIE, la structure complexe « ays » de PAYS. La grande majorité des enquêtés « haut » maîtrisent donc les règles de fonctionnement et de distribution relatives au français et parviennent à proposer une forme graphique plausible lorsqu'ils ne connaissent pas l'orthographe d'un mot : par exemple pour SOLENNEL : *solanel* (18), *solannel* (5), *solanelle*

(2), *solannelle* (2), *sollanel* (1). Les enquêtés « haut » maîtrisent ces règles davantage que les enquêtés « bas » si l'on regarde les réponses recueillies sur cet item avec 84% de réponses conformes à la norme orthographique et au critère phonographique contre 63% chez les enquêtés « bas ». Sur un item dont la forme orthographique est moins complexe, comme *FIANCÉ*, cet écart demeure important avec 18 points d'écart (94% de réponses conformes à la norme orthographique et au critère phonographique chez les enquêtés « haut » contre 76% chez les enquêtés « bas »).

La comparaison des scores de réussite sur les trois seuls items qui permettent de comparer entre les deux populations montre que les difficultés portent sur les mêmes marques : entre les deux marques de nombre c'est l'accord du nom *FROMAGES* qui a été le mieux réussi, même si c'est l'absence de marque sur l'adverbe qui détient les meilleurs taux de réussite chez les enquêtés « bas » comme chez les enquêtés « haut ». En revanche les taux de réussite observés diffèrent beaucoup entre les deux populations, avec jusqu'à 43 points d'écart observé sur la réussite de l'accord en nombre du verbe.

La morphologie flexionnelle du nombre a donc posé problème aux deux populations : ce phénomène s'observe surtout chez les enquêtés « bas », mais il est également présent chez les autres enquêtés, particulièrement lorsqu'ils sont en présence de structures syntaxiques complexes où les marques peuvent passer pour redondantes (ex. *QUATRE PAQUETS DE VIS PLATES, UNE BOITE D'ALLUMETTES*). La maîtrise de cette compétence fonctionne donc comme une ligne de partage, d'une part entre les enquêtés « bas » et les enquêtés « haut », d'autre part au sein des enquêtés « haut » sur le rayon bricolage. Ces résultats montrent que la morphologie (même sans aller chercher des compétences complexes) reste une zone de fragilité en orthographe, même chez les meilleurs scripteurs.

Si le marquage du nombre sur le nom *FROMAGES* et sur le verbe *SENTENT* semble donc être un bon indicateur pour distinguer les personnes qui ont de grandes difficultés en orthographe des autres, ces deux difficultés morphographiques ne suffisent pas, en revanche, pour cerner les compétences en morphologie des personnes qui ont obtenu de bons résultats à la dictée. Pour *FORT*, en revanche, où un taux de réussite élevé a été constaté, c'est l'absence de marque qui correspond à une réussite mais nous ne pouvons déterminer si ceux qui ont réussi ont eu conscience de ne pas devoir marquer le nombre sur un adverbe. Cet exemple pourrait constituer un indicateur de compétence s'il pouvait être comparé avec l'accord en nombre de l'item utilisé comme adjectif (ex. *TROIS FROMAGES QUI NE SENTENT PAS FORT* vs *TROIS FROMAGES FORTS*).

2. COMPÉTENCES ORTHOGRAPHIQUES, PRATIQUES D'ÉCRITURE ET REPRÉSENTATIONS DE L'ÉCRIT DE 87 SCRIPTEURS

Nous observons à présent les compétences orthographiques des 87 enquêtés « haut » sur la partie complémentaire de la dictée. Proposée uniquement aux enquêtés « haut » lors de l'enquête complémentaire de 2014, cette extension de la dictée d'IVQ que nous appellerons désormais le « rayon bricolage » contient des difficultés plus importantes que dans la dictée d'IVQ, avec du lexique et des structures syntaxiques plus complexes.

2.1. RÉSULTATS GÉNÉRAUX DU « RAYON BRICOLAGE »

Nous avons vu que si la dictée d'IVQ avait permis de repérer les personnes les plus en difficulté, les résultats étaient plutôt homogènes chez les enquêtés « haut », avec 88% d'entre eux qui ont obtenu plus de 80% de réussite. Nous regardons maintenant si le taux de réussite reste identique sur la suite de l'exercice ou si certains enquêtés se trouvent en difficulté sur cette partie de l'exercice.

2.1.1. RÉPARTITION DES SCORES

Nous présentons dans la figure 23, la répartition des scores obtenus par les 71¹⁷⁷ enquêtés « haut » qui ont passé l'intégralité de la dictée complémentaire.

Nous rappelons que le barème prend en compte l'orthographe et la morphographie pour un score total de 22 points.

¹⁷⁷ Nous ne disposons que de 71 dictées complètes sur les 87 produites par les enquêtés « haut ». Un des enquêteurs a omis de dicter la dernière phrase du « rayon bricolage » : QUELQUES MÈTRES DE TUYAU D'ARROSAGE à 16 enquêtés.

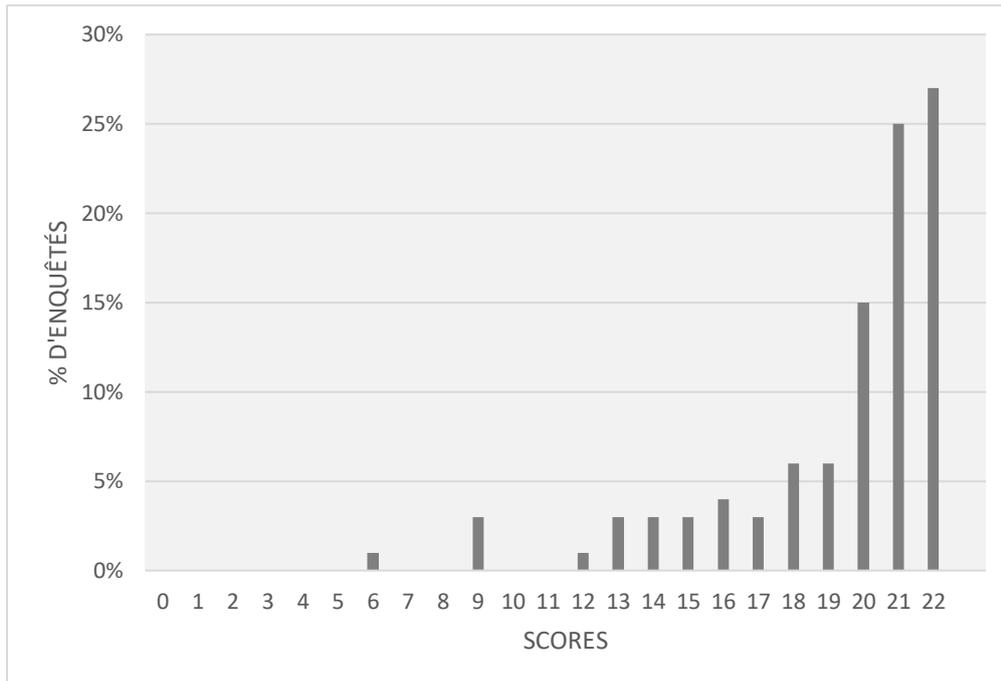


Figure 23. Répartition des scores obtenus par les 71 enquêtés « haut » qui ont passé l'intégralité de la dictée complémentaire (rayon bricolage).

Les écarts qui subsistent entre les enquêtés « haut » sont plus importants sur cette partie de la dictée, avec des scores plus dispersés que sur la partie IVQ, compris entre 6 et 22 malgré une médiane qui reste élevée (21 sur 22). Lorsque l'on compare ce résultat avec celui obtenu sur la dictée d'IVQ, on observe une proportion moins élevée de très bons scores : 79% des enquêtés « haut » ont obtenu plus de 80% de réussite sur cette partie de la dictée (contre 88% pour la partie IVQ). Par ailleurs, une proportion similaire d'enquêtés a obtenu un score maximal sur les deux parties de la dictée (29% pour la première partie et 27% pour le rayon « bricolage »).

Si la proportion d'enquêtés qui n'ont pas rencontré de difficulté particulière avec l'orthographe lexicale et grammaticale des dictées est sensiblement la même, les écarts qui subsistent entre les enquêtés « haut » sont plus importants sur la partie complémentaire de la dictée, avec un résultat moyen plus faible et des difficultés parfois importantes pour quelques enquêtés « haut » (trois scores inférieurs à 10 sur 24) ; difficultés qui n'apparaissent pas sur la dictée d'IVQ.

Pour identifier ce qui a posé problème aux enquêtés « haut » sur cette partie de la dictée, nous observons ensuite le taux de réussite sur chaque mot (score en orthographe lexicale), en isolant la morphographie.

2.1.2. COMPÉTENCES EN ORTHOGRAPHE LEXICALE SUR LES MOTS DU « RAYON BRICOLAGE »

La figure 24 présente les compétences en orthographe lexicale des enquêtés « haut » sur 21 mots du rayon bricolage, classés selon leur pourcentage de réussite orthographique. Les pourcentages de réussite ont été calculés en fonction du nombre de réponses produites¹⁷⁸. Nous rappelons que dans cette analyse, les marques de morphographie ne sont pas prises en considération.

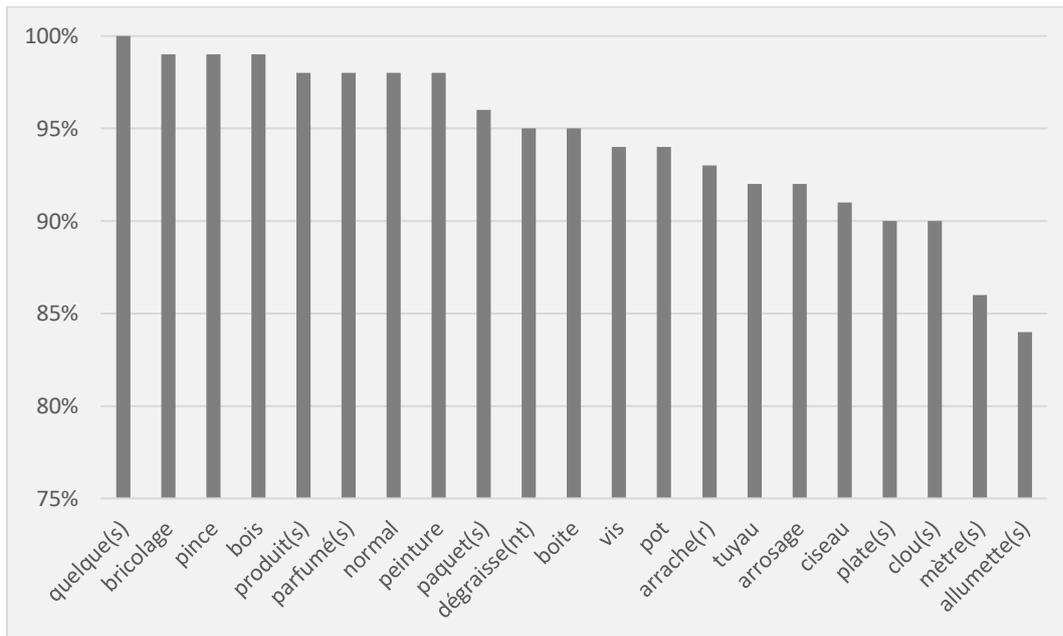


Figure 24. Compétences lexicales des 87 enquêtés « haut » sur 21 mots du rayon bricolage.

Alors que les items de la dictée complémentaire contiennent des difficultés lexicales un peu plus importantes, on observe effectivement une proportion plus importante de mots qui obtiennent moins de 95% de réussite que dans la partie IVQ de la dictée. Néanmoins 8 mots obtiennent entre 98% et 100 de réussite : QUELQUE(S), BRICOLAGE, PINCE, BOIS, PRODUIT(S), PARFUMÉ(S), NORMAL, PEINTURE. Aucun mot n'obtient moins de 84% de réussite alors que dans la partie IVQ, sur l'item SOLENNEL, seules 50% des réponses sont conformes à la norme orthographique.

¹⁷⁸ Nous rappelons que 16 enquêtés sur les 87 n'ont pas produit le dernier syntagme de la dictée (4 items manquants). Ainsi, selon les mots, les taux de réussite sont calculés d'après les résultats de 71 scripteurs ou de 87 scripteurs.

Comme dans la dictée d'IVQ, nous observons un taux de réussite très élevé sur les mots les plus consistants du « rayon bricolage »¹⁷⁹ : PINCE, BRICOLAGE, PRODUIT(S), PARFUMÉ(S) et les plus fréquents : BOIS, PAQUET(s), PEINTURE.

Nous présentons dans le tableau 31, les graphies orthographiquement non conformes relevées sur les quatre mots les moins bien réussis du « rayon bricolage » : CISEAU, PLATE(S), CLOU(S), MÈTRE(S), ALLUMETTE(S).

items	graphies phonographiquement acceptables	graphies approximantes	graphies phonographiquement non acceptables	nombre réponses
CISEAU(X)	<i>cizeau</i> (1) <i>sciseau</i> (1) <i>sciseaux</i> (1) <i>siseau</i> (2) <i>siseaux</i> (1)		<i>sicau</i> (1) <i>sicau</i> (1)	87
PLATE(S)	<i>plâte</i> (1) <i>plâtes</i> (2) <i>platte</i> (3) <i>plattes</i> (1)	<i>plats</i> (1)		87
CLOU(S)	<i>clout</i> (1) <i>clouts</i> (2) <i>cloux</i> (3)		<i>claufs</i> (1)	87
MÈTRE(S)	<i>mêtres</i> (6) <i>mètre</i> (1) <i>mettre</i> (3)			71 ¹⁸⁰
ALLUMETTE(S)	<i>alumette</i> (4) <i>alumettes</i> (8)	<i>allumetts</i> (1)		86

Tableau 31. Relevé des graphies non conformes à la norme orthographique pour les 5 mots les moins bien restitués dans la dictée complémentaire chez les 87 enquêtés « haut ».

¹⁷⁹ Se référer à la section 4.2.2.2, chapitre 3 pour une présentation des indices de consistance calculés dans MANULEX et des indices de fréquence calculés dans LEXIQUE pour chaque mot du rayon bricolage.

¹⁸⁰ Il s'agit des 16 enquêtés qui n'ont pas été interrogés sur cet item.

Nous commentons brièvement les graphies non conformes des cinq mots les moins bien restitués chez les « enquêtés haut » :

- les erreurs relevées sur PLATE(S), MÈTRE(S) et ALLUMETTE(S), concernent principalement le redoublement de consonnes (*platte, mettre*¹⁸¹, *alumette*) et l'ajout d'un accent circonflexe (*mètre, plâte*¹⁸²) ;
- sur CISEAU, les erreurs concernent le choix du graphème pour les phonèmes /s/ (*siseau, sciseau*), /z/ (*cizeau, sicau*) et pour un seul enquêté, le phonème /o/ (*sicau*) ;
- sur CLOU, 3 erreurs concernent l'ajout d'un « -t » final (*clout, clouts*) et 3 erreurs portent sur une marque de nombre erronée (*cloux*).

Les différences de compétences observées sur les scores obtenus par les enquêtés « haut » (de 6 à 22 sur 22) portent davantage sur la morphographie et nous observons dans l'analyse suivante, les compétences des enquêtés dans ce domaine, d'après les difficultés que l'on peut observer sur le « rayon bricolage ».

2.1.3. COMPÉTENCES EN MORPHOGRAPHIE OBSERVÉES SUR L'INTÉGRALITÉ DE LA DICTÉE

2.1.3.1 SCORES EN MORPHOGRAPHIE

La figure 25 présente ensuite les compétences en morphographie des enquêtés « haut » à partir des onze marques de nombre présentes dans l'intégralité de la dictée, avec l'accord en nombre de FROMAGES et SENTENT (partie IVQ), de deux adjectifs et de six noms (« rayon bricolage ») ainsi que l'absence d'accord de l'adverbe FORT (partie IVQ) et de CISEAU et TUYAU (« rayon bricolage »). Les marques de nombre erronées (ex. *cloux* pour CLOUS, 3 occurrences) sont également prises en compte, seule la présence d'un accord nous intéresse.

¹⁸¹ Sur cette graphie, l'erreur peut également provenir d'une confusion entre les formes homophones METTRE et MÈTRE.

¹⁸² L'ajout d'un accent circonflexe rappelle celui qui est présent dans le mot PLÂTRE.

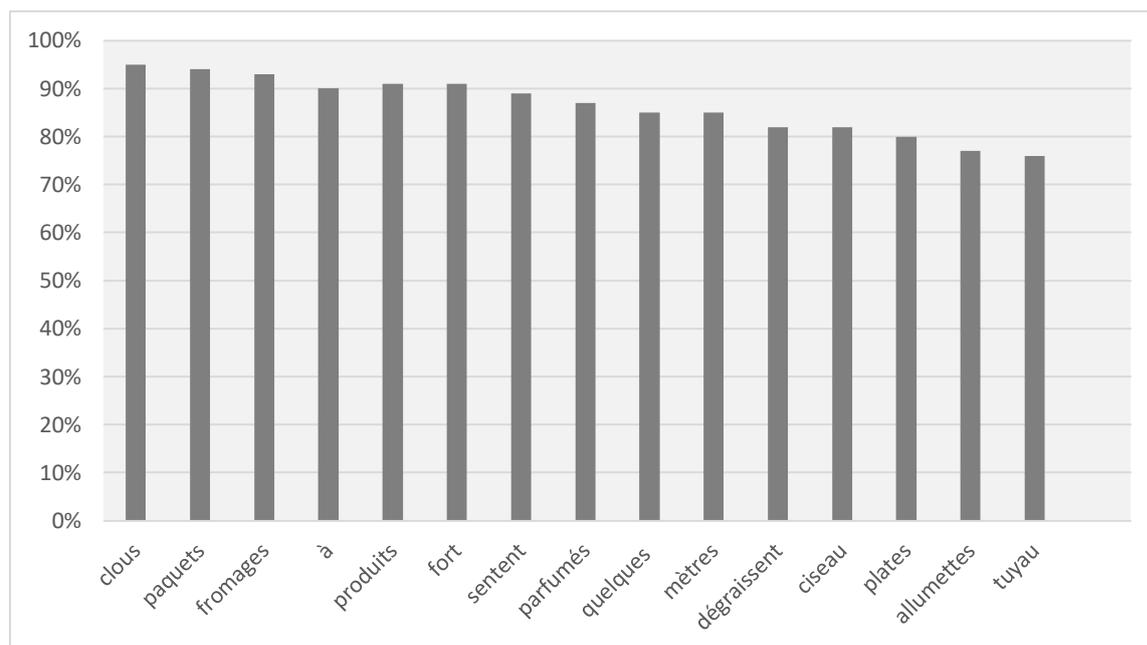


Figure 25. Compétences en morphographie des 87 enquêtés « haut » sur l'intégralité de la dictée.

Nous avons vu chez les enquêtés « bas » que l'accord en nombre du nom FROMAGES et du verbe SENTENT, semblait constituer un bon indicateur pour distinguer les personnes qui présentent d'importantes difficultés en morphographie mais qu'ils ne suffisaient pas, en revanche, pour cerner les compétences dans ce domaine des personnes qui ont de meilleures connaissances orthographiques. 46% des enquêtés « bas » ont réussi l'accord de SENTENT et 59% celui de FROMAGES et nous regrettons de ne pas disposer de davantage de marques d'accord à analyser dans la dictée d'IVQ. La présence de marques supplémentaires aurait été nécessaire pour tester les compétences en morphographie des 33% enquêtés « bas » qui ont obtenu de très bons résultats à la dictée d'IVQ (plus de 80% de réussite).

Ceux-ci ont plutôt bien réussi la morphographie de la dictée, en obtenant 89% de réussite sur FROMAGES, 95% de réussite sur SENTENT (score supérieur à la moyenne des scores obtenus par les enquêtés « haut » sur cet accord).

L'observation du comportement des scripteurs face à la morphologie du « rayon bricolage » nous permet de davantage cerner les compétences des scripteurs dans ce domaine. Chez les enquêtés « haut », on observe en effet une bonne restitution des marques de nombre dans les constructions qui fonctionnent avec un signal clair du pluriel (avec présence d'un déterminant : LES CLOUS, QUATRE PAQUETS, etc.). Mais les résultats baissent lorsqu'il s'agit de tournures complexes où aucun signal ne vient aider au marquage du pluriel : par exemple, le « -s » de ALLUMETTES dans UNE BOITE D'ALLUMETTES a été produit par seulement 77% des enquêtés « haut ». Pour Lucci et Millet (1994), ce type d'omission pourrait s'expliquer par la redondance des marques de pluriel à l'écrit qui pousse les scripteurs à négliger les règles

d'accord, l'idée de pluralité étant déjà exprimée dans UNE BOITE D'. C'est également une difficulté que l'on retrouve en fin de scolarité au collège (Brissaud, 2015). D'après les résultats obtenus par des élèves de troisième dans une épreuve de dictée, c'est surtout le contexte syntaxique qui semble impacter la réussite lorsqu'il s'agit de produire des accords en nombre en fin de scolarité obligatoire, 96 marques de pluriel sur 98 sont présentes sur ARBRES dans DES ARBRES SONT REPLANTÉS contre seulement 60 sur 98 dans LA GAULE ETAIT COUVERTE D'ARBRES. L'absence d'un déterminant qui pourrait fonctionner comme un signal du pluriel devant ARBRES paraît ainsi entraver le marquage du nombre. En revanche, l'absence d'accord pour TUYAU dans QUELQUES METRES DE TUYAU a posé problème à 24% des enquêtés « haut » qui semblent avoir réalisé l'accord entre le déterminant QUELQUES et le nom TUYAU.

Au vu des résultats observés dans l'intégralité de la dictée, il semble que les compétences en morphographie des enquêtés « haut » se distinguent surtout lorsque le marquage du nombre n'est pas facilité par un signal clair. Pour mettre en relief ce phénomène, nous avons établi des profils en morphologie des enquêtés « haut ».

2.1.3.2 PROFIL MORPHOGRAPHIQUE DES ENQUÊTÉS « HAUT »

La figure 26 propose une répartition des enquêtés « haut » selon leur profil morphographique calculé à partir de leur réussite sur les 11 marques de pluriel et les trois marques du singulier. Les enquêtés sont classés en 4 catégories :

- moins de 50% de réussite ;
- entre 50 et 80% de réussite ;
- entre 80 et 100% de réussite ;
- 100% de réussite.

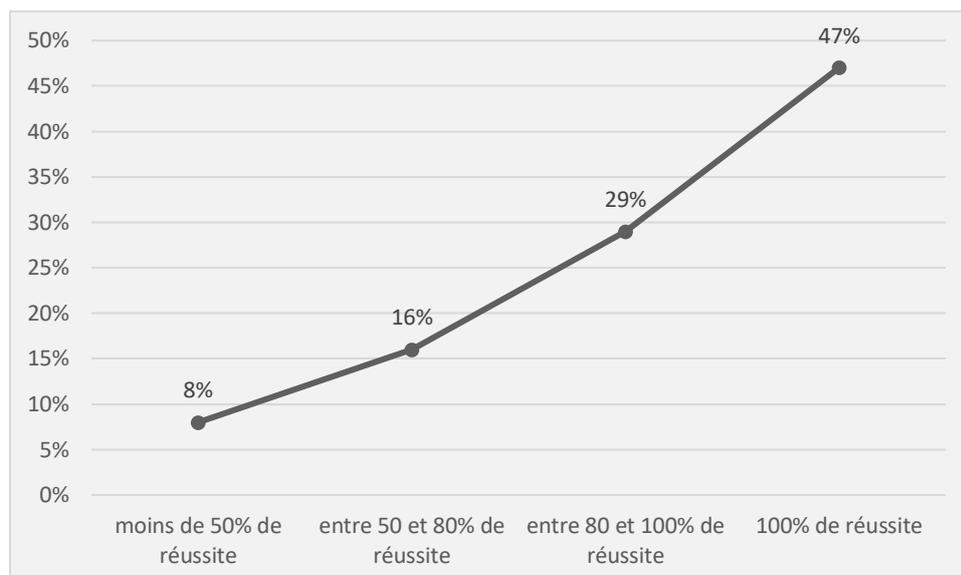


Figure 26. Profil en morphographie des 71 enquêtés « haut ».

Les marques de nombre proposées dans les deux dictées font appel à des connaissances plutôt générales : accord du nom dans des constructions simples avec un marquage du pluriel grâce à la présence d'un déterminant (ex. LES CLOUS, DEUX PRODUITS) ou sans marquage clair (ex. UNE BOITE D'ALLUMETTES), accord de trois verbes au présent de l'indicatif et de trois adjectifs. Les quelques « pièges » pouvaient être déjoués en faisant appel à une réflexion portant sur le sens : QUELQUES (= DES) MÈTRES DE TUYAU, UNE BOITE (QUI CONTIENT) D'(DES) ALLUMETTES. En effet, il n'était pas demandé de connaître et de savoir appliquer des règles complexes comme celle de l'accord du participe passé, etc. Seule l'absence d'accord sur l'adverbe FORT demande une connaissance plus fine des règles de morphographie : il fallait savoir que FORT dans ce contexte était un adverbe et que par conséquent il était invariable. Pour cette raison, nous ne pouvons être certaine que l'absence de marque sur FORT (surtout chez les enquêtés « bas » mais également chez les enquêtés « haut ») corresponde à une réussite.

Nous pourrions nous attendre, dans ce groupe d'enquêtés « haut » à une très bonne réussite sur la morphographie et c'est le cas pour 76% d'entre eux qui obtiennent plus de 80% de réussite sur la morphologie (en ayant réussi plus de 10 marques sur les 14). Cependant, certains enquêtés ont éprouvé des difficultés dans ce domaine. On observe en effet que moins de la moitié des enquêtés « haut » a réussi à restituer les 11 marques de pluriel et les trois absences de pluriel et que la morphologie a posé problème à 24% des enquêtés « haut ». Parmi eux, 8% (soit 7 enquêtés) ont été particulièrement en difficulté, en ayant réussi moins de 7 marques sur les 14.

À titre d'exemple, nous présentons les erreurs relevées chez un scripteur « haut » qui a obtenu un score de 4 sur 14 en morphographie. Nous soulignons les mots qui comportent une erreur en morphographie :

3 fromage qui ne sente pas fort, 2 produit parfumer qui dégraisse bien, 4 paquets de vis plate, un ciseaux, une boîte d'allumette, quelque mètres de tuyaux d'arrosage.

Ici, le scripteur n'a réussi que des accords en nombre sur des noms : PAQUETS, CLOUS, MÈTRES et a respecté l'absence d'accord sur l'adverbe FORT. Les compétences de cet enquêté en morphographie paraissent fragiles : il n'a réussi aucun accord sur les verbes, les adjectifs et le déterminant.

Lorsque l'on observe le reste de sa dictée, il apparaît que la plupart des difficultés en orthographe lexicale a en revanche été surmontée. Les erreurs portent sur les finales muettes de RHUME (*rhum*) et de PAYS (*pay*), sur le graphème « ai » dans LIBRAIRIE (*librerie*) et sur le mot SOLENNEL (*solannelle*), item qui nous le rappelons, a été particulièrement mal restitué dans les deux corpus mais dont la phonographie est néanmoins respectée ici. Nous relevons également des erreurs phonographiques sur trois des six pseudo-mots : *ciraque* pour SIRAP, *cobar* pour GOBAR et *evanord* pour PÉVANORD.

2.2. PRATIQUES D'ÉCRITURE ET REPRÉSENTATIONS DE L'ÉCRIT : UNE HÉTÉROGÉNÉITÉ DE PROFILS ET DE PRATIQUES¹⁸³

Nous présentons dans un premier temps le profil orthographique et les pratiques d'écriture des enquêtés réinterrogés en 2014, à partir non plus des seules productions écrites mais de l'intégralité de l'enquête complémentaire (dictée et entretien). À partir des réponses recueillies aux questions fermées, nous nous intéressons à présent aux pratiques de l'écriture au quotidien, manuscrite ou électronique, à la maison et au travail, ainsi qu'au rapport à l'orthographe des enquêtés. Nous regarderons si certaines pratiques ou certains discours (utiliser un ordinateur au travail, déclarer qu'il est facile de ne pas de faire de faute d'orthographe, tenir des discours plutôt moralisateurs sur les personnes qui font des fautes) prédisent de bonnes compétences en orthographe et en morphologie¹⁸⁴.

¹⁸³ Cette partie a fait l'objet d'un article pour la Délégation Générale aux Langues de France et à la Langue Française (DGLFLF) rédigé par Blondel, Brissaud et Rinck, dont la publication a été acceptée. L'article intitulé « Apport de deux enquêtes sur les connaissances orthographiques et les pratique de l'écrit des Rhônalpins » devrait paraître fin 2019.

¹⁸⁴ La relation entre ces variables a été testée à l'aide du logiciel SPSS. Pour déterminer le niveau de compétences en orthographe ou en morphologie, nous avons déterminé trois profils (sans faute, une erreur, plus d'une erreur).

2.2.1. RAPPORT À L'ÉCRIT ET USAGE DE L'ÉCRITURE

Plusieurs questions avaient trait aux usages de l'écrit, papier et numérique. Les résultats obtenus sont plutôt homogènes : 92 % déclarent écrire quotidiennement sur papier¹⁸⁵. Lorsqu'ils écrivent de manière manuscrite, tous les enquêtés déclarent rédiger des textes (par exemple des courriers) y compris de la copie de texte comme des recettes de cuisine mais également des listes (par exemple de courses).

Presque tous les foyers (92 %) que nous avons visités étaient équipés d'au moins un ordinateur¹⁸⁶, y compris chez les personnes les plus âgées de l'enquête ou vivant à la campagne. 35% des enquêtés en possèdent même deux. La fréquence d'utilisation¹⁸⁷ chez ceux qui possèdent un ordinateur est également homogène, la majorité (73%) l'utilisant quotidiennement (plusieurs fois par jour : 47% et 1 fois par jour : 26%). L'utilisation d'un ordinateur à la maison est surtout liée à l'utilisation d'internet¹⁸⁸ : 88% l'utilisent pour naviguer et écrire, principalement des emails ou des commentaires dans les réseaux sociaux (64%).

Si l'utilisation de l'écriture (manuscrite et électronique) à la maison est relativement homogène, c'est au travail en revanche que les usages de l'écriture diffèrent, avec près de la moitié des enquêtés (49%) qui n'utilisent pas d'ordinateur dans le cadre professionnel¹⁸⁹. En revanche, lorsqu'ils en utilisent un, la fréquence d'utilisation est en général très importante (plusieurs fois par jour pour 88% d'entre eux). Pour autant, les enquêtés qui utilisent un ordinateur dans le cadre de leur travail n'obtiennent pas de meilleurs résultats que les autres en orthographe¹⁹⁰ ou en morphologie¹⁹¹. Nous savons pourtant qu'une exposition quotidienne à l'écrit (y compris s'il s'agit d'écriture électronique) favorise les compétences orthographiques (Paveau et Rosier, 2008). Nous supposons que l'absence de lien dans cette enquête entre les compétences à l'écrit des enquêtés et une pratique fréquente de l'écriture, notamment dans le cadre professionnel, lieu important de préjugés (Paveau et Rosier, 2008) où l'orthographe est particulièrement surveillée, provient probablement du bas niveau de

¹⁸⁵ Question 1 : « Écrivez-vous sur papier ? ».

¹⁸⁶ Question 3 : « Est-ce que vous avez un ordinateur à la maison ? ».

¹⁸⁷ Question 4 : « À quelle fréquence utilisez-vous votre ordinateur à la maison ? ».

¹⁸⁸ Question 5 : « Utilisez-vous Internet à la maison ? » et question 5.1 : « Est-ce que vous consultez Internet ? Est-ce que vous écrivez ? Les 2 ? ».

¹⁸⁹ Question 8 : « Utilisez-vous un ordinateur au travail ? ».

¹⁹⁰ Les résultats du test statistique effectué ne sont pas significatifs (Khi2 = 0,448 ; p = 0,799).

¹⁹¹ Les résultats du test statistique effectué ne sont pas significatifs ((Khi2 = 1,709 ; p = 0,426).

difficulté de la dictée. Celle-ci ne propose pas en effet de difficultés suffisantes, notamment en morphologie (cas de base de l'accord du participe passé par exemple), pour que les enquêtés aient eu besoin de faire appel à des règles complexes, demandant mémorisation et connaissance de la langue.

2.2.2. LIEN ENTRE PERFORMANCES ET DISCOURS : ENTRE CONTRADICTIONS ET RÉALITÉ

La dernière partie de l'enquête est constituée du discours des enquêtés, recueilli par l'intermédiaire d'un entretien (questions ouvertes) faisant appel à leurs souvenirs d'apprentissage de l'orthographe, à leur rapport à la langue française et à l'orthographe ainsi qu'au niveau de compétence à l'écrit qu'ils pensent avoir acquis. 83 entretiens ont été réalisés¹⁹².

Dans cette analyse, nous nous intéressons aux attitudes des enquêtés qui se manifestent dans leurs discours et sont révélatrices de leur rapport à l'écriture et à la norme orthographique.

Entre 1987 et 1989, une enquête réalisée à Grenoble (Millet, Lucci et Billiez, 1990) avait permis de recueillir des discours sur l'orthographe d'adultes ayant un rapport quotidien avec l'écrit (enseignants, secrétaires, etc.). Les enquêtés avaient été interrogés sur leurs souvenirs d'apprentissage, leurs difficultés, et sur leur rapport à l'orthographe et à la norme, lequel apparaît notamment dans les discours sur la faute. Le protocole de cette enquête ne prévoyait pas en revanche de production d'écrit, ce qui empêche de comparer les discours et les pratiques effectives. Ainsi, à 25 ans d'écart, nous comparerons les discours sur la faute recueillis par Millet *et al.* à ceux de l'enquête de 2014.

2.2.2.1 RAPPORT À L'ORTHOGRAPHE ET À LA NORME DANS LES DISCOURS

Une grande majorité d'enquêtés semble accorder de l'importance à l'orthographe lorsqu'ils écrivent puisque 80% affirment qu'ils cherchent toujours à écrire sans faute¹⁹³. 14 enquêtés (17%) disent adapter leur degré d'exigence en matière d'orthographe en fonction de leur interlocuteur ou du type d'écrit.

67% des enquêtés « haut » estiment qu'il est facile pour eux de ne pas faire de fautes quand ils écrivent¹⁹⁴. Ce chiffre peut paraître faible à l'ère du numérique où l'accès aux dictionnaires

¹⁹² 83 entretiens ont été réalisés auprès des 87 enquêtés, les entretiens manquant provenant d'un refus et d'erreurs d'enregistrement.

¹⁹³ Question 12 : « Cherchez-vous à écrire sans faute ? ».

¹⁹⁴ Question 17 : « Finalement, est-ce facile pour vous de ne pas faire de fautes ? ».

et outils électroniques est facilité : ont-ils réellement conscience de leurs compétences orthographiques ? Ou se sous-estiment-ils dans leurs discours pour se protéger d'un jugement négatif en cas d'erreurs sur leur dictée ? Nous sommes pourtant loin du compte puisque lorsque l'on regarde le résultat obtenu sur l'intégralité de la dictée, il apparaît que seuls 8 enquêtés (9%) n'ont fait aucune erreur à l'exercice. Même en nuancant leurs discours et en tolérant quelques écarts à la norme, ils ne sont que 38 (44%) à avoir fait au maximum deux erreurs. Les productions écrites ne vont donc pas nécessairement dans le sens des discours sur la faute et il semble ainsi qu'une partie des enquêtés surestime ses capacités en matière d'orthographe. De plus, les enquêtés qui déclarent qu'il est facile pour eux de ne pas faire de fautes d'orthographe ne sont pas nécessairement ceux qui ont obtenu les meilleurs résultats en orthographe mais nous observons en revanche un lien significatif avec leurs compétences en morphographie¹⁹⁵. Les enquêtés qui pensent ne pas faire de fautes d'orthographe possèdent donc, comme les autres, des zones de faiblesse dans le domaine lexical qui leur sont propres. Leur sentiment de compétence en orthographe paraît davantage provenir d'une bonne connaissance en morphosyntaxe, ce qui leur a effectivement permis d'obtenir, dans la dictée, un bon score dans ce domaine.

Les erreurs produites par les enquêtés qui ont plutôt bien réussi la dictée (deux erreurs maximum) concernent autant l'orthographe lexicale (45%) que l'orthographe grammaticale (42%). Les erreurs de lexique portent principalement sur SOLENNEL, qui, nous le rappelons, a été particulièrement mal restitué dans les deux enquêtes (voir chapitre 5, section 2.2) mais aussi sur les redoublements de consonnes de ALLUMETTES et ARRACHER. Nous relevons enfin des erreurs isolées : *pinture, fillancé, tuyeau, rum*.

En morphographie, nous constatons :

- quelques erreurs récurrentes :
 - accord en nombre erroné sur CISEAU (*ciseaux*), TUYAU (*tuyaux*) et FORT (*forts*) ;
 - accord en genre erroné sur FIANCÉ (*fiancée*) ;
 - absence d'accord sur PLATES (*plate*) ;
- puis des erreurs isolées : *dégraisse, bricolages, rhumes, a*.

La plupart des écarts orthographiques portent sur l'ajout d'une marque de nombre ou de féminin (pour *fiancée*), que nous pouvons imputer, pour certaines, à la situation d'évaluation

¹⁹⁵ Khi2 = 13,014 ; p = 0,001.

qui a pu pousser certains enquêtés à une attitude hypercorrective (Paveau et Rosier, 2008) : le « -s » ajouté à TUYAU, BRICOLAGE et RHUME. D'autres erreurs relevées comme le marquage en nombre de l'adverbe FORT et du nom TUYAU, l'absence d'accord de PLATES ou DÉGRAISSENT, ou encore l'absence d'accent sur la préposition À peuvent révéler un problème de compétences en langue, avec certaines règles non maîtrisées. Nous nous interrogeons enfin sur les graphies *fiancée* (3 occurrences) et *ciseaux* (16 occurrences sur 87), qui reflètent peut-être davantage un effet de la fréquence¹⁹⁶ de la forme fléchie sur les productions : *fiancé* que l'on trouve plus fréquemment avec la marque du féminin¹⁹⁷ et surtout *ciseau* plus souvent écrit au pluriel¹⁹⁸ (Ortége et Lété, 2010).

Ainsi, même chez les meilleurs scripteurs, on observe une hétérogénéité dans les erreurs, dont l'origine semble variée : connaissance de la langue, influence de la fréquence ou effet du contexte d'évaluation. Malgré le lien observé entre les compétences en morphologie et le sentiment de maîtriser le code écrit, il reste une part d'enquêtés qui semble surestimer ses compétences orthographiques, en faisant l'amalgame entre de bonnes connaissances en morphosyntaxe et une maîtrise totale du système orthographique, y compris sur les mots irréguliers dont l'origine rend difficile la prédictibilité de leur graphie (consonnes étymologiques, historiques, diacritiques). Ce décalage observé entre les productions et les discours est également révélateur d'un dialogue faussé entre ceux qui pensent savoir alors qu'ils ne maîtrisent au mieux qu'une partie du système et ceux qui sont conscients de ne pas maîtriser mais qui se savent jugés. Comme le montrent Millet et al (1990), même si la question du rapport à l'orthographe dans notre société reste délicate à définir, les discours font apparaître des contradictions et une tendance à valoriser la norme orthographique.

2.2.2.2 UNE CLASSIFICATION DES TYPES DE DISCOURS

Les discours recueillis dans le cadre de cette enquête ont finalement permis de faire émerger le rapport que les scripteurs entretiennent avec la norme orthographique, notamment lorsque nous les avons questionnés sur ce qu'ils pensent de l'orthographe du français¹⁹⁹, ainsi que sur les images que véhiculent pour eux les personnes qui font des erreurs

¹⁹⁶ Indice de fréquence estimée d'Usage (U) pour 1 million de mots obtenu sur la base Manulex-infra, basé sur les tokens (calculs qui prennent en compte la fréquence réelle des mots dans les textes, soit le nombre de fois où le mot est rencontré), du CP au CM2 (Ortége et Lété, 2010).

¹⁹⁷ FIANCÉ (U = 0,46) est plus fréquent lorsqu'il porte la marque du féminin : FIANCÉE (U = 2,10) (Ortége et Lété, 2010).

¹⁹⁸ CISEAU (U = 0,76) plus fréquent avec la marque du pluriel : CISEAUX (U = 50,81) (Ortége et Lété, 2010).

¹⁹⁹ Question 2 : « Qu'est-ce que vous pensez de l'orthographe du français ? ».

d'orthographe²⁰⁰. C'est principalement dans les discours sur la faute que les enquêtés se positionnent et révèlent, en creux, leur attitude vis-à-vis de la norme.

Dans les réponses fournies par les enquêtés, trois types de discours émergent : des discours plutôt moralisateurs sur l'orthographe (38%), des discours plutôt compréhensifs (28%) et des discours contradictoires ou mitigés de la part d'enquêtés qui déclarent ne pas se positionner sur ces questions (34%). Nous avons construit ces catégories à partir de faits de discours qui nous ont semblé pertinents et qui permettent de cerner les différentes réactions des personnes interrogées.

Dans l'enquête de 1990, Millet *et al.* avaient également relevé différentes attitudes dans les discours des enquêtés, dont la grande majorité se situait dans ce que les auteurs ont appelé l'« *attitude enthousiaste* » et l'« *attitude contradictoire* ». « *L'attitude enthousiaste* » pourrait être rapprochée des discours à tendance moralisatrice de notre enquête.

LES DISCOURS À TENDANCE MORALISATRICE

32 enquêtés (soit 38%) ont tenu des discours plutôt moralisateurs et normatifs. Leurs déclarations sont centrées sur l'importance de l'orthographe pour la communication, l'efficacité et la beauté du français. Les notions de correction, de clarté et d'esthétique sont, comme le rappellent Paveau et Rosier (2008), les traits définitoires du français depuis le 17^{ème} siècle dans les représentations de la langue et fondent le « discours puriste ». Deux attitudes existent chez le puriste : une attitude nuancée qui indique les écarts à la norme et réagit à une hypersensibilité à la langue et une attitude réactionnaire qui se donne un rôle de censeur. Nous n'avons pas distingué ces différentes attitudes parmi les enquêtés que nous avons qualifiés de « plutôt moralisateurs », mais nous rejoignons la définition proposée par Paveau et Rosier (2008 : 12) : « le puriste évalue celui qui parle selon sa maîtrise de la langue, sous l'angle de la richesse lexicale et de la correction grammaticale. Il cultive la nostalgie par l'idéalisation de pratiques antérieures érigées en modèles devenus inaccessibles. Il juge et condamne souvent, et ses positions relèvent parfois de la conservation exclusive du passé et d'un protectionnisme qui peut aller jusqu'au nationalisme ».

Nous avons en effet relevé que ces enquêtés semblaient entretenir un rapport d'attachement fort à l'orthographe, notamment lorsque les locuteurs justifient une complexité avouée de l'orthographe comme une nécessité pour être efficace.

²⁰⁰ Question 6 : « *Que pensez-vous des personnes qui font des erreurs d'orthographe ?* ».

« [...] je pense que c'est une des orthographes les plus compliquées qui soit mais c'est justement ce qui fait sa beauté on va dire, son charme. » (Enquêté 5) ;

« [...] je trouve qu'elle est bonne [...] qu'elle est bien faite, elle est très explicite, elle permet de bien s'exprimer. » (Enquêté 3) ;

« Complexe, mais nous, on a une panoplie de mots qui est fantastique [...] ce n'est pas évident d'écrire comme ça, il faut toujours se poser des questions et tout ça. » (Enquêté 9).

Les arguments avancés pour justifier la complexité de l'orthographe et défendre la langue sans réserve, et parfois avec mauvaise foi, avaient déjà été relevés dans l'enquête de 1990 conduite par Millet et al.

On trouve également des discours conservateurs et lorsque la question des fautes d'orthographe est abordée, les réactions suscitées sont souvent virulentes :

« [...] ça me choque énormément. » (Enquêté 9) ;

« [...] ça m'énerve quand je vois ça. » (Enquêté 42) ;

« [...] ça me scandalise. » (Enquêté 37).

Certains enquêtés tiennent parfois des discours hyper normatifs :

« Voir des fautes [...] c'est à la limite du terrorisme » (Enquêté 5) ;

« Tout dépend ce que c'est, si c'est un 's' quand c'est du pluriel, ça je ne pardonne pas » (Enquêté 17) ;

« [...] moi certaines fautes d'orthographe personnellement ça me fait dresser les cheveux sur la tête. » (Enquêté 73).

La plupart des enquêtés qui adoptent des discours moralisateurs se présentent comme des détenteurs du savoir, allant même jusqu'à déterminer un quota d'erreurs tolérables, un type d'erreurs acceptables ou inacceptables ou encore un âge à partir duquel on ne doit plus faire d'erreurs :

« J'estime qu'à partir de cet âge-là tout le monde doit à peu près savoir écrire correctement » (Enquêté 55) ;

« Je suis assez sévère pour ça, j'estime qu'il y a des fautes qu'il ne faut pas commettre. [...] il y a des fautes grossières que je n'accepte pas. » (Enquêté 63).

On trouve ainsi de nombreux jugements négatifs vis-à-vis du scripteur qui commet des erreurs d'orthographe, ces dernières affectant plus ou moins fortement selon les discours l'image que le scripteur transmet dans ses écrits voire sa crédibilité dans le milieu professionnel. Certains parlent d'impolitesse, comme si les erreurs étaient volontaires ; d'autres d'un manque de sérieux et de crédit :

« *Écrire une lettre, un texto à n'importe qui sans faire de fautes d'orthographe pour moi c'est une marque de respect.* » (Enquête 10) ;

« [...] *ça ne fait pas très sérieux* » (Enquête 1) ;

« *Ça peut donner un mauvais a priori sur la personne [...]* » (Enquête 57).

Le poids de l'orthographe dans le milieu professionnel revient souvent dans les discours :

« [...] *ça fait pas très professionnel.* » (Enquête 57) ;

« *Les erreurs d'orthographe c'est un peu rédhibitoire quand on cherche du job.* » (Enquête 31).

On trouve également de nombreuses allusions à un manque de volonté au moment de la scolarité, comme si la responsabilité de faire des fautes d'orthographe était directement imputable aux personnes concernées :

« *Il faut avoir de la volonté hein, c'est tout !* » (Enquête 3) ;

« *Pour moi ils ne veulent pas ... c'est la simplicité* » (Enquête 10) ;

« *Je me dis qu'elles n'ont pas forcément bien travaillé à l'école* » (Enquête 24).

Paveau et Rosier (2008) soulignent que cette représentation qui consiste à croire que le travail suffit à rendre l'orthographe naturelle et intuitive est fréquente chez les locuteurs ordinaires comme chez les enseignants. Les discours sur la faute n'échappent pas non plus aux stéréotypes, avec la catégorisation en « *classes anthropologiques* » (par exemple « *les jeunes* ») qui sont les stéréotypes les plus ancrés dans la conscience populaire :

« [...] *c'est une très belle langue qui malheureusement est oubliée [...]* c'est générationnel » (Enquête 62) ;

« *C'est un truc qui se perd de plus en plus avec justement tout ce qui est ordinateur et tout ça.* » (Enquête 59) ;

« *Maintenant les jeunes ils savent plus écrire même au lycée, au collège.* » (Enquête 13) ;

« *On va finir par arriver à un jour où les gens se parleront en onomatopées.* » (Enquête 61).

Alors que le parler jeune n'est ni une réalité sociologique ni une réalité linguistique, cette catégorisation permet d'argumenter sur l'évolution du langage et la décadence de la langue. Enfin, certains discours entrent dans l'exagération avec l'argument d'une menace de voir l'écriture alphabétique transformée en écriture phonétique ou la langue prendre la forme d'onomatopées. Pour Paveau et Rosier (2008), ces représentations sont ancrées chez les traditionalistes et les réfractaires aux réformes.

Nous avons également relevé des références, récurrentes mais imprécises dans les discours, à des textes remplis de fautes ; allusions que Millet *et al.* (1990 : 58) avaient également constatées : « *Il semble [...] que la hantise, la véritable phobie de la faute d'orthographe puisse bien entraîner certains à voir des fautes, là où il n'y en a pas, et peut-être aussi à en voir beaucoup là où il n'y en a qu'une ou deux.* ».

À l'inverse des discours « plutôt moralisateurs », qui sont puristes et normatifs, nous avons relevé des discours que nous avons qualifiés de « plutôt compréhensifs », dont nous présentons les grandes caractéristiques.

LES DISCOURS À TENDANCE COMPRÉHENSIVE

Dans les discours de 23 enquêtés (28% des personnes interrogées), nous ne trouvons pas de jugement sur les personnes et c'est ce qui les distingue en particulier des enquêtés qui ont tenu des propos moralisateurs (décrits ci-dessus). Nous avons alors typé leurs discours comme plutôt compréhensifs. Les enquêtés cherchent davantage à justifier les erreurs d'autrui qu'à accabler leurs auteurs et de fait, les responsabilités sont déplacées sur le système linguistique ou le système éducatif :

*« Je n'en pense pas grand-chose parce que moi je peux en faire moi-même. »
(Enquêté 13) ;*

*« [...] ça me met plus en colère par rapport au système d'apprentissage que par rapport à la personne qui fait la faute. Parce que finalement si elle ne sait pas, si elle n'arrive pas à écrire, c'est parce que il y a un problème à la base »
(Enquêté 48).*

Une des raisons invoquée est la difficulté de la langue française ; puisque la langue est compliquée, il est normal de faire des erreurs d'orthographe :

« C'est très très dur en français l'orthographe. » (Enquêté 25) ;

« La langue française elle est tellement riche [...] elle est tellement dense, on n'aurait pas de toute une vie pour tout survoler. » (Enquêté 22).

Mais c'est surtout l'école et le système éducatif en général qui sont incriminés :

« Ils ont mal été éduqués, ils ont mal été enseignés. » (Enquêté 25) ;

« C'est pas de leur faute si on leur apprend mal. » (Enquêté 70) ;

« Au niveau des élèves je pense qu'il y a eu un souci au niveau de l'apprentissage » (Enquêté 43).

On trouve également des discours qui évoquent la fatalité, comme si l'orthographe était innée ou ne l'était pas et que l'on était doué ou on ne l'était pas.

*« J'ai toujours eu des séquelles pour l'orthographe et c'est de famille [rires]. »
(Enquête 39).*

Dans l'enquête de Millet et al. (1990), ce type de discours était présent chez près de 50% des enseignants interrogés.

Nous présentons enfin les discours que nous n'avons pas pu classer dans les deux premières catégories, du fait de la présence de nombreuses contradictions.

LES DISCOURS MITIGÉS OU CONTRADICTOIRES

La dernière catégorie d'enquêtés tient des discours que l'on peut qualifier de mitigés, voire de contradictoires. Ils sont ainsi 28 (soit 34% des enquêtés) dont les déclarations ne peuvent être classées ni parmi les discours moralisateurs ni parmi les discours compréhensifs car des représentations diverses s'y entremêlent.

« Rien de mal, j'en fais moi-même, je trouve que c'est tout à fait excusable. Après ça dépend je dirais du taux de fautes et du niveau de la personne [...] je me dis la personne elle est vraiment en difficulté, alors soit elle n'en a rien à faire et on la changera pas, soit ... et puis ça dépend de l'âge aussi » (Enquête 40).

Millet et al. (1990) avaient également relevé une « attitude contradictoire » de certains enquêtés, notamment dans les discours des enseignants, qui ont une vision morcelée de l'orthographe, tiraillés entre leur opinion personnelle et leur point de vue de professionnel : « [...] pour cette catégorie, l'orthographe en soi n'est ni facile, ni difficile, ce sont les gens qui ont ou qui n'ont pas de difficulté » (Millet et al., 1990 : 21). Ils instaurent dans leurs discours une distance avec l'orthographe en adoptant, le plus souvent, une attitude de consolation ou de résignation :

« Ben je n'en pense pas grand-chose parce que moi je peux en faire moi-même. » (Enquête 13) ;

« [...] ça m'arrive à moi alors je ne peux pas critiquer [...] je ne peux rien dire je ne veux pas me faire incriminer non, je fais du mieux que je peux, on ne peut pas juger. » (Enquête 14).

Cette position leur permet probablement de se rassurer sur leur propre niveau d'orthographe, en s'imaginant que leur indulgence vis-à-vis des erreurs des autres leur assure en retour moins de sévérité de ces derniers. Mais pour Millet et al. (1990 : 23), cette attitude n'est finalement qu'un renforcement du mythe orthographique : « l'orthographe est à ce point une instance supérieure que personne ne peut atteindre sa perfection ».

Ainsi, même si certains déclarent ne rien penser de particulier des erreurs d'orthographe, on trouve cependant de l'ambiguïté dans leurs discours ou une distanciation qui n'est peut-être qu'une façade, comme on peut le percevoir dans l'extrait suivant, qui laisse transparaître le poids de la norme et le souhait de ne pas la transgresser :

*« Ça ne me dérange pas si quelqu'un fait une faute. Après si c'est ma fille, si. »
(Enquêté 47).*

Certains enquêtés admettent que leur position est contradictoire mais d'autres ne se rendent pas compte qu'ils portent un jugement, ce dernier étant généralement peu flatteur :

« Je ne porte pas de jugement je me dis, l'erreur est humaine et en même temps ça m'interpelle vraiment » (Enquêté 37) ;

« [...] ça peut me choquer, ça dépend du niveau aussi ... niveau social de la personne quoi ... des fois je me dis oulala hein ! [Enquêteur : et est-ce que vous portez un jugement ?] Non. » (Enquêté 23).

Dans le dernier extrait, l'enquêté 23 aborde la question du niveau social qui permet généralement de la condescendance mais le jugement est néanmoins présent et là encore, peu favorable. Millet et al. (1990) soulignent que le jugement porté sur la faute n'est pas objectif et que la valeur qui lui est attribuée fluctue en fonction de l'humeur et des préjugés : *« [...] si la faute est permise à certains, pour d'autres la faute reste la faute et ne peut être qu'une marque de bêtise. »* (Millet et al., 1990 : 65).

Nous nous sommes ensuite interrogée sur l'existence d'un lien entre le rapport à la norme recueilli dans les discours des enquêtés et leurs compétences orthographiques observées dans la dictée. Nous avons ainsi regardé si les enquêtés qui ont tenu des discours moralisateurs sur la norme sont également ceux qui obtiennent les meilleurs résultats en morphologie, compétence, nous l'avons vu, particulièrement discriminante. Le test montre en effet que ce sont ces enquêtés qui réussissent le mieux la morphographie²⁰¹, et si les enquêtés plutôt compréhensifs font plutôt partie de ceux qui ont obtenu les moins bons résultats, les scores des enquêtés mitigés sont en revanche hétérogènes. Les enquêtés plutôt moralisateurs sont également ceux qui ont obtenu les meilleurs résultats sur l'intégralité de la dictée : orthographe lexicale, pseudo-mots et morphographie²⁰².

Ainsi, les enquêtés qui tiennent des discours conservateurs voire normatifs sont meilleurs que les autres en morphographie et obtiennent également de meilleurs scores à la dictée. Ces

²⁰¹ Khi2 = 12,422 ; p = 0,014.

²⁰² Khi2 = 13,185 ; p = 0,01.

derniers, probablement davantage sensibilisés à la question de l'orthographe et à son poids social, ont été particulièrement attentifs à la réussite de l'exercice et ils sont effectivement plus compétents. Mais avoir de bonnes connaissances du système orthographique ne signifie pas nécessairement être normatif ou conservateur ; les enquêtés dont les discours sont mitigés ont obtenu des résultats hétérogènes et certains ont très bien réussi l'exercice, notamment sur la morphographie.

2.2.3. BILAN : DES DISCOURS ET DES PRATIQUES CONTRADICTOIRES

L'analyse des discours sur l'orthographe mise en lien avec les pratiques effectives relevées dans les productions écrites montre que lorsqu'il s'agit d'orthographe, tout n'est pas blanc ou noir. Nous avons en effet constaté que les discours avaient tendance à se construire au fur et à mesure de la discussion et des questions, avec des idées qui évoluent et qui s'affirment voire se contredisent, alors que les entretiens ne durent pour la plupart que quelques minutes.

Concernant les questions fermées, les contradictions sont également présentes. Si les enquêtés déclarent, dans un premier temps, être plutôt à l'aise avec l'orthographe puisqu'ils ont peu ou pas de doutes et n'éprouvent que peu ou pas de difficulté lorsqu'ils écrivent (77% déclarent « non » à la question 11²⁰³), 67% admettent finalement hésiter souvent sur des points particuliers d'orthographe²⁰⁴ lorsqu'ils réfléchissent à leurs difficultés.

Par ailleurs, nous avons relevé une hiérarchisation de la faute dans les discours, avec des « *erreurs d'inattention* » auxquelles s'opposent les « *grosses fautes* », « *impardonnables* », de grammaire. Dans les discours plutôt compréhensifs, les erreurs sont banalisées. Lorsqu'ils abordent la question du type de faute, les enquêtés parlent de petites erreurs sans conséquence, pas de grosses difficultés qui affectent la compréhension : « *Ce n'est pas leur faute, ce n'est pas une tare, s'il manque un 'p'* » (Enquêté 22), à l'inverse des discours plutôt « moralisateurs », nous l'avons vu, qui rentrent dans les extrêmes et les stéréotypes.

Malgré l'importance sociale de l'orthographe, notamment dans le milieu professionnel, peu d'enquêtés soulignent que l'on accorde trop d'importance à l'orthographe. Car même si bon nombre s'en défendent, les discours et les réponses des enquêtés tendent à montrer qu'ils ont plutôt une attitude normative, souvent perceptible dès le début de l'entretien. Une

²⁰³ Question 11 : « *Vous arrive-t-il d'avoir des doutes, des difficultés, des hésitations sur la façon d'écrire un mot ?* ».

²⁰⁴ Question 13 : « *Est-ce qu'il y a des points particuliers d'orthographe sur lesquels vous hésitez souvent ?* ».

grande majorité (80%) affirme ainsi toujours essayer d'écrire sans faute, quel que soit l'interlocuteur ou le type d'écrit.

Cela rejoint les propos de Millet *et al.* (1990) qui montrent l'existence d'un tabou lié à la question de l'orthographe, que l'on perçoit notamment par l'absence d'attitude négative. Personne ne critique l'orthographe, ne la juge ouvertement : ce sont les scripteurs qui font des fautes, et ce sont eux que l'on juge. Cette attitude aurait pu s'expliquer par la sélection des enquêtés en 1990, qui étaient tous des professionnels ayant un rapport quotidien avec l'orthographe (enseignants, libraires, secrétaires, etc.) ; mais nous faisons la même observation dans notre corpus, pourtant hétérogène du point de vue des professions et des catégories sociales. Finalement, l'outil n'est pas jugé : il faut s'en accommoder, et, comme le montrent Millet *et al.* (1990), la contradiction générale qui apparaît dans les discours prend la forme d'un renforcement de la norme orthographique. Ce renforcement ne serait alors qu'une solution apportée à la diminution de la sanction orthographique : car si la norme était respectée, il n'y aurait bien évidemment plus de sanction, notamment sociale.

CONCLUSION

Cette conclusion s'articule en deux temps. Nous commençons par un bilan sur l'apport des dictées issues de l'enquête Information et Vie Quotidienne et des dictées de l'enquête complémentaire recueillies dans le cadre du projet « Compétences orthographiques des adultes et contextes sociaux d'usage » sur les connaissances orthographiques des scripteurs en difficulté avec l'écrit. Dans un deuxième temps, nous proposons des pistes d'amélioration de la dictée d'IVQ qui sont issues des résultats de nos analyses. Enfin, nous terminerons par quelques pistes didactiques issues des dernières recherches sur le sujet, concernant l'enseignement de l'orthographe à l'école, là où tout commence, et éventuellement transférables dans d'autres contextes de formation, notamment auprès d'un public adulte. À fortiori avec un public pour lequel l'enseignement de l'orthographe à l'école n'a pas parfaitement porté ses fruits, il est important de miser sur des principes et des dispositifs didactiques dont on sait qu'ils font leurs preuves à la fois au niveau de la maîtrise de l'orthographe et du rapport à l'écrit.

CE QUE LES DICTÉES NOUS APPRENNENT SUR LES CONNAISSANCES ORTHOGRAPHIQUES DE SCRIPTEURS EN DIFFICULTÉS AVEC L'ÉCRIT

Nous avons analysé, dans un premier temps, 441 dictées²⁰⁵ produites par des scripteurs considérés en difficulté avec l'écrit suite à leurs résultats à l'issue du module « d'orientation » de l'enquête Information et Vie Quotidienne (INSEE, 2011). Pour compléter ces analyses et mieux interpréter les résultats, nous avons analysé dans un deuxième temps 87 dictées produites par des scripteurs qui n'ont pas été considérés en difficulté avec l'écrit dans l'enquête IVQ (INSEE, 2011). Ces dictées proviennent d'une enquête complémentaire menée dans le cadre du projet « Compétences orthographiques des adultes et contextes sociaux d'usage », qui a reçu le soutien financier de la Délégation générale à la langue française et aux langues de France. Ce projet, mené conjointement dans les régions Rhône-Alpes et Haute-

²⁰⁵ Il s'agit d'un échantillon du corpus national, qui reprend les dictées produites dans 4 régions : Auvergne (28 dictées), Bourgogne (22 dictées), Provence Alpes Côte d'Azur (285 dictées), Rhône-Alpes (106 dictées)

Normandie en 2014, a permis de recueillir 87 dictées²⁰⁶ en Rhône-Alpes et autant de questionnaires et entretiens.

La première partie de notre travail a consisté à analyser ces deux corpus de dictées : l'un issu de l'enquête IVQ et produit par des scripteurs « bas²⁰⁷ » ; l'autre issu de l'enquête complémentaire et produit par des enquêtés « haut ». Le fichier informatique des graphies produites par les enquêtés « bas » nous a été fourni par l'INSEE et l'ANLCI qui avait déjà procédé à un recodage de certaines données. Nous avons dans un premier temps vérifié la conformité du fichier informatique des graphies avec les dictées manuscrites produites par les enquêtés. Ce travail a consisté à ajouter des données graphiques manquantes sur les mots et les pseudo-mots (comme des marques de nombre, l'accentuation, les majuscules, etc.) et à retirer des données graphiques qui avaient été ajoutées (certaines marques de nombre, mots imputés en cas d'absence de réponse, etc.). Nous souhaitons disposer d'un fichier informatique qui reflète au plus près les textes manuscrits des enquêtés. Nous avons ensuite informatisé le second corpus, composé des dictées que nous avons recueillies²⁰⁸ auprès des enquêtés « haut ». Ce premier travail a permis d'envisager une analyse fine des graphies, avec pour objectif de décrire précisément ce que les enquêtés savent faire lorsqu'on leur demande de transcrire des mots et des pseudo-mots français sous la forme d'une liste s'assimilant à une liste de courses. Nous avons commencé par une analyse « traditionnelle » des dictées, qui consiste à appliquer un score orthographique qui prend en compte l'orthographe lexicale (sur 16 points), la morphographie (sur 2 points) et la phonographie à partir des pseudo-mots (sur 6 points). Ce score (que nous appelons orthographique) permet d'effectuer un classement des enquêtés. À partir du classement par type de difficultés proposé par l'INSEE, il nous permet de rendre compte de la répartition de ces scripteurs lorsque l'on applique un barème orthographique classique.

²⁰⁶ Cette dictée comporte deux parties : la première partie est la dictée d'IVQ et la deuxième, le « rayon bricolage » qui contient du lexique et des structures syntaxiques plus complexes.

²⁰⁷ Pour rappel, le terme « bas » est repris du module de l'enquête IVQ dédié aux personnes repérées en difficultés avec l'écrit suite au « module d'orientation ». Nous avons choisi de nommer les enquêtés issus du module « bas », enquêtés « bas » par contraste avec les enquêtés qui n'ont pas été considérés en difficulté avec l'écrit et qui ont été dirigés vers le module « haut » d'IVQ, que nous appelons enquêtés « haut ».

²⁰⁸ Nous avons participé au recueil des données de l'enquête complémentaire, aidée de Catherine Brissaud (du laboratoire LIDILEM) et de Jean-Pierre Jeantheau (ANLCI).

Nous proposons dans ce qui suit une synthèse des résultats présentés dans les chapitres 5 et 6. La présentation de ces résultats sera déclinée en 4 sous-parties : le score orthographique, les compétences en orthographe lexicale, les compétences en morphographie puis les compétences en phonographie.

SCORE ORTHOGRAPHIQUE

Le score orthographique obtenu par les enquêtés « bas » (sur 24 points) montre une hétérogénéité des compétences à l'écrit chez ces scripteurs, avec des scores répartis de 0 à 24 sur 24. Cette dispersion maximale des scores chez les enquêtés « bas » ne s'observe pas chez les enquêtés « haut », chez qui la note la plus faible (qui ne concerne qu'un seul enquêté) est de 13/24. Avec des scores concentrés entre 20 et 24 (sur 24), la majorité des enquêtés « haut » a donc plutôt réussi cet exercice. La médiane se trouve à 23 sur 24 (contre 18 sur 24 chez les enquêtés « bas »). Ainsi, 92% des enquêtés « haut » ont réussi plus de 80%²⁰⁹ de la dictée, et 40% des enquêtés « bas » ; malgré cette division entre « haut » et « bas », ces enquêtés semblent en fait avoir un profil orthographique similaire. Ce résultat montre qu'un faible niveau de compétences en lecture de mots, en compréhension écrite ou en numération tels que révélés par le test d'orientation pour les enquêtés « bas », n'affecte pas nécessairement les compétences orthographiques (évaluées par les 16 mots et les accords) et phonographiques (évaluées par les 6 pseudo-mots). Si les difficultés en orthographe sont souvent associées à des difficultés dans les autres domaines de l'écrit (lecture, compréhension) ainsi qu'en numération, cette corrélation n'est pas systématique : « De fortes difficultés à l'écrit sont un assez bon révélateur de difficultés en compréhension de l'écrit. En revanche, une bonne réussite en orthographe ne va pas systématiquement de pair avec une bonne compréhension de l'écrit ou de l'oral. » (Blondel et al., 2016 : 37). La dispersion des scores parmi les enquêtés « bas » est révélatrice en effet d'une réalité multiforme des compétences orthographiques, y compris chez un public repéré en difficulté avec l'écrit. Il apparaît en effet que 5% des enquêtés « bas » (ce qui représente 21 personnes) n'ont pas été en difficulté sur cet exercice et ont obtenu 100% de réussite, et 9% (soit 40 personnes) n'ont produit qu'une seule erreur. Ainsi, le module « d'orientation » de l'enquête IVQ a permis de repérer les personnes les plus en difficulté avec l'écrit, notamment les 60% d'enquêtés « bas » qui ont obtenu moins de 80% de réussite sur la dictée et a permis d'écarter les enquêtés « haut » puisqu'aucun de ces enquêtés « haut » n'a obtenu de score inférieur à 13 sur 24. Néanmoins, cette dictée ne comporte pas de difficultés orthographiques majeures et l'on ne

²⁰⁹ Les enquêtés qui ont réussi plus de 80% de la dictée ont obtenu un score compris entre 20 et 24 sur 24.

peut pas affirmer que les enquêtés qui ont obtenu moins de 80% de réussite sur cet exercice (60% des enquêtés « bas » et 8% des enquêtés « haut ») n'ont pas de difficultés en orthographe. Le score de réussite obtenu par les 87 enquêtés « haut » sur la dictée complémentaire (la liste concernant le « rayon bricolage ») montre des écarts entre les enquêtés plus importants que ceux observés sur la partie IVQ de la dictée. Les scores sont plus dispersés, compris entre 6 et 22, malgré une médiane qui reste élevée (21 sur 22). On peut noter notamment que la proportion de très bons scores est moins élevée avec 79% des enquêtés « haut » qui ont obtenu plus de 80% de réussite sur le « rayon bricolage » contre 88% pour la partie IVQ. Par ailleurs, une proportion similaire d'enquêtés a obtenu un score maximal sur les deux parties de la dictée (29% pour la première partie et 27% pour le « rayon bricolage »). Ainsi, 9% des enquêtés « haut » qui n'avaient pas été en difficulté dans la dictée d'IVQ, le sont dans la dictée complémentaire, sur le « rayon bricolage ».

COMPÉTENCES EN ORTHOGRAPHE LEXICALE : L'IMPORTANCE DE LA CONSISTANCE ET DE LA FRÉQUENCE

Dans la dictée d'IVQ, les écarts de réussite sont importants selon les mots, chez les enquêtés « bas » comme chez les enquêtés « haut ». Chez les enquêtés « bas » l'écart maximal se trouve entre le mot SEL, réussi à 93% et SOLENNEL, réussi à seulement 19%. Le classement des mots en fonction du pourcentage de réussite et d'erreurs fait apparaître quasiment le même classement dans les deux populations. Ce sont les mots particulièrement consistants (avec un indice de consistance supérieur à 80 selon Manulex) et dont les correspondances phonie-graphie sont les plus régulières du corpus qui ont été les mieux réussis (SEL, TOMATE, FROMAGE-S). La notion de « consistance » que nous avons présentée dans la partie théorique semble avoir un impact particulièrement important sur la réussite des enquêtés « bas ». Cela semble confirmer que chez les scripteurs en difficulté avec l'écrit, la régularité des correspondances phonie-graphie est un facteur important de réussite orthographique. Chez les enquêtés « haut » aussi, les mots très consistants ont été particulièrement bien restitués mais les écarts de réussite se creusent surtout sur les mots dont le degré de consistance est moindre (CERISE, PAYS). Ainsi, le lexique qui a posé problème aux personnes les plus en difficulté à l'écrit a également davantage posé problème aux enquêtés « haut », bien que chez ces derniers les proportions d'erreurs soient bien inférieures.

Par ailleurs, la consistance orthographique ne semble pas être le seul facteur qui impacte les résultats. Certains mots peu consistants ont été bien restitués par les deux populations : par exemple PHARMACIE, qui combine des graphèmes moins fréquents ou FORT, qui possède une lettre finale muette. Ces deux mots sont des mots plutôt fréquents si l'on se réfère à Manulex,

ce qui semble également confirmer l'importance de la fréquence sur la réussite orthographique.

Ainsi, les erreurs relevées sur les mots très consistants de la dictée chez les enquêtés en difficulté sont révélatrices des difficultés inhérentes à l'écriture du français, à commencer par la variété de graphèmes possibles pour un même phonème. Dès lors que l'on ne se situe pas dans une écriture exclusivement phonographique et parfaitement consistante, de nombreuses possibilités existent et les graphies non conformes à la norme orthographique observées en particulier chez les enquêtés « bas » l'attestent. Tout est possible, en quelque sorte, et seul un apprentissage (qu'il se fasse de manière explicite à l'école ou qu'il soit implicite par l'exposition à l'écrit) peut permettre au scripteur de parvenir à la juste orthographe. Les enquêtés « bas » ont des connaissances, même partielles, comme l'atteste par exemple la restitution du « -e » final sur les mots de la dictée. Les finales sont emblématiques des difficultés de l'écriture du français et le « -e » final n'a pas de valeur de phonogramme, mais sur PHARMACIE, ÉPICERIE et LIBRAIRIE, entre 80 et 89% des réponses selon les mots présentent bel et bien un « -e » final.

En somme, les réussites et les difficultés observées chez les enquêtés « bas » et « haut » portent sur les mêmes mots : ceux qui sont moins consistants ou moins fréquents, même si les taux de réussite sont bien supérieurs chez les enquêtés « haut ». Chez les enquêtés « bas », il semble que la régularité des combinaisons de phonèmes et de graphèmes et donc l'effet de consistance ait un impact particulièrement important sur la réussite. La fréquence des mots impacte également les résultats chez les deux populations mais il semble particulièrement présent chez les enquêtés « haut » lorsqu'il s'agit de mots très rares, comme SOLENNEL que seule la moitié des enquêtés « haut » a correctement restitué. Ces résultats sont intéressants à deux niveaux : d'une part, pour mieux comprendre le système orthographique, car les données sur les réussites et les difficultés des scripteurs « haut » et « bas » vont dans le sens de ce que l'on sait sur le système orthographique et montrent en particulier la pertinence des notions de consistance et de fréquence. D'autre part, ces résultats donnent des indications concrètes sur la manière de tester les compétences orthographiques en sélectionnant soigneusement les items en fonction des critères de consistance et de fréquence. On pourrait imaginer un test « à étages » qui commencerait par tester des mots très consistants (avec des correspondances phonographiques très stables comme « a » pour /a/ ou « i » pour /i/), puis des mots moins consistants (avec l'archiphonème /E/, les finales en /o/, par exemple une finale en « -eau » et, pour les consonnes, typiquement /s/, dans des mots fréquents, puis des mots moins fréquents.

LA MORPHOGRAPHIE : UNE COMPÉTENCE PARTICULIÈREMENT DISCRIMINANTE

Nous avons ensuite observé les compétences en morphographie des enquêtés d'après les deux marques de nombre présentes dans la dictée d'IVQ (l'accord d'un nom, FROMAGES et d'un verbe, SENTENT) et d'après l'absence de marque sur un adverbe (FORT).

Chez les enquêtés « bas », il y a presque autant d'enquêtés qui ont réussi les deux marques de nombre (36%), que d'enquêtés qui n'en ont réussi aucune (38%). Lorsqu'ils n'ont réussi qu'une seule des deux marques, c'est l'accord nominal qui est davantage maîtrisé avec 92 marques réussies contre 29 sur le verbe. Chez les enquêtés « haut » en revanche, 85% ont réussi les deux marques de nombre attendues et la différence de réussite selon la catégorie grammaticale des mots est très peu marquée. Les deux accords en nombre présents dans la dictée d'IVQ permettent de distinguer les enquêtés qui ont d'importantes difficultés en morphographie (ceux qui n'ont réussi aucune des deux marques). On pourrait rapprocher ces résultats de ceux obtenus avec des écoliers. En CE1, dans la recherche *Lire et écrire au CP* : le marquage en nombre nominal est réussi par moins de 50% des élèves en CE1 et l'accord verbal réalisé par environ 8% des élèves (Geoffre et al., 2019). Les scores fournis par la DEPP pour la fin de la scolarité élémentaire sont plus élevés (Andreu, Steinmetz, 2016) : dans la dictée proposée à un grand nombre d'élèves de CM2 en 2015, le nom GAMINS dans LES GAMINS est écrit conformément à la norme par 85,5% d'élèves et le nom GARÇONS dans LEURS QUATRE GARÇONS par 78% des élèves.

Les deux accords présents dans la dictée d'IVQ ne suffisent pas, en revanche, pour cerner les compétences dans ce domaine des personnes qui ont de meilleures connaissances orthographiques. En effet, nous ne pouvons pas affirmer que les 85% d'enquêtés « haut » et les 36% d'enquêtés « bas » qui ont réussi les deux marques de nombre attendues dans IVQ, n'ont pas de difficulté en morphographie. La dictée complémentaire du « rayon bricolage », qui comporte des structures syntaxiques plus complexes, montre que 76% des enquêtés « haut » ont obtenu plus de 80% de réussite, avec plus de 10 marques réussies sur les 14 attendues. Pour ces enquêtés, la morphographie ne semble pas constituer une difficulté orthographique importante. Sur le « rayon bricolage », certains enquêtés « haut » en revanche ont éprouvé des difficultés qui n'étaient pas apparues dans IVQ, notamment dans les tournures plus complexes où aucun signal clair ne vient aider au marquage du pluriel (par exemple, le « -s » de ALLUMETTES dans UNE BOITE D'ALLUMETTES). On retrouve cette difficulté, également décrite par Lucci et Millet (1994), dès la fin du CP, lorsque les élèves entrent dans le marquage morphosyntaxique. L'effet de position du déterminant apparaît alors comme primordial pour la réussite de l'accord (Geoffre et al., 2019). Cet effet de la

position du déterminant et de la clarté du signal qui permet de déclencher l'accord est toujours observé en fin de scolarité au collège (Brissaud, 2015). La redondance des marques de pluriel à l'écrit qui rend complexe, pour le scripteur, la gestion en continu et de manière systématique des chaînes d'accord serait une explication pour ces différents auteurs. Selon l'apport de la psycholinguistique, le lien serait également à chercher au niveau de la surcharge cognitive qu'implique la gestion d'une chaîne d'accord. Ici, la structure demande au scripteur une réflexion (que l'on sait coûteuse en situation de scripture) pour réaliser l'accord : la forme « de », préposition ou déterminant, ne permet pas de voir ni d'entendre le pluriel et la structure du groupe nominal avec complément du nom demande à être analysée (il y a une seule boîte, mais à l'intérieur plusieurs allumettes et la pluralité appelle le pluriel).

Lorsque l'accord en nombre n'était pas explicitement demandé, notamment pour les items dictés sans déterminant et sous forme de liste, 34% des enquêtés « bas » ont marqué le nombre sur CERISES, où il est fait appel à la réflexion des enquêtés en lien avec le contexte. Ainsi, chez les enquêtés « bas », il y a presque autant d'enquêtés qui ont réalisé les deux accords en nombre attendus que d'enquêtés qui ont marqué le nombre là où il était « implicitement » attendu. Sur les noms de rayon ou de magasin (pharmacie, librairie), nous ne relevons que 7 graphies au pluriel. Cette absence de marque, à opposer à celle ajoutée sur CERISES, semble montrer une conscience des règles de fonctionnement en matière de morphographie chez les enquêtés « bas ». Nous soulignons que cette analyse appelle à la prudence, car on l'a vu avec le cas de l'adverbe FORT, l'absence de marque peut être une réussite par défaut dans la mesure où en français il n'y a en général pas de marquage du singulier. Par ailleurs, nous relevons un comportement hétérogène des enquêtés vis-à-vis de la présence ou de l'absence de marques de nombre : certains enquêtés qui ont réussi les deux marques attendues sur le nom et le verbe, n'ont pas marqué le pluriel sur CERISES et, inversement, parmi les enquêtés qui n'ont réussi aucune des deux marques attendues, certains ont marqué les pluriels implicites (souhaitables). Chez ces enquêtés, on peut se demander s'il ne s'agit pas d'une récupération d'instances (plutôt que d'une volonté de marquer le nombre). CERISES avec la marque de flexion apparaît en effet dans Manulex comme deux fois plus fréquent que la graphie au singulier CERISE ($U = 25,41$ vs $U = 11,67$). L'ajout d'un « -s » est d'ailleurs massivement observé sur les graphies qui sont correctes d'un point de vue orthographique, avec 90% de « -s » observés sur la forme normée CERISE. Chez les enquêtés « haut », ce phénomène ressort de manière encore plus forte avec 56% des graphies de CERISES qui ont une marque de nombre et toutes (sauf une) sont observées sur la graphie CERISE conforme à la norme.

Que les formes produites soient issues d'une récupération d'instance ou d'un calcul morphographique, ces observations vont dans le sens de ce qui a été noté concernant le lien entre orthographe lexicale et morphographie. Chez les enquêtés « bas » comme chez les enquêtés « haut », ce sont les scripteurs qui réussissent le mieux à restituer la forme orthographique des mots qui ont le comportement le plus adapté s'agissant de la morphographie (réalisation des accords en nombre et ajout des marques que nous avons appelées marques « implicites »).

LA PHONOGRAPHIE : UNE COMPÉTENCE DANS L'ENSEMBLE MAÎTRISÉE

Enfin, la dictée d'IVQ nous permet de rendre compte des compétences en phonographie des enquêtés d'après leur capacité à fournir une graphie qui restitue la forme phonologique de 6 pseudo-mots. Nos résultats montrent qu'ils s'avèrent peu discriminants pour rendre compte des compétences en phonographie, chez les enquêtés « bas » comme chez les enquêtés « haut ». Chez les enquêtés « bas », les scores de réussite sont relativement élevés (compris entre 68% et 85%) et les phénomènes phonographiques à observer sont limités. Chez les enquêtés « haut », l'analyse des réponses non conformes à la norme orthographique sur les mots, montre que leurs erreurs sont principalement orthographiques. Tester leurs compétences en phonographie ne semble donc pas pertinent.

En revanche, dans le cadre de ce travail, l'observation de la variation graphique observée sur les pseudo-mots chez les enquêtés « bas » et « haut » révèle des résultats intéressants, notamment concernant les comportements de scripteurs. Sur des mots inventés, les possibilités de graphies sont importantes car elles sont liées à la diversité de certains graphèmes pour transcrire un même phonème, ce qui, comme nous l'avons vu dans la partie théorique, représente une difficulté caractéristique du système orthographique du français. L'observation des choix graphiques (variantes) opérés par les enquêtés « bas » et « haut » qui ont proposé une transcription acceptable du pseudo-mot et l'analyse de la récurrence des réponses montrent une même tendance : lorsqu'ils ont le choix, les scripteurs « bas » comme les scripteurs « haut » choisissent préférentiellement une graphie *minimale*. Par minimale, nous entendons une orthographe épurée (composée essentiellement d'archigraphèmes), qui utilise des suites de graphèmes relativement stables et fréquentes (Catach, 1980) et avec une morphographie minimale, dont la tendance consiste à ne pas ajouter de finale muette, sauf pour GOBAR (voir paragraphe suivant). Les enquêtés « bas » et « haut » ont ainsi en majorité, fait les mêmes choix en utilisant des combinaisons de graphèmes similaires, qui sont sans doute les plus caractéristiques du lexique du français : *micatol* pour MICATOL, *sirap* pour SIRAP, *jock* pour JOC, etc.

Nous avons observé plus en détail le comportement des scripteurs quant à l'ajout ou non d'une finale muette sur les pseudo-mots. Pour voir si des comportements en particulier sont à relier avec des compétences orthographiques, nous avons réparti les enquêtés en quatre groupes de scripteurs (dont le classement dépend principalement de leur score en morphographie). Sur les 4 pseudo-mots de la dictée utilisés pour cette analyse (MICATOL, SIRAP, PÉVANOR, GOBAR), aucune finale muette n'était nécessaire pour que la consonne finale se prononce. Sur MICATOL, SIRAP et PÉVANOR, l'analyse des réponses montre très clairement que les deux groupes les plus compétents en orthographe optent pour une graphie minimale sans ajout de lettre muette. Il semble qu'ils aient davantage enregistré que le « -e » muet n'est pas obligatoire pour que le « -l », le « -p » et le « -r » en finale se prononcent. Sur GOBAR en revanche, plus les enquêtés sont compétents et plus ils privilégient une finale muette et le choix de la finale chez les deux groupes les plus compétents correspond à la fréquence des finales observée sur les mots se terminant par /ar/ en français (le « -d »).

Si les pseudo-mots représentent une opportunité de rendre compte des compétences en phonographie, cette compétence peut également être saisie dans l'analyse des graphies non conformes à la norme orthographique relevées dans les dictées. La prise en compte de la phonographie sur les mots nous a permis de mettre en relief le fait qu'une grande partie des enquêtés « bas » qui ont des difficultés avec la norme orthographique parvient néanmoins à produire un énoncé lisible. Ainsi, lorsqu'ils essayent de transcrire phonétiquement les mots dictés, la plupart y parviennent en appliquant les règles de position et de distribution spécifiques au français. Quelques enquêtés montrent qu'ils possèdent des connaissances partielles en phonographie, en proposant des formes graphiques qui sont plausibles mais en dehors des valeurs de position des graphèmes, car leurs graphies ne tiennent pas compte de l'environnement graphémique (ex. *épiserie*). Enfin, les compétences en phonographie n'apparaissent pas seulement dans la restitution de la forme phonologique des mots mais aussi dans le choix de graphèmes qui attestent qu'ils ont enregistré certaines régularités (absence de combinaisons de graphèmes non attestées en français, préférence pour les graphèmes les plus fréquents, etc.). Par ailleurs, comme pour les mots très consistants, certaines formes graphiques montrent que le scripteur ne va pas toujours dans le sens d'une relation simplifiée entre phonèmes et graphèmes (on trouve par exemple « mm » dans la graphie *tommate* pour TOMATE). Généralement, les choix proposés sont des choix plausibles qui respectent les grands principes de l'écriture du français et qui montrent que le scripteur a intériorisé les difficultés de l'orthographe. Mais les enquêtés sont parfois comme perturbés même quand il s'agit d'orthographier des mots simples « comme ils se prononcent » (*tommate, thomate, tomat*). Ce sont des difficultés comparables à celles des élèves au cours

de leur apprentissage ; par exemple, le fait de découvrir « gu » peut interférer avec « g » et conduire à la surgénéralisation du premier, de même que pour la cédille sur « c ». C'est aussi ce qu'ont montré Thévenin et al. (1999) à propos des marques de nombre et des étapes de l'apprentissage : la marque « -s » sur les noms va conduire à leur surgénéralisation sur les verbes.

Par ailleurs, nous avons vu qu'une prise en compte large du principe phonographique (avec les graphies considérées comme « approximantes » c'est-à-dire plausibles en mettant de côté certaines contraintes comme la valeur de position dans *épicerie* évoquée dans le paragraphe précédent) ne change pas les résultats moyens : le score moyen est légèrement supérieur et gagne 2 points en moyenne (87%). La prise en compte de ce critère n'a en effet d'incidence que sur peu de mots : SENTE(NT) (6%), CERISE (2%), ÉPICERIE (5%), PAYS (4%), FIANCÉ (4%) et RHUME (14%). Pour SOLENNEL, il ne s'agit que de deux graphies, ce qui est tout à fait marginal dans notre corpus. À la lecture de ces graphies, les problèmes de correspondances entre phonèmes et graphèmes peuvent faire naître un sentiment d'étrangeté au niveau visuel qui peut gêner le déchiffrement du mot comme dans *saile* pour SEL, *solanaille* pour SOLENNEL. Nous avons néanmoins souhaité classer ces graphies dans une catégorie qui leur est propre parce qu'elles révèlent des compétences qui sont à distinguer des graphies très éloignées de ce qui est attendu : par exemple *nhamacia* pour PHARMACIE, *outri* pour ANTI, *mump* pour RHUME. Dès que l'on sort du principe phonographique, on obtient un nombre important de formes graphiques différentes, parfois indécodables dans un contexte qui ne serait pas celui d'une dictée et ne serait-ce que parce qu'au niveau visuel, le mot est très éloigné de ce avec quoi le lecteur est familier. Ce point représente donc une priorité au niveau didactique : s'appuyer sur le principe phonographique et les correspondances phonèmes-graphèmes les plus régulières.

QUE DISENT LES ENQUÊTÉS DE L'ORTHOGRAPHE ?

L'enquête complémentaire a permis de recueillir, en plus des dictées, 87 questionnaires et entretiens. L'analyse des entretiens (qui ont été enregistrés) effectués auprès des enquêtés « haut » montre de nombreuses contradictions que l'on a pu observer entre les discours sur l'orthographe et les pratiques effectives observées dans les écrits. Ces contradictions ont également été relevées dans les discours eux-mêmes, avec des idées qui évoluent, qui s'affirment ou parfois se contredisent.

Ce corpus représente un matériau riche. Nous avons souhaité faire ressortir les principales représentations de l'orthographe et les mettre en lien avec les scores des enquêtés « haut ». Il apparaît qu'une grande majorité des enquêtés interrogés ne sont pas au clair avec leurs

pratiques de l'écrit et leurs compétences orthographiques. Ils estiment (à 77%) ne pas avoir (ou avoir très peu) de doutes lorsqu'ils écrivent et ne pas éprouver de difficultés particulières lorsqu'ils écrivent. Pourtant, parmi eux, beaucoup (67% des 77%) admettent finalement hésiter souvent sur des points d'orthographe particuliers.

Lorsque l'on aborde l'orthographe, la « faute » apparaît systématiquement dans les discours et celle-ci relève d'une hiérarchisation qui semble fonction de la personne (si ce sont des personnes extérieures qui font des erreurs d'orthographe, c'est mieux que l'entourage et encore plus que les enfants) et du domaine de l'orthographe, avec les fautes dites « de grammaire » souvent considérées comme « impardonnables ». Une partie des enquêtés se révèlent compréhensifs vis-à-vis des difficultés orthographiques (ou semblent l'être à première vue), tandis que d'autres peuvent se montrer virulents. Il est alors apparu que c'est principalement dans les discours sur la faute que les enquêtés se positionnent et qu'ils révèlent, en creux, leur attitude vis-à-vis de la norme. Malgré l'importance que la société accorde à la maîtrise de la norme, notamment dans le milieu professionnel, peu d'enquêtés en font la critique, comme si l'orthographe était une institution sacrée à laquelle on n'a pas le droit de s'attaquer. Nous avons en particulier observé que même si bon nombre d'enquêtés s'en défendent, leurs discours et leurs réponses tendent à montrer qu'ils ont plutôt une attitude normative, souvent perceptible dès le début de l'entretien. Une grande majorité (80%) affirme ainsi toujours essayer d'écrire sans faute, quel que soit l'interlocuteur ou le type d'écrit. Cette attitude était déjà décrite par Millet et al. (1990) et souligne le poids de la norme et l'insécurité des scripteurs.

APPORT DE NOTRE TRAVAIL SUR L'ÉVALUATION À L'ÉCRIT DES ADULTES : QUELQUES PISTES POUR AMÉLIORER LA DICTÉE D'IVQ

Nous avons vu que la composante écriture est souvent absente des tests destinés à évaluer les compétences à l'écrit des adultes, notamment dans les enquêtes réalisées à grande échelle et destinées à une comparaison des compétences entre les pays participants (comme IALS). Pourtant la capacité d'écriture est une compétence essentielle en littéracie et nous avons vu que la maîtrise de l'orthographe n'est pas nécessairement liée à la maîtrise des autres composantes de l'écrit, qui doivent donc être évaluées indépendamment les unes des autres. Il semble que l'absence de test d'écriture soit surtout liée aux contraintes que sa mise en œuvre implique, et à fortiori dans une enquête destinée à comparer les performances entre les pays. On sait aussi que les scripteurs entretiennent un rapport à l'écriture et à l'orthographe très différent et très inégal selon les pays. Certains systèmes orthographiques, comme celui du français, nécessitent un long apprentissage qui continue de poser des

difficultés à l'âge adulte. À ces difficultés de comparabilité (qui avaient d'ailleurs été soulevées par la France dans IALS), s'ajoutent des contraintes techniques liées à la saisie des données si les tests sont manuscrits, à l'exploitation des réponses et surtout aux critères d'évaluation. Si chez les enfants, il existe des outils permettant de cibler les compétences attendues en fonction de la classe d'âge et du niveau d'acquisition, chez les adultes, il existe peu de données sur lesquelles se baser lorsque l'on cherche à déterminer les compétences à l'écrit. Le test d'orthographe, proposé dans l'enquête IVQ en 2004 puis en 2011, qui comporte 16 mots que nous considérons comme exploitables et 6 pseudo-mots est une vraie originalité en terme d'évaluation des adultes : il ne s'agit pas de quelques mots isolés comme certaines enquêtes le proposent. Cependant, on voit bien, lorsque l'on se penche sur les données, les difficultés que pose l'exploitation d'un corpus de textes manuscrits, difficultés auxquelles nous avons dû faire face pour nous assurer de la fiabilité des données. Par ailleurs, le fait que la population enquêtée ne concerne que les personnes considérées en difficulté avec l'écrit se révèle être un premier obstacle en termes d'interprétation des réponses. Quand un mot a été mal restitué par les enquêtés « bas » et « haut » (comme SOLENNEL dans IVQ), il convient de s'interroger sur la pertinence d'un tel item.

Enfin, le fait que le repérage des difficultés à l'écrit provienne d'un test d'orientation qui repose sur le postulat qu'il existe un lien entre les compétences orthographiques et les autres compétences évaluées (lecture, compréhension et numératie) pose également question. Le risque est que certains enquêtés en difficulté avec l'écriture (et non avec la lecture ou la compréhension) ne soient pas repérés (et inversement), ce que semble indiquer le taux de 40% d'enquêtés « bas » et « haut » pour lesquels nous avons montré qu'ils avaient un profil orthographique similaire sur la dictée d'IVQ. Ensuite, les compétences évaluées par l'INSEE dans IVQ, concernent essentiellement l'orthographe lexicale (et le respect de la norme orthographique) ; compétence qui n'est pas la seule qui entre en compte lorsque l'on évalue la capacité d'un scripteur à communiquer par écrit. L'une des réussites à valoriser représente aussi le fait que le message soit lisible pour le lecteur (la graphie *tommate* pour TOMATE ne gêne en rien le message). Il nous semble qu'un test d'écriture devrait prendre en compte les deux dimensions : la dimension orthographique et la dimension phonographique. La première est essentielle parce que c'est effectivement celle qui est attendue et qui est socialement valorisée. Mais nous le verrons dans nos propositions, les compétences lexicales évaluées doivent faire l'objet d'une sélection qui tienne compte de la consistance et de la fréquence des mots. Il nous semble aussi que la morphographie devrait tenir une place plus importante, qui permette de rendre compte de compétences minimales dans ce domaine, comme dans le « rayon bricolage » de l'enquête complémentaire.

La dimension phonographique doit selon nous permettre de rendre compte de la capacité d'un scripteur à produire un message lisible, au-delà de la seule contrainte liée à la norme orthographique. La correction des graphies non normées pose des problèmes de méthodologie liés à l'acceptabilité ou non de certaines formes et dont nous avons discuté dans ce travail, au gré des réponses des enquêtés. Une graphie comme *serise* n'empêche pas le lecteur de comprendre le message ; la graphie *solanaille* en revanche perturbe la lecture et la frontière entre ce qui est acceptable et ce qui ne l'est pas se situe selon nous ici. Cette dimension phonographique devrait être davantage prise en considération dans les critères d'évaluation qui servent à déterminer une situation d'illettrisme ou de faible compétence en littéracie. Nous avons montré que de ce point de vue, un scripteur qui a mis en place un système d'écriture suffisamment efficace pour transmettre un message lisible ne devrait pas être considéré comme illettré, ou, du moins, que cette catégorie demande à être affinée. Il faut en effet analyser plus précisément les compétences et reconnaître que certains scripteurs possèdent des connaissances phonographiques non négligeables. Les connaissances orthographiques relèvent d'un niveau supérieur de maîtrise que certains scripteurs n'ont pas acquis dans sa totalité.

L'analyse des dictées réalisées par les deux populations enquêtées et l'apport du « rayon bricolage » pour rendre compte des connaissances orthographiques des enquêtés « haut » nous montrent plusieurs éléments qui nous semble-t-il, doivent être pris en compte lorsque l'on évalue des adultes à l'écrit. Nous proposons à présent quelques pistes pour améliorer la dictée d'IVQ, qui sont issues de nos résultats

QUELQUES PISTES D'AMÉLIORATION

Nous discutons pour commencer du choix des mots pour évaluer l'orthographe lexicale, qui tient compte de la consistance et de la fréquence des mots, ainsi que des contraintes liées au recueil de graphies manuscrites.

Dans le corpus de mots présents dans la dictée d'IVQ, il manque de contrastes plus nets pour pouvoir mesurer l'impact de l'effet de consistance et de fréquence : les mots les plus rares sont aussi les moins consistants. Les résultats obtenus par les enquêtés « bas » montrent que chez une population en difficulté, l'effet de consistance et donc de la régularité des combinaisons de phonèmes et de graphèmes est un facteur important de réussite. Mais la présence de correspondances phonie-graphie irrégulières fait baisser les scores chez tous les enquêtés, y compris chez les enquêtés « haut ». Ce phénomène a notamment été observé sur l'item SOLENNEL. Cet effet de la consistance orthographique est présent dès le début de l'apprentissage de l'orthographe par les élèves. Geoffre et al. (2019) montrent qu'en CE1, 80%

des élèves écrivent conformément à la norme orthographique des mots consistants (comme PATIN, CAPUCHON) et que les difficultés d'écriture augmentent avec la présence d'accents et de graphèmes complexes (comme le « t » de RÉCRÉATION). Geoffre et al. (2019) soulignent par ailleurs qu'en CP, l'exigence de l'écriture orthographique de mots dans une phrase est difficilement réaliste, même si la phrase est courte et qu'elle ne comporte pas de marque de morphographie. Enfin, l'effet de fréquence joue un rôle non négligeable sur la réussite orthographique puisque certains mots particulièrement irréguliers comme PAYS, montrent que même les scripteurs en difficulté ont mémorisé certaines formes graphiques complexes, par exemple sous l'effet de la scolarisation ou la fréquentation des mots au quotidien). Dans une étude menée par Pérez et Giraud (2019), des mots choisis pour leur complexité orthographique et leur familiarité ont été soit lus, soit lus et copiés, puis dictés à des élèves de CE1²¹⁰. Dans cette étude, pour choisir les mots utilisés dans l'expérimentation, c'est le critère de la familiarité qui a été préféré à celui de la fréquence textuelle : « [...] tout en étant corrélés, fréquence et familiarité se distinguent par leur caractère objectif vs subjectif. » (Pérez et Giraud, 2019 : 203). Ce critère de la familiarité nous paraît intéressant pour rendre compte des difficultés rencontrées par les adultes sur des mots du quotidien (ceux-ci n'étant pas nécessairement des mots « fréquents » au sens strict du terme, comme nous l'avons vu avec PHARMACIE dans la dictée d'IVQ). Les résultats de cette étude montrent un effet significatif de l'effet de familiarité sur la réussite orthographique chez des élèves de CE1.

Il n'existe pas à notre connaissance, d'outil qui recense les mots selon le critère de la familiarité. Le cas échéant, nous recommandons l'utilisation de deux outils :

- Manulex, pour disposer d'indices de fréquence et de consistance, en choisissant des mots dont l'indice de fréquence serait supérieur à $U = 20$;

Pour des mots très consistants, comme FROMAGE, il n'est pas nécessaire de se reporter à leur indice de fréquence. Ainsi dans la dictée d'IVQ, $U > 20$, aurait exclu : RHUME, SOLENNEL, LIBRAIRIE, ÉPICERIE, ALCOOL, FIANCÉ, FORT utilisé comme adjectif. Malgré cette recommandation, nous conservons PHARMACIE parce que la fréquence de cet item dans Manulex ne semble pas correspondre à la fréquence d'exposition aux mots d'adultes en difficulté avec l'écrit (nous rappelons qu'il a été plutôt bien orthographié, malgré le fait que ce mot soit peu consistant et notre hypothèse est que ce résultat s'explique par un critère de fréquence dans l'environnement quotidien) ;

²¹⁰ 4 catégories d'items ont été proposés aux élèves : mots familiers, mots non familiers, mots complexes/familiers, mots complexes/non familiers.

- EOLE pour vérifier que les mots choisis correspondent aux acquis des élèves en fin de cycle 3.

L'ORTHOGRAPHE LEXICALE

Dans une dictée destinée à évaluer les compétences en littéracie, nous recommandons, pour tester les compétences en phonographie des enquêtés ainsi que leur connaissance du lexique, des mots fréquents qui jouent sur le degré de consistance : des mots fréquents / consistants et des mots fréquents / peu consistants puis en dernière instance de recourir à des mots moins fréquents qui sont ceux qui vont distinguer entre eux les scripteurs de niveau plus avancé. Pour rendre compte de la fréquence d'exposition aux mots d'une population adulte, il manque pour la recherche, d'un outil axé sur l'exposition aux mots du quotidien : ceux qui sont visibles dans la rue, dans le métro et dans les écrits grand public (magazines, publicités, etc.). C'est ce qu'on a vu avec l'exemple de PHARMACIE : sa fréquence dans Manulex mériterait d'être couplée à un indice sur sa fréquence dans la vie quotidienne des personnes interrogées (effet de familiarité).

Une attention doit également être portée au choix de mots qui représentent les difficultés inhérentes à l'écriture du français. Nous proposons des mots qui permettent d'illustrer notre propos mais ils ne sont que des propositions.

- En premier lieu, la variété de graphèmes possibles pour un même phonème, en proposant des mots qui transcrivent différemment un même phonème :
 - /f/ = « f » vs « ph » : PHARMACIE vs FROMAGE ;
 - /o/ = « o » vs « eau » : POT, FAUX, EAU ;
 - /ɔ/ : FORT, GORGE, POMME ;
 - /e/ : « é », « ai », « et » : ÉPICERIE, FRAISE, PAQUET ;
 - /s/ : CERISE, PHARMACIE, SEL, SALADE²¹¹ ;
 - /é/ : PAIN, PEINTURE ;
 - /ø/ : « e » vs « eu » vs « œu » : LE, ŒUFS, BLEU.

Nous avons vu dans les réponses non conformes, que certains enquêtés ont ajouté une consonne double alors qu'elle n'était pas attendue. Cette possibilité de doubler certaines consonnes est connue des enquêtés en difficulté avec l'écrit ; elle vient même perturber la graphie de mots très consistants comme TOMATE que l'on a observé avec les réponses avec *tommate*, *tomatte* et même *tommatte*. Les modalités de doublement de certaines consonnes

²¹¹ « SALADE VERTE » est issu de la dictée de la DEPP (Goigoux, 2016).

semblent également maîtrisées puisqu'aucune consonne double non attestée en français n'a été relevée dans les dictées. Aussi, tester les connaissances des enquêtés sur les mots qui contiennent des consonnes doubles ne nous semble pas prioritaire, et que ce type d'erreurs est peut-être moins stigmatisé que d'autres.

- le « -e » muet final ; cette difficulté étant emblématique de l'orthographe du français :
 - o dans des séries régulières en « -ie » : BOULANGERIE, ÉPICERIE ;
 - o « -e » muet final qui permet la prononciation de la consonne finale : CERISE ;
- des mots qui contiennent des schwas (ÉPICERIE), des finales muettes (CANARD, HARICOT, PUIS), des lettres historiques ou étymologiques (BIBLIOTHÈQUE).

LA MORPHOGRAPHIE

Nous avons vu dans les résultats des dictées, que les compétences en morphographie sont particulièrement discriminantes et que sa maîtrise (dans des configurations simples) semble liée aux compétences orthographiques, en général.

Les deux accords en nombre dans la dictée d'IVQ pour rendre compte des compétences en morphographie nous semblent insuffisantes, comme le montre les 9% des enquêtés « haut » qui n'avaient pas été en difficulté dans la dictée d'IVQ mais qui le sont dans la dictée complémentaire, sur le « rayon bricolage ».

Il nous semble par ailleurs plus intéressant d'ajouter tous les déterminants dans la dictée, pour que l'on puisse davantage observer l'accord nominal avec des déterminants simples (UN, LES, DES, LA) et des déterminants plus complexes (DEUX, DE L', DU, QUELQUES, DE, AUX). Ces déterminants devront être alternés entre singulier et pluriel.

Aussi, nous proposons pour améliorer la dictée d'IVQ d'ajouter :

- des marques de nombre sur les noms avec présence d'un signal (pluriel audible), en utilisant des déterminants simples (comme LES, DES) :
 - o noms : DES CACHETS, DES HARICOTS ;
 - o adjectifs : DES SALADES ;
- des structures complexes sans marqueur de pluriel ou avec des marques redondantes :
 - o nom et adjectifs : UNE TARTE AUX POMMES, QUELQUES BANDES DESSINÉES, QUATRE GROSSES TOMATES, UN PAQUET DE PÂTES ;
- des marques de genre : BANDE DESSINÉE, PEINTURE BLEUE ;
- des finales muettes issues de la morphologie dérivationnelle : PHARMACIE , FORT , TROIS , VERT ;

- des verbes à finales en /E/ (ELLE EST FERMÉE) qui permet d'opposer le passé composé à l'infinitif (SI LA BOULANGERIE EST FERMÉE, PASSER AU SUPERMARCHÉ) ;

Concernant la morphographie, il pourrait être envisagé une dictée qui comporterait plusieurs niveaux de difficulté, permettant de distinguer différents niveaux de compétence. Les difficultés morphographiques plus importantes (comme les structures complexes sans marqueur de pluriel, les marques redondantes ou les verbes à finales en /E/) ne seraient proposées qu'aux personnes qui ne montreraient pas de difficultés particulières en phonographie, après avoir testé cette compétence sur des mots d'abord consistants et fréquents, puis moins consistants

LA PHONOGRAPHIE À PARTIR DE PSEUDO-MOTS

Les résultats obtenus sur les pseudo-mots se sont révélés peu discriminants et nous questionnent sur la nécessité d'intégrer des pseudo-mots dans une dictée pour analyser les compétences d'ordre phonographique chez les scripteurs. Certains enquêtés « bas » ont montré de réelles difficultés dans ce domaine qui peuvent en effet être saisies dans l'analyse des graphies erronées des mots de la dictée. L'analyse des erreurs chez les enquêtés « haut » montre en revanche que ces derniers semblent être concernés par d'autres problématiques, d'origine lexicale et/ou morphographique, de sorte que leurs compétences en phonographie n'ont pas, selon nous, à être testées de manière ciblée, par des pseudo-mots. Si les pseudo-mots sont conservés, on pourrait envisager une dictée sous forme enregistrée et proposée dans de bonnes conditions d'audition (bruits extérieurs, qualité du son, etc.), avec des arrêts possibles et des segmentations de l'enregistrement pour permettre des réécoutes mot par mot.

La présence de pseudo-mots dans ce corpus reste intéressante pour analyser les choix que font les scripteurs par rapport à leur connaissance des contraintes du système d'écriture du français (analyse de la variation graphique, de la présence ou non d'une finale muette).

LA PHONOGRAPHIE À PARTIR DES MOTS

Les compétences en phonographie sont importantes à cerner en formation d'adulte, notamment chez les scripteurs qui ont une orthographe qui s'écarte de la norme. Cette compétence doit servir de point de départ, de « levier » pour améliorer les compétences orthographiques. Par ailleurs, nous avons vu que les enquêtés qui ont des graphies qui s'écartent de la norme mais qui répondent aux règles de phonographie du français ont des compétences communicationnelles : lorsqu'ils écrivent un message, celui-ci peut être compris. La remédiation pour eux, dans le domaine lexical porte alors sur l'accès à la norme orthographique.

Au contraire, ceux qui, en transcrivant phonétiquement les items, ne fournissent pas des graphies plausibles, soit parce qu'elles sont complètement éloignées de la forme attendue (*rme* pour Rhume) soit parce qu'elles ne respectent pas les règles de phonographie du français (*épiserie* pour ÉPICERIE), ont besoin d'une remédiation spécifique et prioritaire sur ce point.

Aussi, la correction des dictées devrait permettre d'appliquer deux niveaux de compétence sur le lexique : respect de la forme phonographique et respect de la forme orthographique.

RECOMMANDATIONS DIVERSES

Nous listons à présent quelques recommandations générales, qui proviennent d'observations que nous avons faites durant l'exploitation des dictées :

- Problèmes liés à la typographie : ne pas mettre de « e » accentué en début de phrase, au risque que l'enquêté mette une majuscule non accentuée qui ne signifie pas nécessairement qu'il s'agit de « e » (EPICERIE = *epicerie* ou *épicerie* ?) ;
- Organisation matérielle :
 - une dictée enregistrée, qui permette autant de réécoutes que souhaité par l'enquêté (pour éviter les différences de prononciation entre les enquêtés, les biais, les oublis) ;
 - une feuille qui contienne une marge (pour encadrer les graphies) et un format de liste déjà matérialisé avec des consignes qui indiquent à l'enquêté de revenir à la ligne après chaque syntagme. Il nous semble que la liberté laissée à l'enquêté sur l'organisation « spatiale » de la dictée n'apporte pas d'informations exploitables et qu'elle pourrait même gêner le scripteur (notamment ceux pour qui le geste graphique semble déjà difficile, et pour qui l'organisation de l'espace sur une feuille semble également représenter une difficulté supplémentaire).

... AUX PISTES DIDACTIQUES

La connaissance du système orthographique du français qui s'est développée grâce aux travaux des chercheurs cités tout au long de ce travail est essentielle pour la formation des enseignants : les apports de la linguistique avec la notion de plurisystème de Catach et al. (1980), les analyses de Blanche-Benveniste et Chervel (1969) mais aussi les travaux récents issus de la psycholinguistique sont désormais un point incontournable de la formation pour l'enseignement/ apprentissage de l'orthographe à l'école primaire, au collège, et, depuis la rentrée 2019, au lycée. À ce sujet, les récents travaux de Lété (2018) en psycholinguistique apportent de nouvelles pistes pour l'enseignement/apprentissage qui pourront être

développées en didactique. En ce qui concerne la grammaire, il est admis que son acquisition ne peut se faire que par l'intermédiaire de son apprentissage en contexte scolaire car l'orthographe du français se caractérise en particulier par une morphologie silencieuse, qui en fait sa principale difficulté. Plus spécifiquement, les apports de la recherche dans le domaine de l'orthographe sont essentiels pour la formation des enseignants du premier degré : pour permettre aux futurs enseignants de comprendre le code écrit et favoriser l'entrée dans l'écrit des enfants (conscience phonologique, découverte du principe alphabétique, travail sur les correspondances entre phonèmes et graphèmes etc.) puis le développement de leurs compétences.

Brissaud (2011) au terme de son état des lieux publié dans la revue *Pratiques* sous le titre « *Didactique de l'orthographe : avancées ou piétinements ?* » montre que la didactique du français est un domaine qui a beaucoup évolué depuis les années 1980. Les travaux issus de la recherche en psycholinguistes ont quant à eux apporté des connaissances sur les procédures d'acquisition par les élèves, le temps que prend l'acquisition, les difficultés liées à la gestion de l'orthographe notamment de la morphologie flexionnelle en situation de production écrite. L'apport de ces connaissances sur les pratiques d'enseignement est considérable car elles ont permis de développer de nouveaux dispositifs d'apprentissage. Citons en particulier l'intérêt d'un enseignement explicite de l'orthographe lexicale pour l'acquisition du lexique comme décrit par Fayol, Grimaud et Jacquier (2013) ou la mise en place d'un apprentissage réflexif des concepts grammaticaux. Lors de différentes situations d'apprentissage, les élèves sont conduits à la réflexion et à la manipulation tout en utilisant un métalangage grammatical rigoureux, qui montre bien que les élèves sont capables de manipuler des concepts complexes de la langue (Brissaud, Pasa, Ragano et Totereau, 2016 ; Fisher et Nadeau, 2014) alors que les discours alarmistes sur les compétences orthographiques des élèves semblent dire le contraire.

La diffusion des travaux savants sur le système orthographique passe par les grammaires et par les travaux des didacticiens du français. La *Grammaire méthodique du français* de Riegel, Pellat et Rioul (1994 pour la 1ère édition) sert de référence dans les cursus de Lettres et de Linguistique ainsi que dans la formation des enseignants. Cette grammaire intègre tout un chapitre sur l'orthographe française. Il en est de même dans l'ouvrage récent *Le Grevisse de l'enseignant*, de Pellat et Fonvielle (2016), qui adopte la même perspective d'être une grammaire d'orientation linguistique, mais qui se veut plus synthétique et clairement orientée vers le monde scolaire. L'intérêt est de proposer une grammaire au sens large, pas seulement dédiée à l'enseignement grammatical au sens strict, et de faire évoluer la tradition scolaire par la prise en compte d'apports récents et rigoureux sur l'étude de la langue.

Les travaux des didacticiens contribuent eux aussi à une meilleure connaissance du système orthographique du français, des difficultés des élèves dans l'appropriation de ce système complexe et également du rapport des élèves à l'orthographe, source d'insécurité comme nous l'avons vu dans la partie théorique. Nous pouvons citer les travaux de Brissaud et Cogis (2011), ou encore la parution récente de l'ouvrage dirigé par Mortamet, *L'orthographe. Pratiques d'élèves, pratiques d'enseignants, représentations* (2019).

La question de l'orthographe est aussi abordée dans d'autres domaines de la didactique, notamment en didactique de la grammaire car il est bien admis aujourd'hui que la grammaire est pour une part au service des apprentissages orthographiques, comme l'avait montré Chervel (2008). Les travaux présentés lors de la récente conférence de Consensus sur le thème « *Écrire et rédiger : comment guider les élèves dans leurs apprentissages ?* » (CNESCO²¹², 2018), soulignent également que les pratiques de classe sont importantes à prendre en compte ainsi que les pratiques de formation pour comprendre ce qui va permettre de former des enseignants qui aient des gestes professionnels efficaces, ajustés aux besoins de leurs élèves. Les nombreuses contributions présentées lors de cette conférence montrent que la recherche est particulièrement active dans ce domaine.

Une des solutions, qui représente un chantier ouvert depuis la fin du XIX^{ème} siècle consisterait à simplifier l'orthographe, dans la lignée de ce qui a été amorcé en 1990, pour qu'elle soit davantage fonctionnelle et accessible à tous. Pour cela, Fayol et Jaffré (2014) soulignent qu'il faudrait également que la formation des enseignants tienne compte des connaissances de la recherche dans le domaine de l'acquisition de l'orthographe et de son histoire. La psycholinguistique a plus récemment permis de souligner que la complexité orthographique ne s'arrête pas au mot mais qu'elle surgit surtout lorsqu'il s'agit de produire du texte. Ainsi, même si les scripteurs disposent de connaissances suffisantes sur les formes orthographiques et sur les procédures d'accord, la production de textes induit l'apparition d'erreurs alors que les savoirs et savoir-faire requis sont pourtant disponibles. Il est donc désormais admis que la maîtrise orthographique n'est pas qu'une question de compétences. Plusieurs travaux dont ceux de Fayol et Largy (1992) ont montré qu'il faut distinguer les connaissances et la mise en œuvre de ces connaissances. Les élèves, comme les adultes, ne peuvent pas se concentrer sur les tâches qui sont requises lors d'une activité d'écriture et il peut apparaître des erreurs d'orthographe alors même que le scripteur possède les connaissances, sait les mettre en œuvre et est également capable de les réviser. Il convient alors d'apprendre aux élèves, en

²¹² CNESCO : Conseil national d'évaluation du système scolaire.

situation d'apprentissage, à comprendre ce processus (lié à l'attention et à la mémoire de travail) et à leur fournir des outils pour en tenir compte. La révision orthographique, souvent nécessaire dans les constructions syntaxiques complexes où il est indispensable de réfléchir pour procéder aux accords serait plus efficace si les scripteurs savaient repérer les configurations à risque : celles où la survenue d'erreurs est davantage propice, et qui peut être propre à chacun, mais qui est largement liée aussi aux difficultés inhérentes au système orthographique.

Malgré l'importance des travaux réalisés dans ce domaine, les idées reçues sur cette question sont tenaces, notamment en orthographe grammaticale. Les discours sur l'origine des erreurs orthographiques qui incriminent le scripteur lui-même plutôt que le système subsistent aujourd'hui encore.

Enfin, de nombreux auteurs, depuis longtemps (Catach ; Blanche-Benveniste et Chervel) et plus récemment (Jaffré, Brissaud, Pellat, etc.) montrent que l'enseignement de l'orthographe et notamment de l'orthographe grammaticale est à réinventer. Fayol et Jaffré (2014) parlent d'une nécessité de rationaliser de l'enseignement de l'orthographe : c'est-à-dire améliorer les apprentissages notamment grâce à de nouvelles manières de l'enseigner tout en n'y consacrant pas plus de temps. En effet, la ré-augmentation du nombre d'heures au détriment d'autres disciplines ne semble pas aujourd'hui envisageable. L'école a aujourd'hui de nombreuses missions pour former le futur citoyen en accord avec les demandes de la société : maîtrise des langues étrangères, des nouvelles technologies de l'information et de la communication, etc., ce que précisent les nouveaux programmes de 2015 pour l'école primaire (BO du 23 avril 2015), même si la maîtrise de la langue française est inscrite comme « la priorité » du cycle 2 et un « objectif central » du cycle 3.

De nombreux didacticiens se sont penchés sur des pédagogies et des activités innovantes en orthographe, notamment pour développer la réflexion métalinguistique et favoriser l'appropriation des concepts grammaticaux en fournissant aux élèves des outils comme la manipulation syntaxique comme aide à l'analyse grammaticale. En somme, il s'agit d'aider les élèves à s'approprier la grammaire de la langue en favorisant les activités qui permettent de manipuler la phrase et donc d'envisager la grammaire de manière réflexive. D'autres travaux réalisés au Québec montrent que dispositifs d'apprentissage innovants profitent particulièrement aux élèves les plus en difficulté, notamment lorsqu'ils se trouvent dans une situation où la perspective est la résolution de problème et lorsqu'on leur donne la possibilité de justifier leurs choix orthographiques et de parler de leurs erreurs. Les discours métagraphiques sont également d'une grande aide pour les enseignants, car ils peuvent faire

émerger des représentations qui sont parfois des obstacles pour l'apprentissage (Fisher et Nadeau, 2014 ; Nadeau et Fisher, 2012). De nombreux auteurs se sont engagés dans cette voie. Citons par exemple les propositions de Fisher et Nadeau (2012, 2014), Cogis, Fisher, Nadeau (2015) ou même des propositions de pédagogues comme Maurice Laurent la grammaire en couleurs. Nous ne présenterons pas ici les ressources qui existent et qui se développent dans cette voie, mais nous soulignons leur importance pour la formation des élèves ainsi que des enseignants.

Il faut bien entendu tenir compte de la spécificité d'un public adulte, mais il nous semble que les principes et dispositifs qui font leur preuve par rapport au développement du raisonnement orthographique chez les élèves peuvent être mis à profit chez des adultes aussi. Le fait de travailler le raisonnement orthographique à partir de situations-problèmes est un levier pour améliorer les productions mais également pour favoriser un meilleur rapport à l'écrit.

Pour conclure, il nous semble important de redire, dans la lignée des chercheurs et notamment dans la tradition du laboratoire LIDILEM (université Grenoble Alpes) au sein duquel nous avons effectué ce travail, qu'il convient de tout faire pour que la maîtrise de l'orthographe ne soit plus une compétence aussi discriminante au quotidien. Pour cela, il est indispensable que le regard porté par la société évolue, pour que progressivement, les discours conservateurs de l'orthographe non fondés mais encore tenaces, s'assouplissent et que l'orthographe ne soit plus un objet de tensions aussi vives.

TABLE DES FIGURES

Figure 1. Architecture de l'enquête IVQ 2011 (INSEE, 2011)	67
Figure 2. Capture d'écran de FROMAGES, Consistance OP (Manulex-Infra).....	140
Figure 3. Capture d'écran de POINT, Consistance OP (Manulex-Infra).....	142
Figure 4. Capture d'écran de PHARMACIE, Consistance OP (Manulex-Infra)	142
Figure 5. Classement des mots d'IVQ selon leur pourcentage de réussite dans EOLE, chez des élèves de CM2.	184
Figure 6. Classement des mots du « rayon bricolage » selon leur pourcentage de réussite dans EOLE, chez des élèves de CM2.	216
Figure 7 : Répartition des 447 enquêtés « bas » du quart Sud-Est par type de difficulté, sur l'intégralité du module « bas », selon l'INSEE	238
Figure 8. Répartition des scores (sur 24) obtenus à la dictée d'IVQ par les 447 enquêtés « bas ».	243
Figure 9. Répartition des 447 enquêtés « bas » du quart Sud-Est par type de difficulté, sur la dictée.	248
Figure 10. Compétences en orthographe lexicale des 441 enquêtés « bas » sur 16 mots de la dictée d'IVQ.....	249
Figure 11. Comparaison des scores de réussite des mots de la dictée d'IVQ, par les 441 enquêtés "bas" et par des élèves de CM2 (EOLE).....	252
Figure 12. Proportion de réponses phonographiquement acceptables et approximantes obtenues par les 441 enquêtés « bas » pour chaque pseudo-mot.	254
Figure 13. Classement des mots de la dictée d'IVQ chez les 441 enquêtés « bas » selon le type de réponse : orthographe conforme, phonographie acceptable, phonographie approximante, graphies non conformes.	271
Figure 14. Réussite en morphographie flexionnelle du nombre observée sur trois items chez les 441 enquêtés « bas ».	286
Figure 15. Répartition des scores obtenus sur la dictée d'IVQ par les 157 enquêtés « bas » qui ont marqué le nombre sur FROMAGES et sur SENTENT.	287
Figure 16. Comparaison du taux de réussite de 6 mots entre les 215 enquêtés « bas » en difficulté dans la dictée (score compris entre 35% et 80%) et les 176 enquêtés « bas » qui ne sont pas en difficulté (plus de 80% de réussite).	300
Figure 17. Scores moyens des 447 enquêtés « bas » obtenus en orthographe, en phonographie (sur les pseudo-mots) et en morphographie, répartis selon 5 profils de compétences.....	320

Figure 18. Répartition des scores (sur 24) obtenus à la dictée d'IVQ par les 447 enquêtés « bas » et les 87 enquêtés « haut ».....	330
Figure 19. Compétences en orthographe lexicale des 441 enquêtés « bas » et des 87 enquêtés « haut » sur 16 mots de la dictée d'IVQ.	332
Figure 20. Comparaison de la proportion de réponses acceptables et approximantes obtenues par les 441 enquêtés « bas » et les 87 enquêtés « haut » pour chaque pseudo-mot.....	334
Figure 21. Répartition des réponses sur les 7 mots qui ont obtenus entre 52 % et 93% de réussite orthographique chez les 87 enquêtés « haut ».....	337
Figure 22. Comparaison du taux de réussite en morphographie chez les 441 enquêtés « bas » et les 87 « haut ».....	340
Figure 23. Répartition des scores obtenus par les 71 enquêtés « haut » qui ont passé l'intégralité de la dictée complémentaire (rayon bricolage).	357
Figure 24. Compétences lexicales des 87 enquêtés « haut » sur 21 mots du rayon bricolage.	358
Figure 25. Compétences en morphographie des 87 enquêtés « haut » sur l'intégralité de la dictée.	361
Figure 26. Profil en morphographie des 71 enquêtés « haut ».....	363

TABLE DES TABLEAUX

Tableau 1. Taux de participation des enquêtés dans les épreuves du module « bas ».....	76
Tableau 2. Répartition des archigraphèmes « s » et « z » en trois niveaux selon Catach et al. (1980).....	113
Tableau 3. Classement des mots de la dictée IVQ et de certaines graphies produites par les enquêtés selon leur degré de consistance (consistance PO totale, du CP au CM2, calcul par token) et leur indice de fréquence, calculés dans Manulex-Infra.	179
Tableau 4. Classement des mots de la dictée IVQ selon leur indice de fréquence calculé dans la base LEXIQUE 3.	182
Tableau 5. Pourcentages de réussite de mots de la dictée d'IVQ d'après EOLE.	183
Tableau 6. Relevé des graphies considérées comme orthographiquement conformes (16 mots).....	196
Tableau 7. Graphies considérées comme phonographiquement acceptables chez les 441 scripteurs « bas ».....	200
Tableau 8. Graphies considérées comme phonographiquement approximantes chez les 441 scripteurs « bas ».....	202
Tableau 9. Graphies des pseudo-mots considérées comme phonographiquement acceptables chez les 441 scripteurs « bas ».....	204
Tableau 10. Classement des mots de la dictée d'IVQ et du « rayon bricolage » selon leur degré de consistance (consistance PO totale, du CP au CM2, calcul par token) calculé dans Manulex-Infra-Infra.	211
tableau 11. Classement des mots du « rayon bricolage » selon leur indice de fréquence calculé dans la base lexicque 3.	213
Tableau 12. Pourcentages de réussite de mots du "rayon bricolage" d'après EOLE.	215
Tableau 13. Relevé des graphies considérées comme orthographiquement correctes dans la score lexical (16 mots).	220
Tableau 14. Nombre de réponses, d'absences de réponses et de graphies illisibles recueillies pour chaque mot et pseudo-mot de la dictée IVQ dans les 441 dictées produites par les enquêtés « bas ».	241
Tableau 15. Détail de la variation sur les pseudo-mots chez les 441 enquêtés « bas ».	259
Tableau 16. Relevé de 20 graphies de chaque pseudo-mot considérées comme phonographiquement non acceptables chez les 441 enquêtés « bas ».	260

Tableau 17. Relevé des graphies considérées comme phonographiquement acceptables pour chaque pseudo-mot chez les 441 enquêtés « bas ».	262
Tableau 18. Relevé des graphies orthographiquement non conformes du mot ÉPICERIE chez les 441 enquêtés « bas », classées selon la restitution du phonème /s/ et selon le nombre d'erreurs phonographiques.	275
Tableau 19. Relevé des graphies orthographiquement non conformes du mot LIBRAIRIE chez les 441 enquêtés « bas » classées selon le nombre d'erreurs phonographiques : phonographie acceptable, 1 erreur phonographique, 2 erreurs phonographiques ou plus.	278
Tableau 20. Relevé des 40 graphies de LIBRAIRIE qui contiennent plus d'une erreur phonographique chez les 441 enquêtés « bas ».	281
Tableau 21. Proportion de réussite observée sur 16 mots en fonction du type de critère : orthographique ou phonographique et du niveau de réussite attendu chez les 441 enquêtés « bas ».	284
Tableau 22. Détail des graphies qui présentent une marque de genre ou de nombre sur FIANCÉ et SOLENNEL chez les 441 enquêtés « bas ».	291
Tableau 23. Détail des erreurs produites chez les 53 scripteurs en difficulté sur la dictée (score compris entre 9 et 19 sur 24) et ayant fait au moins une erreur sur la sélection des 3 mots les plus consistants.	303
Tableau 24. Proportion de graphies qui contiennent un « -e » final parmi les graphies phonographiquement acceptables et les graphies phonographiquement non acceptables/approximantes sur 3 mots qui contiennent une marque de dérivation chez les 441 enquêtés « bas ».	311
Tableau 25. Proportion de graphies qui contiennent un « -e » final parmi les graphies phonographiquement acceptables et les graphies phonographiquement non acceptables/approximantes sur 5 mots qui contiennent un « -e » final chez les 441 enquêtés « bas ».	314
Tableau 26. Relevé de la variation présente sur les mots, relevé d'après les réponses phonographiquement acceptables, approximantes et phonographiquement non acceptables chez les 441 enquêtés « bas » et les 87 enquêtés « haut ».	345
Tableau 27. Relevé de la variation sur les pseudo-mots dont la forme est considérée comme phonographiquement acceptable chez les 441 enquêtés « bas » et les 87 enquêtés « haut ».	347
Tableau 28. Relevé des formes les plus récurrentes des pseudo-mots parmi les variantes phonographiquement acceptables chez les 441 enquêtés « bas » et les 87 enquêtés « haut ».	348

Tableau 29. Proportion de « -e » final muet sur MICATOL et SIRAPE, calculée pour chacun des 4 groupes de scripteurs.	352
Tableau 30. Proportion de finale muette sur GOBAR et PÉVANOR, calculée pour chacun des 4 groupes de scripteurs.	354
Tableau 31. Relevé des graphies non conformes à la norme orthographique pour les 5 mots les moins bien restitués dans la dictée complémentaire chez les 87 enquêtés « haut ».	359

BIBLIOGRAPHIE

- Andreu, S., Steinmetz, C. (2016). Les performances en orthographe des élèves en fin d'école primaire (1987-2007-2015). *Note d'information n°28, MEN-DEPP*.
https://cache.media.education.gouv.fr/file/2016/07/5/depp-ni-2016-28-performances-orthographe-eleves-fin-ecole-primaire_658075.pdf
- ANLCI. Lutter ensemble contre l'illettrisme. Cadre national de référence [en ligne], consultée le 30 novembre 2015, URL <<http://www.anlci.gouv.fr/Mediatheque/Cadre-national-de-reference-sept-2003>>
- Barré-De-Miniac, C., Lété, B. (1997). Comprendre la complexité des situations d'illettrisme : des nécessaires interactions entre les théories et les pratiques. In *L'illettrisme, de la prévention chez l'enfant aux stratégies de formation chez l'adulte*, (7-18). De Boeck Université.
- Bastyns, C. (2012). Les enquêtes de l'OCDE sur la littéracie et les compétences des adultes font-elles avancer la cause de l'alpha ? *Journal de l'alpha*, (185) : 26-39.
- Béguin, C., Saint-Laurent, L., Giasson, J. (2010). Le traitement morphologique dans l'écriture chez les élèves de 6^{ème} au primaire. *L'année psychologique*, (110) : 275-297.
- Bentolila, A. (1996). De l'illettrisme en général, de l'école en particulier. Paris : Plon.
- Besse, J-M. (1995). *L'Écrit, l'École et l'Illettrisme*. Paris : Magnard.
- Besse, J-M. (1997). Le rapport à l'écrit chez des adultes illettrés. In *L'illettrisme, de la prévention chez l'enfant aux stratégies de formation chez l'adulte*, (263-278). De Boeck Université.
- Besse, J-M., Guérin-Pace, F. (2002). Une évaluation des compétences sur l'écrit : l'enquête « Information et vie quotidienne ». *Économie et Humanisme*, (363), n° spécial Vaincre l'illettrisme : 17-21.
- Besse, J-M., Petiot-Poirson, K., Petit Charles, E. (2003). *Qui est illettré ?* Paris : Retz.
- Besse, J-M., Luis, M-H., Dumoulin-Paire, K., Petiot-Poirson, K., Petit Charles, E. (2004). *Évaluer les illettrismes. Diagnostic des modes d'appropriation de l'écrit : guide pratique*. Paris : Retz.

- Besse, J-M., Luis, M-H., Bouchut, A-L., Martinez, F. (2009). La mesure des compétences en traitement de l'écrit chez des adultes en grande difficulté. *Économie et statistique*, (424-425) : 31-48.
- Bessin, M. (1999). L'armée et l'illettrisme : remarques sur la fonction de dépistage de la conscription militaire. *Débats/jeunesses*, coll. Agora, (15) : 57-67.
- Blanche-Benveniste, C., Chervel, A. (1969). *L'orthographe*. Paris : Maspéro.
- Blondel, C. (2011). *Les étudiants entrant à l'université : analyse des besoins en français écrit*. Mémoire de Master. Grenoble : Université Stendhal, Grenoble 3.
- Blondel, C. (2012). Quelle formation en grammaire et en orthographe pour des étudiants de licence : de l'étude des besoins aux propositions didactiques. Mémoire de Master. Grenoble : Université Stendhal, Grenoble 3.
- Blondel, C., Conseil, J. (2015). La dictée dans l'enquête IVQ : variation et compétences d'écriture des scripteurs en difficulté avec l'écrit dans les régions Haute-Normandie et Rhône-Alpes. In Brissaud, C., Mortamet, C. : La dictée, une pratique sociale emblématique, *Glottopol*, (26) : 207-225, <<http://glottopol.univ-rouen.fr>>
- Blondel, C., Brissaud, C., Conseil, J., Jeantheau, J-P., Mortamet, C. (2016). Les difficultés des adultes face à l'écrit : l'apport de l'épreuve de production écrite de l'enquête *Information et Vie Quotidienne*. *Économie et Statistique*, (490) : 1-24.
- Blondel, C., Brissaud, C., Rinck, F. (2016). Description des pratiques orthographiques de scripteurs ordinaires et de scripteurs en difficulté avec l'écrit à partir de l'analyse d'une dictée. *Congrès Mondial de Linguistique Française*, <https://www.shsconferences.org/articles/shsconf/pdf/2016/05/shsconf_cmlf2016_07003.pdf>
- Blondel, C., Brissaud, C., Rinck, F. (à paraître). Apports de deux enquêtes sur les connaissances orthographiques et les pratiques de l'écrit des Rhônalpins. *Délégation générale à la langue française et aux langues de France (DGLFLF)*, 29 pages.
- Blum, A., Guérin-Pace, F. (2000). Des lettres et des chiffres - Des tests d'intelligence à l'évaluation du « savoir lire », un siècle de polémiques. Paris : Fayard.
- Boch, F., Frier, C. (2015). *Écrire dans l'enseignement supérieur : des apports de la recherche aux outils pédagogiques*. Grenoble : ELLUG.

- Brendler, J., Sueur, C., Belmans, C., Folléa, G., Mercier, J-P., Thibaudeau, P. (2013). Les Haut-Normands face aux savoirs de base : lire – écrire – compter. Dossier d'Aval, <http://www.epsilon.insee.fr/jspui/bitstream/1/19076/3/dosaval_IVQ.pdf>
- Brissaud, C. (1999). La réalisation de l'accord du participe passé employé avec avoir. De l'influence de quelques variables linguistiques et sociales. *Langage et société*, (88) : 5-24.
- Brissaud, C. (2011). Didactique de l'orthographe : avancées ou piétinements ? *Pratiques*, (149/150) : 207-226.
- Brissaud, C. (2015). Quels enseignements tirer de quatre-vingt-dix-huit dictées de troisième : du décalage entre prescription et acquisition des élèves. *Le français aujourd'hui*, (190), 61-70.
- Brissaud, C, Bessonnat, D. (2001). *L'orthographe au collège : pour une autre approche*. Grenoble : CRDP.
- Brissaud, C., Chevrot, J.P, Lefrançois, P. (2006). Les formes verbales homophones en /E/ entre 8 et 15 ans : contraintes et conflits dans la construction des savoirs sur une difficulté orthographique majeure du français. *Langue française*, (151) : 74-93.
- Brissaud, C., Cogis, D. (2011). *Comment enseigner l'orthographe aujourd'hui ?* Paris : Hatier.
- Brissaud, C., Pasa, L., Ragano, S., Totereau, C. (2016). Effets des pratiques d'enseignement de l'écriture en cours préparatoire. *Revue française de pédagogie*, (196), 85-100.
- Brissaud, C. et Fayol, M. (2018). *Étude de la langue et production d'écrit*, conférence de consensus « Écrire et rédiger », Cnesco, <<http://www.cnesco.fr/wp-content/uploads/2018/04/VDEF-rapport-Brissaud-Fayol.pdf>>
- Brunot, F. (1905). La réforme de l'orthographe. Lettre ouverte à Monsieur le Ministre de l'Instruction publique. Paris : Armand Colin.
- Buisson, F. (1911). Nouveau dictionnaire de pédagogie et d'instruction primaire de Ferdinand Buisson, édition électronique : <<http://www.inrp.fr/edition-electronique/lodel/dictionnaire-ferdinand-buisson/document.php?id=2888>>
- Brunot, F. (1930). *Histoire de la langue française, des origines à 1900*. Paris : Armand Colin.
- Catach, N. (1973). Table ronde sur la structure de l'orthographe française : compte rendu. *Langue française*, (20) : 6-10.
- Catach, N. (1979). Le graphème. *Pratiques*, (25) : 21-32.

- Catach, N. (1980). L'enseignement de l'orthographe. Paris : Nathan.
- Catach, N. (1986). The grapheme : Its position and its degree of autonomy with respect to the system of the language. In Gerhard August (Éds), *New Trends in graphemics and orthography*, (1-10). Berlin : De Gruyter.
- Catach, N. (1995). Dictionnaire historique de l'orthographe française. Paris : Larousse.
- Catach, N., Gruaz, C., Duprez, D. (1980). *L'orthographe française*, (Nouvelle présentation, 2016). Paris : Nathan Université.
- Chervel, A. (1977). Et il fallut apprendre à écrire à tous les petits Français : histoire de la grammaire scolaire, rééd. 1981. Paris : Payot.
- Chervel, A. (1991). L'école républicaine et la réforme de l'orthographe. *Mots : Orthographe et société*, (28) :35-55.
- Chervel, A. (2008). L'orthographe en crise à l'école : et si l'histoire montrait le chemin ? Paris : Retz.
- Chervel, A., Manesse, D. (1989). La dictée : les Français et l'orthographe, 1873-1987. Paris : INRP.
- Cogis, D. (2005). Pour enseigner et apprendre l'orthographe. Paris : Delagrave.
- Cogis, D., Fisher, C. et Nadeau, M. (2015). Quand la dictée devient un dispositif d'apprentissage. *Glottopol*, (26) : 69-91.
- Conseil, J. (2017). Situations d'illettrisme et difficultés à l'écrit en Haute-Normandie : le cas de l'orthographe. Thèse de doctorat. Université de Rouen, Rouen.
- Content, A., Mousty, P., Radeau, M. (1990). BRULEX : une base de données lexicales informatisée pour le français écrit et parlé. *L'Année Psychologique*, (90) : 551-566.
- Cousin, M-P., Thibault, M-P., Largy, P., Fayol, M. (2006). Apprentissage de la morphologie flexionnelle du nombre nominal : étude de la récupération d'instances chez des enfants tout-venant et des enfants présentant un trouble de l'apprentissage de l'écrit. *Rééducation orthophonique*, (225) : 93-109.
- David, J., Guyon, O., Brissaud, C. (2006). Apprendre à orthographier les verbes : le cas de l'homophonie des finales en /E/. *Langue française*, (151) : 109-126.
- Dabène, M. (1987). L'adulte et l'écriture : contribution à une didactique de l'écrit en langue maternelle. Bruxelles : De Boeck Université.

- Dabène, M. (2000). Le monde de l'écrit : pratiques et représentations. In J-L. Poueyto (dir.), *Illettrismes et cultures*, (45-54). Paris : L'Harmattan.
- Degorre, A., Murat, F. (2009). La mesure des compétences des adultes, un nouvel enjeu pour la statistique publique. *Économie et statistique*, (424-425) : 3-30.
- Dehaene, S. (2007). *Les neurones de la lecture*. Paris : Odile Jacob.
- DEPP, Ministère de l'Éducation Nationale et de la jeunesse. Journée défense et citoyenneté 2018 : plus d'un jeune Français sur dix en difficulté de lecture. *Note d'information*, (19-20), juin 2019 : 1-4, <https://cache.media.education.gouv.fr/file/2019/74/9/depp-ni-2019-19-20-Journee-defense-et-citoyennete-2018-plus-de-un-jeune-Francais-sur-dix-en-difficulte-de-lecture_1135749.pdf>
- Dickes, P., Vrignaud, P. (1995). Rapport sur les traitements des données françaises de l'enquête internationale sur la littéracie. *Rapport pour le ministère de l'Éducation nationale*, Paris : Ministère de l'Éducation Nationale.
- Eme, E. (2010). Cognitive and psycholinguistics skills of adults who are functionally illiterate : Current state of research and implications for adult education. *Applied Cognitive Psychology*, (24-5) : (753–762).
- Espérandieu, V., Lion, A., Bénichou, J-P. (1984). *Des illettrés en France : rapport au Premier Ministre*. Paris : La Documentation Française.
- Fayol, M. (2013). *L'acquisition de l'écrit*. Paris : PUF (collection Que sais-je ?).
- Fayol, M., Jaffré, J-P. (2008). *Orthographier*. Paris : Presses universitaires de France.
- Fayol, M., Jaffré, J-P. (2014). *L'orthographe*. Paris : Presses universitaires de France (collection Que sais-je ?).
- Fayol, M., Jaffré, J-P. (2016). L'orthographe : des systèmes aux usages. *Pratiques [en ligne]*, (91-2) : 169-170. URL <<http://journals.openedition.org/pratiques/2984>> ; DOI : 10.4000/pratiques.2984.
- Fayol, M., Got, C. (1991). Automatisation et contrôle dans la production écrite : les erreurs d'accord sujet verbe chez l'enfant et l'adulte. *L'Année psychologique*, (91-2) : 187-205.
- Fayol, M., Grimaud, F., Jacquier, M. (2013). Une expérience d'enseignement explicite de l'orthographe lexicale. *Approche Neuropsychologique des Apprentissages chez l'Enfant*, 31, 22-29.

- Fayol, M., Largy, P. (1992). Une approche fonctionnelle de l'orthographe grammaticale. *Langue française*, numéro thématique, L'orthographe : perspectives linguistiques et psycholinguistiques, (95) : 80-98.
- Fayol, M., Miret, A. (2005). Écrire, orthographier et rédiger des textes. *Psychologie française*, (50) : 391-402.
- Fijalkow, J. (1986). *Mauvais lecteurs, pourquoi ?* Paris : PUF.
- Frier, C., (1992). Les représentations sociales de l'illettrisme. In J-M.Besse, M-M. De Gaulmyn, D. Ginet, B. Lahire (Éds.), *L'« illettrisme » en questions*, (47-57). Lyon : PUL.
- Frier, C. (1997). Portraits d'illettrés : au-delà des discours, quel message pour quelle société ? In *L'illettrisme, de la prévention chez l'enfant aux stratégies de formation chez l'adulte*, (37-54). De Boeck Université.
- Gak, V-G. (1976). *L'orthographe du français : essai de description théorique et pratique*. Paris : Société d'études linguistiques et anthropologiques de France.
- Garcia-Debanc, C. (2012). Genres et normes dans l'enseignement de la production écrite à l'école primaire de 1880 à 2007. In Grinshpun, Y., et Nyée-Doggen, J. (Éds.), *Regards croisés sur la langue française : usages, pratiques, histoire*, (183-194). Paris : Presses Sorbonne Nouvelle. Tiré de <<http://books.openedition.org/psn/3324>>
- Geoffre, T., Brissaud, C. (2012). L'accord sujet-verbe : acquis en fin d'école primaire, vraiment ? 3^{ème} Congrès Mondial de Linguistique Française, <https://www.shs-conferences.org/articles/shsconf/pdf/2012/01/shsconf_cmlf12_000196.pdf>
- Geoffre, T., Totereau, C., Brissaud, C. (2019). Évolution des performances orthographiques du début du CP à la fin du CE1 : une étude longitudinale. In Kervyn, B., Dreyfus, M., Brissaud, C., *L'écriture dès le début de l'école primaire*. Pratiques enseignantes et performances des élèves, coll. « Études sur l'Éducation », (185-200). Pessac : Presses Universitaires de Bordeaux.
- Guérin-Pace, F. (2009). Illettrismes et parcours individuels. *Économie et statistique*, (424-425) : 49-62.
- Goigoux, R. (2016). *Lire et écrire au CP : étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages*. Rapport de recherche, Institut français de l'Éducation, Université de Lyon, <<http://ife.ens-lyon.fr/ife.ens-lyon.fr/ife/recherche/lire-ecrire/rapport/rapport-lire-et-ecrire>>

- Hauser, J. (1988). Chapitre d'introduction. *Hommes et Migrations*, dossier Illettrisme, (1117) : 12-13.
- Holmes, V. (1996). Skilled reading and orthographic processing. *Australian Journal of Psychology*, (48) : 149-154.
- Irwin, S., Kirsch, Jungeblut, A. (1986). Literacy : Profiles of America's Young Adults, Educational Testing Service, (800), 223-0267. Educational Testing Service : Princeton, NJ.
- Jaffré, J-P. (1992). *Didactique de l'orthographe*. Paris : Hachette.
- Jaffré, J-P. (2000). Écriture et sémiographie. *Linx*, (43) : 15-28.
- Jaffré, J-P. (2003). La morphographie du français : un cas sémiographique. *Rééducation orthophonique*, (213) : 13-26.
- Jaffré, J-P. (2004). La littéracie : histoire d'un mot, effets d'un concept. In Barré-De Miniac, C., Brissaud, C., Rispaïl, M. *La littéracie. Conceptions théoriques et pratiques d'enseignement de la lecture-écriture* (21-41). Paris : L'Harmattan.
- Jaffré, J-P. (2005). L'orthographe est un château fort. Commentaires sur l'enquête du Collectif « Sauver les Lettres », [en ligne], consulté le 28 octobre 2019, <www.cafepedagogique.net/lesdossiers/Pages/jaffre_index.aspx>
- Jaffré, J-P. (2008). La mixité des orthographes. In M. Fayol et J-P. Jaffré (Éds.), *Orthographier*, (85-101). Paris : Presses Universitaires de France.
- Jaffré, J-P., Pellat, J-C. (2008). Sémiographie et orthographes : le cas du français. In C. Brissaud, J-P Jaffré, J-C. Pellat (Éds.), *Nouvelles recherches en orthographe*, actes des journées d'étude des 14 et 15 juin 2007, Université de Strasbourg, (9-30). Limoges : Lambert-Lucas.
- Jeantheau, J.P. (2010). L'orthographe dans les enquêtes sur l'illettrisme en France. In C. Gruaz, C. Jacquet-Pfau (Dir.), *Autours du mot : pratiques et compétences*, tome II, séminaire 2006-2008, (223-239), Limoges : Lambert-Lucas.
- Jeantheau, J.P. (2015). La dictée dans les enquêtes sur la « littéracie » des adultes : pratiques, résultats, exemples d'analyse, perspectives. *Glottopol*, (26) : 191-206.
- Jonas, N. (2012). Pour les générations les plus récentes, les difficultés des adultes diminuent à l'écrit, mais augmentent en calcul. *Insee Première*, (1426), <<https://www.insee.fr/fr/statistiques/1281410>>

- Kirsch, I-S., Jungeblut, A. (1986). Literacy : Profiles of America's Young Adult, final report. *NAEP report*, (16-PL-01), Educational Testing Service, Princeton, NJ.
- Lahire, B. (1997). Usages sociaux de l'écrit et « illettrisme ». In Andrieux, F., Besse, J-M., Falaize, *Illettrisme : quels chemins vers l'écrit ?*, (11-20). Paris : Magnard.
- Lahire, B. (1999). L'invention de l'illettrisme. Rhétorique publique, éthique et stigmates. Paris : La découverte.
- Largy, P. (2001). La révision des accords nominal et verbal chez l'enfant. *L'Année Psychologique*, (101) : 221-245.
- Largy, P., Fayol, M., Lemaire, P. (1996). The homophon effectis in written French : the case of verb-noun inflections errors. *Language and cognitive processes*, (11) : 217-255.
- Laurent, M. (2004). *Les jeunes, la langue, la grammaire : d'une grammaire implicite à une grammaire explicite. Tome 1, catégories de mots, constituants de la phrase*. Besançon : UEPD.
- Laurent, M. (2009). *Les jeunes, la langue, la grammaire : d'une grammaire implicite à une grammaire explicite. 2, Orthographe grammaticale, expression du temps, conjugaison*. Besançon : UEPD.
- Leclercq, V. (1997). Comment transmettre des savoirs sur la langue dans des actions de formation de base pour adultes. In *L'illettrisme, de la prévention chez l'enfant aux stratégies de formation chez l'adulte*, (279-296). De Boeck Université.
- Leclercq, V. (1999). Face à l'illettrisme, enseigner l'écrit à des adultes. Paris : ESF.
- Leclercq, V. (2008). Connaître les adultes peu scolarisés et peu qualifiés : des recherches pour comprendre et agir. *TransFormations, Recherches en éducation des adultes*, (1) : 3-18.
- Légifrance (2017). Journal Officiel du 10 janvier 2017, n°8, <https://www.legifrance.gouv.fr/jo_pdf.do?id=JORFCONT000033842856>
- Lété, B. (2004). Manulex-Infra : une base de données du lexique écrit adressée aux élèves. *Didactique du lexique : Contextes, démarches, supports*, (241-257). Louvain-la-Neuve : De Boeck Supérieur.
- Lété (2008). La consistance phonographique : une mesure statistique de la complexité orthographique. In Brissaud, C., Jaffré, J-P., Pellat, J-C. (Éds.), *Nouvelles recherches en orthographe*, actes des journées d'étude des 14 et 15 juin 2007, Université de Strasbourg, (85-99). Limoges : Lambert-Lucas.

- Lété, B. (2018). Quelles sont les relations entre lecture et écriture pour l'apprentissage du lexique ? Cnesco, Écrire et rédiger : comment guider les élèves dans leurs apprentissages, (70-78), Notes des experts, <http://www.cnesco.fr/wp-content/uploads/2018/04/180411_CCEcrire_notes_experts.pdf>
- Lété, B., Sprenger-Charolles, L., Colé, P. (2004). Manulex-Infra : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, (36) : 156-166.
- Lété, B., Peereman, R., Fayol, M. (2008). Consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study. *Journal of Memory and Language*, (58) : 952-977.
- Lucci, V., Nazé, Y. (1979). Enseigner ou supprimer l'orthographe. *Langue française, théorie et pratique* (1). Paris : CEDIC.
- Lucci, V., Millet, A. (1994). L'orthographe de tous les jours. Enquête sur les pratiques orthographiques des Français. Paris : Champion.
- Manesse, D., Cogis, D. (2007). *Orthographe : à qui la faute ?* Issy les Moulineaux : ESF.
- Martinet, A. (1960). *Éléments de linguistique générale*. Paris : Armand Colin.
- Millet, A. (1990). *Quelques aspects sociolinguistiques de l'orthographe française*. Thèse de doctorat. Université Stendhal Grenoble 3, Grenoble.
- Millet, A., Lucci, V., Billiez, J. (1990). *Orthographe mon amour*. Grenoble : Presses Universitaires de Grenoble.
- Ministère de la défense. *Journée défense et citoyenneté*, <<https://www.defense.gouv.fr/jdc/parcours-citoyennete/jdc>>
- Ministère de l'Éducation Nationale et de la jeunesse. Horaires et programmes d'enseignement de l'école primaire. *Bulletin officiel de 2002*, Hors-série n°1, 14 février, <<https://www.education.gouv.fr/bo/2002/hs1/default.htm>>
- Ministère de l'Éducation Nationale et de la jeunesse. Programme de l'enseignement commun de français en classe de seconde générale et technologique et en classe de première des séries générales et programme de l'enseignement de littérature en classe de première littéraire. *Bulletin officiel spécial n°9*, 30 septembre 2010, <<https://www.education.gouv.fr/cid53318/mene1019760a.html>>

- Ministère de l'Éducation Nationale et de la jeunesse. *Programme international pour le suivi des acquis des élèves (PISA, 2018)* : <<https://www.education.gouv.fr/pid37635/pisa-programme-international-pour-le-suivi-des-acquis-des-eleves.html>>
- Ministère de l'Éducation Nationale et de la jeunesse. Nouveaux programmes du lycée général et technologique à la rentrée 2019. *Bulletin officiel spécial n°8*, 25 juillet 2019, <https://www.education.gouv.fr/pid285/bulletin_officiel.html?pid_bo=39051>
- Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner, G.W, Humphreys (Eds.), *Basic processes in reading: Visual word recognition* : 148-197. Hillsdale, NJ: Erlbaum.
- Mortamet, C. (2019). *L'orthographe. Pratiques d'élèves, pratiques d'enseignants, représentations*. Mont-Saint-Aignan : Presses universitaires de Rouen et du Havre.
- Murat, F. (2005). La compétence des adultes à l'écrit, en calcul et en compréhension orale. *INSEE Première*, (1044), <https://www.epsilon.insee.fr/jspui/bitstream/1/224/1/ip1044.pdf>
- Murat, F., Rocher T. (2016). L'évolution des compétences des adultes : effet « génération » et effet « cycle de vie ». *Économie et Statistique*, (490) : 61-83.
- Nadeau, M., Fisher, C. (2012). La phrase dictée du jour et la « grammaire nouvelle » : pour des progrès en orthographe grammaticale. *La Lettre de l'AIRDF*, (52) : 31-35.
- Nadeau, M., Fisher, C. (2014). Expérimentation de pratiques innovantes, la dictée 0 faute et la phrase dictée du jour, et étude de leur impact sur la compétence orthographique des élèves en production de textes. *Rapport de la recherche FQRSC*, <<http://www.frgsc.gouv.qc.ca/fr/parteneriat/rapports-de-recherche>>
- Nadeau, M., Fisher, C. (2014). Usage du métalangage et des manipulations syntaxiques au cours de dictées innovantes dans les classes du primaire. *Repères*, (49) : 169 : 191.
- New, B., Pallier, C., Ferrand, L., Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'Année Psychologique*, (101) : 447-462, <<http://www.lexique.org>>
- New, B., Pallier, C., Brysbaert, M., Ferrand, L. (2004). Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, (36-3) : 516-524.
- OCDE. Le cadre d'évaluation de PISA 2015 : Les compétences clés en compréhension de l'écrit, en mathématiques et en sciences. PISA, Éditions OCDE, Paris, <<https://doi.org/10.1787/9789264259478-fr>>

- OCDE. PISA 2015. *Les défis du système éducatif français et les bonnes pratiques internationales*, <<http://www.oecd.org/pisa/PISA-2015-Brochure-France.pdf>>
- Ortége, É., Lété, B. (2010). *EManulex-Infra : Electronic version of Manulex-Infra and Manulex-Infra-infra databases*, <<http://www.Manulex-Infra.org>>
- Pacton, S., Fayol, M., Perruchet, P. (2002). Acquérir l'orthographe du français : apprentissages implicites et explicites. In Florin, A., Morais, J. (dir.), *La maîtrise du langage*, (95-118). PUF : Rennes.
- Pacton, S., Casalis, S. (2006). L'utilisation d'informations morphologiques en production écrite rend-elle moins cauchemar(d) des lettres muettes moins cauchemardesque ? *Rééducation orthophonique*, (225) : 129-144.
- Paveau, M-A, Rosier, L. (2008). *La langue française : passions et polémiques*. Paris : Vuibert.
- Peereman, R., Lété, B., Sprenger-Charolles, L. (2007). Manulex-Infra-infra : Distributional characteristics of grapheme-phoneme mappings, infra-lexical and lexical units in child-directed written material. *Behavior Research Methods*, (39) : 593-603 et 579-589.
- Peereman, R., Sprenger-Charolles, L., Messaoud-Galusi, S. (2013). The contribution of morphology to the consistency of spelling to sound relation : A quantitative analysis based on French elementary school readers. *L'Année Psychologique*, (113-1) : 3-33.
- Pellat, J.-C. (1988). Indépendance ou interaction de l'écrit et de l'oral ? Recensement critique des définitions du graphème. In N. Catach (Éds.), *Pour une théorie de la langue écrite, Actes de la Table Ronde internationale*, CNRS-HESO, (133-146). Paris : Éditions du CNRS.
- Pellat, J.-C. (2001). Repères pour l'histoire de l'orthographe française : le XVIIe siècle. In C. Gruaz et R. Honvault (Éds.), *Variations sur l'orthographe et les systèmes d'écriture. Mélanges en hommage à Nina Catach*, (241-256). Paris : Honoré Champion.
- Pellat, J.-C., Fonvielle, S. (2016). *Le Grévisse de l'enseignant*. Paris, Magnard.
- Penloup, M-C. (2000). *La tentation du littéraire : essai sur le rapport à l'écriture littéraire du scripteur ordinaire*. Éditions Didier.
- Penloup, M-C. (2006). Vers une didactique de l'écriture centrée sur l'apprenant et ses pratiques. In Lafont J., Colin D. (coord.), *Didactique de l'écrit - La construction des savoirs et le sujet écrivant*, coll. Diptyque, (81-104). Namur : Presses Universitaires de Namur.
- Péret, C., Sautot, J-P., Brissaud, C. (2008). Les professeurs entrant dans le métier et la norme orthographique. In Brissaud, C., Jaffré J-P., Pellat J-C., *Nouvelles recherches en Orthographe*,

- actes des journées d'étude des 14 et 15 juin 2007 - Université de Strasbourg, (203-214).
Limoges : Lambert Lucas.
- Pérez, M., Giraudo, H. (2019). Familiarité, complexité orthographique et encodage dans les dictées de mots en 2^{ème} élémentaire (CE1). In Kervyn, B., Dreyfus, M., Brissaud, C., *L'écriture dès le début de l'école primaire*. Pratiques enseignantes et performances des élèves, coll. « Études sur l'Éducation », (201-214). Pessac : Presses Universitaires de Bordeaux.
- Pothier, B., Pothier, P. (2003). EOLE : Échelle d'acquisition en orthographe lexicale du CP au CM2. Paris : Retz.
- Petit Robert de la langue Française (2011). *Version numérique*. Édition Le Robert.
- Riegel, M., Pellat, J.C., Rioul, R. (1994, première édition). *Grammaire méthodique du français* (Édition revue et augmentée de 2009). Paris : Presses universitaires de France.
- Sénéchal, M. (2000). Morphological effects in children's spelling of french words. *Canadian Journal of Experimental Psychology*, (54) : 76-85.
- Sénéchal, M., Basque, M.T., Leclaire, T., (2006). Morphological knowledge as revealed in children's spelling accuracy and reports of spelling strategy. *Journal of Experimental Child Psychology*, (95) : 231-254.
- Ters, F., Mayer, G., Reichenbach, D. (1977). *Vocabulaire orthographique de base*. Paris : éditions OCDL.
- Ters, F., Mayer, G., Reichenbach, D. (1995). *L'échelle « Dubois-Buyse »*. Paris : éditions OCDL.
- Thévenin, M-G, Totereau, C., Jarousse, J-P, Fayol, M. (1999). L'apprentissage/enseignement de la morphologie écrite du nombre en français. *Revue française de pédagogie*, (126), L'acquisition/apprentissage de l'orthographe » : (39-52).
- Thimonnier, R. (1967). *Le système graphique du français*. Paris : Plon.
- TLFi. Trésor de la langue Française informatisé, CNRS, Université de Lorraine : <http://atilf.atilf.fr/>
- Totereau, C., Thévenin, M-G, Fayol, M. (1997). The development of the understanding of number morphology in written French. In Rieben, M., Fayol, M., Perfetti, C. (Éds.), *Learning to spell*, (97-114). Hillsdale, NJ : L.E.A.
- Totereau, C., Brissaud, C., Reilhac, C., Bosse, M-L. (2013). L'orthographe grammaticale au collège : une approche socio-différenciée. *ANAE*, (123) : 164-171.

UNESCO (2009). La nouvelle génération de statistiques sur les compétences en alphabétisme : mise en œuvre du Programme d'évaluation et de suivi de l'alphabétisation (LAMP), 10-12 juin 2003 : <<http://uis.unesco.org/sites/default/files/documents/the-next-generation-of-literacy-statistics-implementing-the-literacy-assessment-and-monitoring-programme-lamp-fr.pdf>>

UNESCO. *Rapport mondial de suivi sur l'éducation pour tous*, (155-170) : <http://www.unesco.org/education/GMR2006/full/chap6_fr.pdf>

Villechaise-Dupont, A., Zaffran, J. (2002). Le « drame » de l'illettré : analyse d'une fiction sociologique à succès politique. *Langage et Société*, (102) : 71-96.

Weekes, B. S., Castles, A. E., Davies, R.A (2006). Effects of consistency and age of acquisition on reading and spelling among developing readers. *Reading and Writing*, (19- 2) : 133-169.

ANNEXES

(original)

pharmacie

enti rutme Mikatol arape
alcool 9° du sce

épicerie tomate de pays

confiture cerise sel gobarté

krois fromage qui sente pas fort

librairie de fiancé ~~sonnet~~
sonanel

du jouque pèfanor

82 01 507 585

Annexe 2 : Dictée produite par l'enquêté « haut » n° 187

Pharmacie - Anti Rhume - Nicathole - Cirepe
 Alcool à 90° ~~deux~~ - Epicerie - tomates de pays
 Confiture - Cerises - sel Gobard - 3 fromages qui
 ne sentent pas fort - Librairie - le fiancé solennel
 de Jacques Pesanord - Bricolage - 2 produits
 parfumés qui dégraisent bien - 4 paquets de vices
 flâtes - un ciseau à bois normal - une boîte
 d'allumettes - une pince pour arracher les clous
 un pot de peinture - quelques mètres de
 tuyau d'arrosage.

A travaillé au ministère de la Défense → formé au français des
 jeunes recrues illettrées puis responsable du groupe illettrisme
 avec A. Bentolila.

Annexe 3 : Classement des graphies des enquêtés « bas » et « haut » relevées dans la dictée d'IVQ

ALCOOL																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	alcool	267	79	alcol	52	3	alcohol	2		acool	9		26	4		0
				alcole	12		alchol	1		acole	6					
				alccol	5		alecoole	1		accol	5	1				
				alcoole	3		aalcool	1		acol	3					
				allcol	3					acoll	3					
				alcoll	3					la cole	2					
				alcolle	2					alcoul	2					
				alccool	2					alcele	2					
				alcols	1					olcol	2					
				âlcol	1					alcc	1					
				àlcoole	1					alcolol	1					
				al'coole	1					alc	1					
				alkol	1					alccceal	1					
				alkole	1					aclool	1					
				alcôl	1					alcoil	1					
				allcole	1					alceul	1					
				allcolle	1					alkl	1					
				al col	1					aleceule	1					
										alcocol	1					
										accole	1					
									aclol	1						
									olkol	1						
									alokol	1						
									allol	1						

										alool	1					
										alocl	1					
										laritala	1					
										acoole	1					
										hacole	1					
										cole	1					
										oh	1					
										ol	1					
total	orthographe correcte	267	79	phonographie acceptable	92	3	phonographie approximante	5	0	phonographie non acceptable	57	1	26	4	0	0

ANTI																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	anti	283	85	enti	61	2	aen ti	1		inti	4		23	0	1	0
				entie	9					onti	3					
				antie	5					entu	3					
				en ti	4					antu	2					
				enthi	4					oti	2					
				hanti	4					ani	2					
				anthi	2					enté	2					
				anty	1					ente	2					
				enty	1					ati	2					
				henti	1					anteu	1					
				hentie	1					antem	1					
				antis	1					ant	1					
				anthie	1					entil	1					
				antti	1					enter	1					
									hnti	1						

										int	1					
										onté	1					
										neti	1					
										ont	1					
										omti	1					
										anté	1					
										enth	1					
										outri	1					
										adi	1					
										un tirere	1					
										yanti	1					
										un ta	1					
										anlyti	1					
										anet	1					
										andi	1					
total	orthographe correcte	283	85	phonographie acceptable	96	2	phonographie approximante	1	0	phonographie non acceptable	43	0	23	0	1	0

CERISES																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			Non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	cerise	184	34	serise	20	1	seriz	3		cerice	14	1	21	3	2	0
	cerises	130	46	serize	4		seris	3		cerices	8	1				
				çerise	3		ceriz	1		cerisse	5					
				cerize	2		criz	1		cerie	2					
				se rises	1		ceris	1		cherige	2					
				ceurises	1					cérise	2					
				cerizes	1					serice	2	1				
				cerrise	1					sourice	2					

								cerige	1					
								cerisye	1					
								sourisi	1					
								caurise	1					
								cerrisse	1					
								cerisie	1					
								cirise	1					
								cerisece	1					
								cherries	1					
								c�erises	1					
								coric	1					
								courize	1					
								cericie	1					
								curize	1					
								serese	1					
								seriere	1					
								suris	1					
								cerese	1					
								ceil	1					
								serisse	1					
								sorise	1					
								sorice	1					
								sirige	1					
								soris	1					
								sourise	1					
								zerisse	1					
								sourix	1					
								cerecies	1					
								sorizi	1					
								sorisi	1					

										sugüz	1					
total	orthographe correcte	314	80	phonographie acceptable	33	1	phonographie approximante	9	0	phonographie non acceptable	68	3	21	3	2	0

CONFITURE																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	confiture	330	73	comfiture	12		confiteure	1		confiture	6		15	0	1	0
	confitures	26	14	confitur	7					confeture	3					
				comfitur	2					conféture	3					
				comfitures	1					confitue	2					
				konfitur	1					cofiture	2					
										confitire	2					
										colour	1					
										conetur	1					
										confector	1					
										conféteure	1					
										confetoure	1					
										canfiture	1					
										canffeuteur	1					
										confitures	1					
										confidür	1					
										conficture	1					
										confirture	1					
										confi	1					
									confit	1						
									coinfuture	1						
									caunfieture	1						
									cftir	1						

										confitour	1					
										confitoure	1					
										confituire	1					
										confretere	1					
										confitura	1					
										confiturun	1					
										confutire	1					
										confouteur	1					
										confutier	1					
										confutre	1					
										contitue	1					
										kmfutur	1					
										coufituor	1					
										convetur	1					
										counfiture	1					
										convtor	1					
										kouifetoure	1					
total	orthographe correcte	356	87	phonographie acceptable	23	0	phonographie approximante	1	0	phonographie non acceptable	51	0	15	0	1	0

DUX													
	phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	dux	203	64	ducs	19	2	duce	7		33	4	1	0
	duxe	76	10	duks	5	1	deux	6					
	ducse	7		ducx	3		dus	5					
	dukse	4	1	ducks	2		dusque	4					
	disc	3		duchs	1		dix	2					

duſce	3		duqs		1	dups	2					
dûx	1					duſſe	2					
duxce	1					doux	2					
duqce	1					duques	2					
d'uxe	1	1				luxe	2	2				
duxs	1					dixſe	1					
duckſe	1					dks	1					
dukce	1					dics	1					
ducſſe	1					dexe	1					
						digux	1					
						diquxe	1					
						diquesi	1					
						disc	1					
						dexu	1					
						dixe	1					
						dixs	1					
						ducaé	1					
						drex	1					
						duc	1					
						ducce	1					
						doex	1					
						ducſer	1					
						du	1					
						duct	1					
						due	1					
						duex	1					
						dupps	1					
						deqc	1					
						dupsne	1					
						durcque	1					

						tux	1						
						durc	1						
						de ux	1						
						duquese	1						
						d'us	1						
						duxeue	1						
						bux	1						
						dcu	1						
						duxet	1						
						dusk	1						
						duske	1						
						dusqueu	1						
						de quse	1						
						decece	1						
						dust	1						
						dx	1						
						lux	1						
						ükus	1						
						tus	1						
						tuse	1						
						lude		1					
total	phonographie acceptable	304	76	phonographie approximante	30	4	phonographie non acceptable	79	3	33	4	1	0

ÉPICERIE																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
gra phi	epicerie	203	55	epiceri	8		episerie	8	1	epecerie	4		25	1	1	0
	épicerie	99	22	episserie	4		episeri	4		epecirie	2					

èpicerie	4		epiçerie	2		et piserie	2		epicirie	2					
Epicerie		8	épisserie	2		épiecrie	2		epcerie	2					
			epicery	1		et piseri	1		epicérie	2					
			épicerrie	1		epicrie	1		episiri	2					
			épiceri	1		epicri	1		ipiceri	2					
			episseris	1		épiserie	1		picherie	1					
			hépicerie	1		épisrie	1		epecieri	1					
						épicerie	1		epecèrie	1					
									epècrer	1					
									epeiseire	1					
									éperciei	1					
									épercié	1					
									epesorie	1					
									epesseri	1					
									epesseries	1					
									ephicerie	1					
									épèsserie	1					
									éphisserie	1					
									epicaïri	1					
									épi ssiri	1					
									épice	1					
									épicercaie	1					
									épícíri	1					
									eipiceraï	1					
									epicir	1					
									epicire	1					
									epcorio	1					
									epiciri	1					
									pepisire	1					
									ipsore	1					

									pisr	1				
									pisarie	1				
									épiscie	1				
									épiciarie	1				
									epircérie	1				
									episary	1				
									et pusseris	1				
									episéri	1				
									ale pissiri	1				
									épicière	1				
									epicérie	1				
									epucrai	1				
									et piro	1				
									épouséri	1				
									epuisre	1				
									épisséure	1				
									ibisiri	1				
									ipecerie	1				
									ibsouri	1				
									ibsori	1				
									éupece	1				
									épicierie	1				
									ipicierie	1				
									ipiseri	1				
									épecerie	1				
									ipicerie	1				
									ipisseri	1				
									les pesrir	1				
									picsirie	1				
									pisri	1				

										picri	1					
total	orthographe correcte	306	85	phonographie acceptable	21	0	phonographie approximante	22	1	phonographie non acceptable	72	0	25	1	1	0

FIANCÉ																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	fiancé	199	76	fiancer	18	1				fiancie	8		20	0	4	0
	fiance	35	1	fiançé	9	1	fianse	2		fiancier	6	1				
	fiancée	33	3	fiançais	4		fiancei	1		fiancié	5					
	FIANCE	14		fience	4					fiyonse	4					
	fiancee	2		fiencer	3		fillansais	1		fance	1					
	fiancè	2	1	fillancé	3	1	FIYANSE	1		fiaicé	1					
	fianceé	1		fiençé	2		fiençé	1		faiençe	1					
	fiancès	1		fiançee	1		fiençe	1		faiance	1					
				fiançais	1					fauisi	1					
				fiançaie	1					fiancial	1					
				fiançés	1					fiallicer	1					
				fiençée	1					faincé	1					
				phiancée	1					fianciée	1					
				fiençai	1					faince	1					
				fiençais	1					fiané	1					
				fiançai	1					fiancillé	1					
				fiençe	1					fiancille	1					
				fiancai	1					fiane	1					
				fiançée	1					fense	1					
				fillancée	1					ficiancer	1					
			fillencée	1					fiasion	1						
			fiancey			1			ficanse	1						

								fensé	1					
								fiacé	1					
								fiencçais	1					
								fiace	1					
								fensse	1					
								filensser	1					
								filianse	1					
								fieslance	1					
								feyanse	1					
								féancé	1					
								finances	1					
								fimance	1					
								finosé	1					
								fincie	1					
								fiomcé	1					
								fioncé	1					
								fionçé	1					
								fionce	1					
								fionsé	1					
								fioncier	1					
								FIONSE	1					
								FIONCE	1					
								foujoci	1					
								feyancé	1					
								fiyncie	1					
								fiacne	1					
								fyoncé	1					
								françai	1					
								fuiancer	1					
								viensé	1					

										fensie	1					
										financieru		1				
total	orthographe correcte	287	81	phonographie acceptable	57	4	phonographie approximante	7	0	phonographie non acceptable	72	2	20	0	4	0

FORT																	
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible		
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut	
graphies	fort	309	78	for	13					forte	8		37	0		0	
	forts	43	8	ford	10	1				four	2						
				fore	4					fonre	1						
				fors	4					forde	1						
				fohr	1					fat	1						
				faure	1					f	1						
				forr	1					feure	1						
										force	1						
										fora	1						
										vor	1						
										fot	1						
										fours	1						
										fere	1						
										fror	1						
									furt	1							
									frot	1							
total	orthographe correcte	352	86	phonographie acceptable	34	1	phonographie approximante	0	0	phonographie non acceptable	24	0	37	0	0	0	

FROMAGES

FROMAGES																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	fromages	241	81	frommage	2					formage	9		24		1	0
	fromage	129	6	fromaje	2					fomage	4					
				fromaj	2					formages	2					
				fraumage	1					formoge	2					
				phromaje	1					framages	2					
										fromges	2					
										farmagi	1					
										formaj	1					
										feuranege	1					
										formaje	1					
										fomages	1					
										fouromage	1					
										fouramg	1					
										français	1					
										fomof	1					
										framje	1					
										fremage	1					
										farmage	1					
										freumage	1					
										fromace	1					
									fohmos	1						
									fomja	1						
									fromege	1						
									fromane	1						
									frome	1						
									feremaja	1						
									froumoge	1						

										fronage	1					
										frommge	1					
total	orthographe correcte	370	87	phonographie acceptable	8	0	phonographie approximante	0	0	phonographie non acceptable	44	0	24	0	1	0

GOBAR														
	phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible		
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut	
graphies	gobar	139	10				cobar	12	3	21	0	1	0	
	gobard	102	40				cobare	10						
	gobare	32	3				goubar	6						
	gobart	27	15				goubare	3						
	gaubar	10					goba	3						
	gaubard	7	3				coba	2						
	gobarre	4	1				gombar	2						
	gaubart	3					copar	2						
	guobar	2	1				gabard	2	1					
	gobbar	1	1				cobard	2						
	gaubarre	1					quobar	2						
	guobard	1					cobarre	2						
	guobare	1					cobart	2						
	gaubart		1				coubare	1						
	gobbart		1				caubare	1						
	gabard		1				caubar	1						
							co"bas"	1						
							copa	1						
						combar	1							
						codar	1							
						copard	1	1						

					courra	1					
					copra	1					
					gabab	1					
					gabare	1					
					gambar	1					
					gauba	1					
					gaubrare	1					
					gebare	1					
					gepari	1					
					gobab	1					
					gobad	1					
					gobale	1					
					gobat	1					
					gobor	1					
					gobû	1					
					gobr	1					
					godard	1					
					godart	1					
					gomar	1					
					gomard	1					
					baubard	1					
					gommares	1					
					gorbart	1					
					gopar	1	1				
					goubaire	1					
					bobard	1					
					cofar	1					
					govar	1					
					gouber	1					
					gova	1					

							gros bare	1					
							gowlbar	1					
							gubahe	1					
							gubar	1					
							guobair	1					
							jobard	1					
							colmar	1					
							copare		1				
							copart		2				
							ocbard		1				
total	phonographie acceptable	330	77	phonographie approximante	0	0	phonographie non acceptable	95	10	21	0	1	0

JOC													
	phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	jock	83	26	goc	6		jox	8		25	0	1	0
	joc	77	21	goque	5		jacque	5					
	jok	54	10	gok	4	1	jacques	5	6				
	joque	43		goq	2		jaque	4					
	jocque	16	2	goke	2		choc	3					
	jauque	11	1	gock	1		jouque	2					
	joke	9	5	gooke	1		jos	2					
	jocques	6	9	djoc	1	1	jouk	2					
	jocq	6	1	djoque	1		soce	1					
	joq	2					foc	1					
	joqu	2					geaque	1					
	jauc	2					geod	1					
	jocke	1	1				georges	1					

joques	1	1				gogue	1						
jauques	1					gopenor	1						
joch	1					gonque	1						
jokc	1					gorge	1						
joks	1					gouch	1						
jauke	1					jace	1						
jauq	1					gque	1						
joqe	1					j.	1						
geoc	1					gorque	1						
						gox	1						
						jack	1						
						choque	1						
						chonque	1						
						jake	1						
						jec	1						
						choq	1						
						jece	1						
						jek	1						
						jeuq	1						
						joage	1						
						joco	1						
						jode	1						
						jof	1						
						jope	1						
						jonc	1						
						jor	1						
						jouce	1						
						jote	1	1					
						jorque	1						
						gac	1						

							jougue	1					
							geage	1					
							dioke	1					
							chonpe	1					
							sog	1					
							jouki	1					
							joxe	1					
							poke	1					
							juc	1					
							jugue	1					
							vok	1					
							jo		1				
total	phonographie acceptable	321	77	phonographie approximante	23	2	phonographie non acceptable	77	8	25	0	1	0

LIBRAIRIE

	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	librairie	254	81	librerie	43	4				liberie	6		29	0	2	0
	librairies	3		librerie	11	2				librarie	5					
				libreri	10					libririe	4					
				librairi	6					librire	3					
				libreri	4					libraire	3					
				librerie	2					libairie	2					
				librairis	1					lebriri	2					
				l'hibrerie	1					library	2					
				libraierie	1					libriri	2					
				l'ibrairie	1					livrairie	2					
										livrirer	1					

								lévarie	1					
								les brire	1					
								luiprerie	1					
								libererie	1					
								libere	1					
								colebé	1					
								libourie	1					
								l'iberie	1					
								libeuri	1					
								liberiryai	1					
								libracrie	1					
								elrairie	1					
								librere	1					
								librére	1					
								librier	1					
								libregi	1					
								epririe	1					
								librairie	1					
								lebrairie	1					
								libre	1					
								libraiser	1					
								epréi	1					
								l'abirerie	1					
								imprepie	1					
								libreririe	1					
								librérié	1					
								libreuri	1					
								les birire	1					
								librille	1					
								libriré	1					

										laibrerie	1					
										lébraire	1					
										libririt	1					
										librri	1					
										liphagi	1					
										l'imprimerie	1					
										liperi	1					
										liprerie	1					
										lebrerie	1					
										lirerie	1					
										lipriri	1					
										livrarir	1					
										livraraire	1					
										livreire	1					
										mipari	1					
										lubierie	1					
										livveri	1					
total	orthographe correcte	257	81	phonographie acceptable	80	6	phonographie approximante	0	0	phonographie non acceptable	79	0	29	0	2	0

MICATOL													
	phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	micatol	166	46				nicatol	8		31	1	2	0
	micatole	56	11				mecatol	5	1				
	mikatol	38	6				mécatol	2					
	micathol	16	2				mecatole	2					
	mikatole	8					mécatole	2					
	micatolle	7					mekatol	2					

mycatol	7	9				micadole	2						
micathole	4	3				micadol	2	2					
miquatol	4					makatol	1						
mycathol	4	3				mecatile	1						
micatôle	3					mecalole	1						
miquatole	3					medical	1						
mycathole	3					makatale	1						
mycatole	3					megatole	1						
mickatol	2					mequatole	1						
miccatol	2					méquatolle	1						
mykatole	2					mercatole	1						
micatoll	1					licatale	1						
mickatole	1					miatol	1						
miquatolle	1					micacol	1						
mi ka tol	1					licatole	1						
mikathol	1					micadone	1						
mykattol	1					micale	1						
mi catol	1					micadore	1						
mycatoll	1					micaltol	1						
miqatole	1					micaltole	1						
micateùl	1	1				micatau	1						
miekatele	1					micaltome	1						
						micatofe	1						
						micatloul	1						
						micathurl	1						
						micatope	1						
						micatome	1						
						micator	1						
						micatora	1						
						micatou	1						

						micatoul	1						
						micotole	1						
						micotol	1						
						migatol	1						
						migotal	1						
						mihatol	1						
						mik taule	1						
						mikadol	1	1					
						mikal	1						
						mikatorle	1						
						mikotol	1						
						mille catole	1						
						micatoul	1						
						mitatol	1						
						mit-cole	1						
						mucatul	1						
						mktn	1						
						mugatole	1						
						mucatole	1						
						nicathol	1						
						picatol	1						
						nicatole	1						
						mica		1					
total	phonographie acceptable	339	81	phonographie approximante	0	0	phonographie non acceptable	75	5	31	1	2	0

PAYS																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	pays	303	83	péie	1		pei	5		pay	34	3	20	0	1	0
							payi	3		paye	15	1				
							payie	3		pais	6					
							peyi	2		payé	5					
							peie	1		payer	3					
							payis	1		pey	3					
							péyi	1		péye	2					
										payes	2					
										peye	2					
										paiy	1					
										paire	1					
										Pai	1					
										pey	1					
										paie	1					
										payee	1					
										<i>payée</i>	1					
										payés	1					
										payier	1					
										peaye	1					
										peié	1					
									peiér	1						
									pell	1						
									peuy	1						
									peri	1						
									peilhe	1						
									pille	1						
									paiu	1						

										peyer	1					
										paiss	1					
										peyé	1					
										pieni	1					
										pieyu	1					
										<i>pié</i>	1					
										pie	1					
										piy	1					
										pyer	1					
										pye	1					
										poye	1					
										piye	1					
										piyer	1					
										poye	1					
										piyi	1					
										pvi	1					
total	orthographe correcte	303	83	phonographie acceptable	1	0	phonographie approximante	16	0	phonographie non acceptable	106	4	20	0	1	0

PÉVANOR													
	phonographie acceptable			phonographie approximante			phonographie non acceptable			non réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	pevanor	90	17				pivanore	4		25	0	1	0
	pévanor	48	23				prévanor	3					
	pevanore	37	6				pevanol	3					
	pevanord	30	13				evanor	3					
	pévanord	19	9				peranor	3					
	pévanore	16	3				péranor	3					
	pévanort	11	4				puvanore	2					

pevanort	6				prévanord	2						
pèvanor	3				pavanor	2						
pèvanore	3				pavanore	2						
pevanores	2				penavor	2						
pévannor	2	1			pénavol	2						
pévanaure	2				penavore	2						
paivanore	2				pevanare	2						
p.vanor	2				pévanol	2						
pevanaure	1				pevanoir	2						
pevanorr	1				péfanor	1						
pévanorre	1				pefanero	1						
paivanor	1				pèfanor	1						
pévanaur	1				péfanore	1						
pe vanore	1				pémanore	1						
p.evanor	1				pefanort	1						
pevannor	1				pélanord	1						
p vanor	1				pelanor	1						
petvanor	1				penanour	1						
pey va nor	1				péna vore	1						
pevanorre	1				pénalor	1						
pevanors	1				penalord	1						
pe va nore	1				pebano	1						
pevaneur	1				pebanort	1						
P Van Ore		1			pev	1						
					penarvort	1						
					paivanr	1						
					pénavore	1						
					penavort	1						
					penvarore	1						
					penvannor	1						

					pepanort	1					
					hévanoe	1					
					p. vanoque	1					
					peragor	1					
					perranor	1					
					peranore	1					
					perevadole	1					
					pervanor	1					
					pervanau	1					
					pervanaure	1					
					pevamor	1					
					ben vante	1					
					pevalnore	1					
					pevanère	1					
					pèvanel	1					
					pevaner	1					
					pévano	1					
					pevanoc	1	1				
					et banore	1					
					evenort	1					
					gopenor	1					
					pévanole	1					
					pévanoru	1					
					pevenor	1					
					pevenore	1					
					pevavoure	1					
					pevarone	1					
					pevanory	1					
					pévanur	1					
					peynavaure	1					

						pfinra	1					
						pevunro	1					
						pevwannol	1					
						pévonord	1					
						pevonore	1					
						pevonord	1					
						pevnore	1					
						pevaroeur	1					
						pevaur	1					
						pivano	1					
						pirang	1					
						pilanore	1					
						pinalar	1					
						pifannre	1					
						pifanor	1					
						pifamer	1					
						pfand	1					
						pivanorre	1					
						pivanort	1					
						pivanort	1					
						be va noir	1					
						pivanord	1					
						pivanor	1					
						pivanneur	1					
						pivanol	1					
						pivanar	1					
						povanor	1					
						porevka	1					
						povenor	1					
						pivolare	1					

						pivanot	1						
						pivanur	1						
						trés vanore	1						
						tevanaurt	1						
						sovalor	1						
						prevonor	1						
						privanerre	1						
						prevanore	1						
						ben vante	1						
						bévano	1						
						prevanord	1						
						bivanore	1						
						prevanor	1						
						p.e	1						
						bevanor		1					
						bivabor		1					
						pavanord		1					
						pevarnor		1					
						bivenne		1					
						Evanord		1					
						P.VNOR		1					
						pevevanort		1					
						Sépavanor		1					
total	phonographie acceptable	288	77	phonographie approximante	0	0	phonographie non acceptable	133	10	25	0	1	0

PHARMACIE																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	pharmacie	320	86	farmacie	10					phamacie	9		18	0	1	0
				farmaci	8					phamaci	3					
				pharmaçie	7					famaci	2					
				pharmaci	6					parmacie	2					
				fharmacie	3					phamasi	2					
				fharmaçie	2					pharmace	2					
				pharmacy	1					pharmacie	2					
				farhmaci	1					phramacie	2					
				farmaçie	1					faramaci	1					
				farmassi	1					famachi	1					
				pharmaçi	1					famacie	1					
										famansi	1					
										phrmace	1					
										famasie	1					
										faramaci	1					
										farimace	1					
										farmac	1					
										farmace	1					
										farmacé	1					
										farmacille	1					
										fhrimas	1					
										flammacie	1					
										formation	1					
									framacie	1						
									framasi	1						
									fromachie	1						
									hpar mac	1						

										nhamacia	1					
										paharmacé	1					
										paremocé	1					
										parmasi	1					
										pehamce	1					
										ph	1					
										phamarci	1					
										pharmacie	1					
										phamocie	1					
										pharm	1					
										pharma	1					
										pharmacé	1					
										pharmaces	1					
										pharmacei	1					
										pharmacice	1					
										pharmage	1					
										pharmaicie	1					
										pharmcie	1					
										pharmeï	1					
										pharmicie	1					
										pharmfie	1					
										phomarcie	1					
										phavumachie	1					
										qharmacie	1					
										farmaice		1				
total	orthographe correcte	320	86	phonographie acceptable	41	0	phonographie approximante	0	0	phonographie non acceptable	67	1	18	0	1	0

RHUME																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	rhume	225	73	rume	55		rhum	51	8	rime	6		26	0	1	0
	rhumes	2	1	ruhme	17	2	reume	2		rome	3					
				rum	13	2	reum	1		roume	3					
				ruhm	2		rhûm	1		rhyme	2					
				hrume	1		rheum	1		reme	2					
				rhûme	1		rueme	1		hume	2					
				rhumme	1					rom	2					
				rumh	1					rune	2					
				rumme	1					rhime	2					
				rhûme		1				mump	1					
										rem	1					
										hum	1					
										rhune	1					
										rhym	1					
										rme	1					
										rm	1					
										rihme	1					
										me	1					
										ruhe	1					
										ru	1					
										roum	1					
										royeme	1					
										reyme	1					
									rmo	1						
									rumer	1						
									rhme	1						
									na	1						

										tirere	1					
										urm	1					
total	orthographe correcte	227	74	phonographie acceptable	92	5	phonographie approximante	57	8	phonographie non acceptable	44	0	26	0	1	0

SEL																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	sel	394	86	cel	6					sal	2		18	1	0	0
	sels	3		selle	4					sol	2					
	sèl	2		sele	3					salle	1					
	sél	2		sell	1					cil	1					
				seile	1					sl	1					
				<i>cél</i>	1					cal	1					
				<i>cels</i>	1					se	1					
			cele	1					suel	1						
total	orthographe correcte	401	86	phonographie acceptable	18	0	phonographie approximante	0	0	phonographie non acceptable	10	0	18	1	0	0

SENTENT																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	sentent	166	75	sante	27		sent	17	1	sonte	8		41	1	1	0
	sente	128	9	son	6		sant	3		sans	3					
	sont	10	1	cente	4		scent	1		san	2					
	sentes	6		santes	1		sents	1		se	2					
	semblent	1		sons	1		cent	1		sende	2					
				santte	1					sontent	1					
									s	1						

									con	1						
									sàn	1						
									sanc	1						
									sa	1						
									chente	1						
									seat	1						
									sieta	1						
									sate	1						
									sens	1						
									cen	1						
									sen	1						
									centre	1						
total	orthographe correcte	311	85	phonographie acceptable	40	0	phonographie approximante	23	1	phonographie non acceptable	31	0	41	1	1	0

SIRAPE													
	phonographie acceptable			phonographie approximante			phonographie non acceptable			non réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	sirap	111	41	ssirape	1		ciraque	3	1	27	0	0	0
	sirape	96	17	sirapeux	1		sirope	3					
	cirape	54	4	sirapeu	1		cérape	2					
	cirap	33	8	sixrape	1		cirat	2					
	cyrape	10	4	six.rappes	1		sirate	2					
	syrape	9		cyrapp		1	sirup	2					
	syrap	9	5				surape	2					
	sirappe	8					cerape	2					
	cyrap	5	1				cirapar	1					
	sirape	5					cirapeure	1					
	çirape	3					cypar	1					
	cirappe	3	1				cyrak	1					

cirrape	2					siraph	1					
si rape	2					cirapre	1					
cyrappe	1					ciralé	1					
cyrrape	1					cirope	1					
siraps	1					dicirppe	1					
cirrap	1					firape	1					
sirrape	1					shyrape	1					
syrrap		1				cirabe	1					
						rirape	1					
						serap	1	1				
						sirable	1					
						simate	1					
						sirab	1					
						sirah	1					
						sirame	1					
						sirant	1					
						sirapa	1					
						sirapae	1					
						siraphe	1					
						siraple	1					
						sirapt	1					
						siraqe	1					
						sirapy	1					
						sireble	1					
						siraque	1					
						sirhp	1					
						cipar	1					
						sirop	1					
						chirapé	1					
						sirra	1					

							sirpe	1					
							sirp	1					
							chirap	1					
							sqpa	1					
							surappe	1					
							cera	1					
							syraph	1					
							syrop	1					
							sira		1				
							sirabe		1				
total	phonographie acceptable	355	82	phonographie approximante	5	1	phonographie non acceptable	60	4	27	0	0	0

SOLENNEL

	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	solennel	10	43	solanel	90	18	solanaille	1		solenel	21	10	24		3	0
	solennelle	6	2	solanelle	42	2	colanelle	1		solonel	10					
	solennèl	1		solannel	28	5				solenelle	6	1				
	solennél	1		solanele	8					solonnel	4					
				solannelle	5	2				solonelle	4					
				sollanelle	5					soulanel	3					
				solanèle	3					soulalane	2					
				sollannel	2					sonanel	2					
				solanéle	2					sonanelle	2					
				saulanel	2					solnale	2					
				solennele	2					selon elle	2					
				solanell	2					sonaelle	2					
				sollanel	1	1				sous la nelle	2					

		sollanell	1					selanel	2					
		sollanel	1					sol en L	1					
		solanelles	1					solané	1					
		sollenele	1					solane	1					
		solanéle	1					solanatel	1					
		saulanelle	1					solanlle	1					
		sol à nel	1					solanne	1					
		solanail	1					solanielle	1					
		solanèl	1					solanle	1					
								soleneel	1					
								solénel	1					
								selamve	1					
								solanéé	1					
								solenl	1					
								solésnel	1					
								sollenel	1	1				
								sollenele	1					
								sollonel	1					
								selon nelle	1					
								sola laine	1					
								solnel	1					
								solonaile	1					
								solocarel	1					
								selanniele	1					
								solonl	1					
								solaenel	1					
								son anel	1					
								soml	1					
								somrallane	1					
								sona lel	1					

								sonael	1					
								sonaelle	1					
								selanalle	1					
								sonalle	1					
								sor la nelle	1					
								sonale	1					
								sonalés	1					
								selaluni	1					
								se lin	1					
								sor la nel	1					
								sonnelle	1					
								sonellene	1					
								sonoll	1					
								sou la milé	1					
								selenel	1					
								sou la nelle	1					
								se la nêlle	1					
								soulanal	1					
								soulanele	1					
								soulanelle	1					
								sounanl	1					
								sovanel	1					
								sous lenmenle	1					
								su laline	1					
								sur nel	1					
								se lanel	1					
								sulenelle	1					
								sur le mile	1					
								sulanil	1					
								soanne	1					

										ce la neil	1					
										celanul	1					
										celannelle	1					
										celenle	1					
										ceulanele	1					
										celemanle	1					
										salanel	1					
										cont	1					
										s la naire	1					
										sans laine	1					
										salennele	1					
										saunelel	1					
										se elani	1					
										solénnel	1					
										sollenelle		1				
										sonalell		1				
total	orthographe correcte	18	45	phonographie acceptable	201	28	phonographie approximante	2	0	phonographie non acceptable	137	14	24	0	3	0

TOMATE																
	orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	tomate	237	39	tomat	10					tomato	3		15	0	0	0
	tomates	143	48	tomatte	5					tomade	3					
				tommate	4					toumatte	2					
				thomate	3					tomote	2					
				taumatte	1					toman	1					
				tommatte	1					tomt	1					
				thomat	1					tomant	1					

				tomats	1					toma	1					
										tomahe	1					
										tomâtas	1					
										tomatate	1					
										tonates	1					
										tomatos	1					
										tommade	1					
										tomete	1					
										tomati	1					
										temate	1					
										toumat	1					
										toumate	1					
										toumata	1					
total	orthographe correcte	380	87	phonographie acceptable	26	0	phonographie approximante	0	0	phonographie non acceptable	26	0	15	0	0	0

TROIS

		orthographe correcte			phonographie acceptable			phonographie approximante			phonographie non acceptable			non-réponse		illisible	
		graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	graphies	bas	haut	bas	haut	bas	haut
graphies	trois	145			troi	6					tois	3		23		1	
	3	258			troix	5					tros	1					
					troie	1					turo	1					
					troit	1					trox	1					
					troies	1											
total	orthographe correcte	403	0	phonographie acceptable	14	0	phonographie approximante	0	0	phonographie non acceptable	6	0	23	0	1	0	

Annexe 3 bis : Classement des graphies des enquêtés « haut » relevées dans le « rayon bricolage »²¹³

ALLUMETTES						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphies	occ.	graphies	occ.	graphies	occ.
graphies	allumettes	58	allumette	15	alumettes	8
					alumette	4
					allumetts	1
					non-réponse	1
total		58		15		13
TOTAL ²¹⁴	86					

ARRACHER						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	arracher	75	arraché	4	aracher	4
			arrachées	1	accrocher	1
					non-réponse	2
total		75		5		5
TOTAL	85					

ARROSAGE							
	ortho conforme/ morpho restituée		ortho conforme/ morpho non restituée		ortho non conforme		non dicté ²¹⁵
	graphies	occ.	graphies	occ.	graphies	occ.	occ.
graphies	arrosage	65			arrossage	1	16
					arosage	3	
					arossage	1	
					orrosage	1	
total		65			6	16	
TOTAL	71						

²¹³ Les mots sont classés par ordre alphabétique.

²¹⁴ Le TOTAL représente le nombre de graphies recueillies.

²¹⁵ Ce mot fait partie de ceux qui n'ont pas été dictés par un des enquêteurs.

BOIS						
	ortho conforme /morpho restituée		ortho conforme/ morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	bois	86			Boi	1
total		86				1
TOTAL	87					

BOITE						
	ortho conforme/morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphies	occ.	graphies	occ.	graphies	occ.
graphies	boite	38			B	1
	boîte	45			bte	1
					boitte	1
					Non-réponse	1
total		83				3
TOTAL	86					

BRICOLAGE						
	ortho conforme/morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphies	occ.	graphies	occ.	graphies	occ.
graphies	bricolage	85	bricolages	1	bricolage	1
total		85		1		1
TOTAL	87					

CISEAU						
	ortho conforme/morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	ciseau	65	ciseaux	14	siceau	1
					sciseaux	1
					sicau	1
					siseaux	1
					sciseau	1
					siseau	2
					ciseau	1
total		65		14		8
TOTAL	87					

CLOUS						
	ortho conforme/morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	clous	78			cloux	3
					clout	1
					clouts	2
					claufs	1
					Non-réponse	2
total		78				7
TOTAL	85					

DÉGRAISSENT						
	ortho conforme/morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	dégraissent	63	DÉGRAISSE	6	DEGRASSE	1
	degraissent	7	degraisse	7	dégrèsse	1
					degressent	1
					décraisse	1
total		70		13		4
TOTAL	87					

MÈTRES							
	ortho conforme/morpho restituée		ortho conforme/morpho non restituée		ortho non conforme		non dicté
	graphie	occ.	graphie	occ.	graphie	occ.	occ.
graphies	mètres	41	mètre	4	mêtres	6	16
	metres	11	metre	1	mêtre	1	
	mètres	2	mètre	2	mettre	3	
total		54		7		10	16
TOTAL	71						

NORMAL						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	normal	83			normale	2
					droit	1
					Non-réponse	1
total		83				3
TOTAL	86					

PAQUETS						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	paquets	80	paquet	4	Poqué	1
					pacquets	1
					packets	1
total		80		4		3
TOTAL	87					

PARFUMÉS						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	parfumés	75	Parfumé	8	PAFURME	1
			parfume	1	porfumés	1
			parfumer	1		
total		75		10		2
TOTAL	87					

PEINTURE						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	peinture	83	peintures	2	pinture	2
total		83		2		2
TOTAL	87					

PINCE						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	pince	86			Non-réponse	1
total		86				1
TOTAL	86					

PLATES						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	plates	65	plate	13	plâtes	2
					PlatTE	3
					plâte	1
					plats	2
					Plattes	1
total		65		13		9
TOTAL	87					

POT						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	pot	82			pôt	5
total		82				5
TOTAL	87					

PRODUITS						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	produits	78	produit	7	prodduits	1
					prodt	1
total		78		7		2
TOTAL	87					

QUELQUES							
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme		non dicté
	graphie	occ.	graphie	occ.	graphie	occ.	occ.
graphies	quelques	59	quelque	10	quels	1	16
					quel que	1	
total		59		10		2	16
TOTAL	71						

TUYAU							
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme		non dicté
	graphie	occ.	graphie	occ.	graphie	occ.	occ.
graphies	tuyau	49	tuyaux	16	tuio	1	16
					tuyo	1	
					tuyeau	3	
					tyaux	1	
total		49		16		6	16
TOTAL	71						

VIS						
	ortho conforme/ morpho restituée		ortho conforme/morpho non restituée		ortho non conforme	
	graphie	occ.	graphie	occ.	graphie	occ.
graphies	vis	82			vices	3
					visses	1
					Visse	1
total		82				5
TOTAL	87					

Annexe 4 : Projet Compétences orthographiques des adultes et contextes sociaux d'usage

Dossier de candidature – présentation du projet

Candidat
<p>Brissaud Catherine Fonction actuelle : Professeure des universités Institution de rattachement : Université Joseph Fourier Laboratoire de rattachement : Laboratoire <i>LIDILEM</i>, EA 609 Adresse personnelle : 20 bis rue Wilson, 38610 Gières Tel professionnel : 04 56 52 04 99 – portable : 06 49 62 00 27 Courriel : catherine.brissaud@ujf-grenoble.fr</p> <p>MINI CV du responsable du projet</p> <p>Catherine Brissaud est professeure de Sciences du langage à l'université Joseph Fourier – ESPE de Grenoble où elle est responsable, pour les quatre universités de l'académie de Grenoble, du master <i>Métiers de l'enseignement scolaire</i> (professorat des écoles). Ses travaux de recherche portent sur l'acquisition et l'enseignement du français écrit : acquisition de la morphologie écrite du verbe ; didactique de l'orthographe et de la grammaire ; usages écrits des langues. Elle est responsable, avec François Mangenot, de l'axe « Didactique des langues et ingénierie éducative » du laboratoire Lidilem. Elle a été chercheuse invitée à l'université de Montréal en 2010 (un mois), à l'université d'Oxford en 2009 (3 mois), elle a passé une année post-doctorale à l'université d'Oxford Brookes (2001-2002). Elle a enseigné le français et les Lettres au lycée et en collège de 1980 à 1998.</p> <p>Elle est titulaire d'une habilitation à diriger des recherches (<i>Acquisition et didactique de l'orthographe</i>), d'un doctorat de Sciences du langage de l'Université Stendhal - Grenoble 3 (<i>Acquisition de l'orthographe du verbe au collège : le cas des formes en /E/. Invariants et procédures</i>), d'un DEA « Linguistique et phonétique », d'une maîtrise de français langue étrangère et d'une maîtrise de Lettres Modernes, d'un certificat européen d'anglais des affaires, d'un CAPES de Lettres Modernes et d'un baccalauréat de série scientifique.</p> <p>Elle a été rédactrice en chef de la revue <i>Lidil</i> de 2006 à 2012. Elle est membre du comité de rédaction de la revue <i>Repères</i> depuis 2012.</p> <p>Elle s'intéresse aux standards en matière de langue adoptés par différents systèmes d'éducation et à l'acquisition comparée des langues en milieu scolaire. Elle participe actuellement au projet <i>Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages</i> du groupe de travail « Lire-écrire à l'école primaire », projet initié par l'Institut français de l'Éducation, coordonné par Roland Goigoux. Elle entreprend un travail sur l'orthographe des personnes en difficulté avec l'écrit.</p>

Activités de recherche

Thèmes de recherche : acquisition et didactique du français écrit.

1/ modélisation de l'acquisition de la morphologie écrite du verbe ;

2/ acquisition de l'écriture : approches comparatives ;

3/ didactique de l'écrit, de l'orthographe et de la grammaire.

Activités d'enseignement

Depuis mon recrutement à l'IUFM de Grenoble (2002), d'une part je prépare les étudiants aux concours de recrutement de professeurs des écoles (épreuves de français), d'autre part j'assure la formation professionnelle des étudiants de master se destinant au professorat des écoles ou des professeurs stagiaires des écoles (didactique de la langue et de l'écriture).

J'encadre tous les ans des mémoires professionnels de professeurs stagiaires. J'assure depuis 2010 un cours intitulé « Connaissances linguistiques pour enseigner », pour des étudiants de première année de master *Métiers de l'enseignement scolaire*.

Depuis 2004, j'assure à l'université Stendhal-Grenoble 3 un cours sur l'écriture et la production d'écrits (Master 2 Sciences du langage, spécialité « Sociolinguistique et développement langagier » et depuis 2011, Master 2 Sciences du langage, spécialité « Didactique du français »)

Les cinq publications récentes les plus significatives

• Mout, T. et Brissaud, C. (2013). « Acquisition de la morphographie flexionnelle du français : retard de jeunes adultes en formation professionnelle », *Journal of French Language Studies*.

• Brissaud, C., Negro, I, Fisher, C (2012). The relation between spelling and pronunciation: the case of French /E/ verbal endings, *Written Language and Literacy*, 15.1., 46-64.

• Brissaud, C., Chevrot, J.-P. (2011). Acquisition of a major difficulty of French inflectional orthography: the homophonic /E/ verbal endings, *Writing Systems Research*, 3, n2: 129-144

• Brissaud, C. (2011). Didactique de l'orthographe : avancées ou piétinements ? *Pratiques*, 149-150, 207-226.

• Brissaud, C., Cogis, D. (2011) *Comment enseigner l'orthographe aujourd'hui ?* Paris: Hatier.

Qualifications

Doctorat en Sciences du langage (1998)

Habilitation à diriger des recherches (2008)

Recrutement en qualité de professeure à l'université Joseph Fourier (2009).

Composition de l'équipe
<p>Carole Blondel Fonction : doctorante (sous la direction de Catherine Brissaud et Fanny Rinck) Institution de rattachement : Université Stendhal Laboratoire de rattachement : Laboratoire <i>LIDILEM</i>, EA 609 Adresse personnelle : 29 chemin des Ebavous, 38660 La Terrasse Tel portable : 06 15 56 13 10 Courriel : cblondel38@yahoo.fr Publication :</p> <ul style="list-style-type: none"> • Boch, F., Buson, L., Blondel, C. (2012). Orthographe et grammaire à l'université. Quels besoins ? Quelles démarches pédagogiques ? in M.-C. Pollet, <i>De la maîtrise du français aux littéracies dans l'enseignement supérieur</i>, Presses Universitaires de Namur, collection Diptyque, 139-153.
<p>Fanny Rinck Fonction : maître de conférences Institution de rattachement : Université Joseph Fourier Laboratoire de rattachement : Laboratoire <i>LIDILEM</i>, EA 609 Adresse personnelle : 10 rue René Thomas, 38000 Grenoble Tel portable : 06 85 21 87 55 Courriel : fanny.rinck@ujf-grenoble.fr Publications récentes en lien avec le projet :</p> <ul style="list-style-type: none"> • Rinck F., Quintiliano Guimarães Silva J. et Assis J. Alves (2012). Quel cadre pour penser les pratiques de lecture et d'écriture dans la formation universitaire et/ou dans la vie professionnelle? <i>Revista Scripta</i>, 16(30), 16-25. Traduit en portugais : Qual abordagem erigir para pensar as praticas de leitura e escrita na formação acadêmica e/ou na vida profissional?, même numéro, 7-15. • Rinck F. et Sitri F. (2012). Pour une formation linguistique aux écrits professionnels, <i>Pratiques</i>, 153-154, 71-83.

Rattachement institutionnel du projet
<p>Laboratoire d'accueil du projet : LIDILEM, EA 609</p> <p>Institution et adresse de l'institution : Université Stendhal Grenoble III, Direction de la recherche BP 25 - 38040 Grenoble cedex 9 Tél. : +33 (0) 4 76 82 43 74 ; Fax : +33 (0) 4 76 82 43 95</p>

Projet de recherche

Titre du projet : *Compétences orthographiques des adultes et contextes sociaux d'usage*

Résumé : L'enquête porte sur les compétences orthographiques de base de la population résidant dans la région Rhône-Alpes. Elle consiste à tester des personnes qui n'ont pas passé le module ANLCl en 2011, à leur proposer la dictée augmentée de quelques items, à leur poser des questions sur leur utilisation de l'écrit, notamment électronique, et à les interroger sur leur rapport à l'écrit et à l'orthographe.

Cette enquête devrait permettre :

1. de mieux connaître les compétences orthographiques de base de la population adulte « ordinaire », non réduite aux personnes en difficulté à l'écrit et, par comparaison, de mieux évaluer la particularité dans ce domaine des personnes jugées en difficulté à l'écrit ;
2. d'évaluer la pertinence de la dictée pour mesurer les compétences en orthographe des personnes en difficulté à l'écrit. Les items ajoutés à la dictée (rayon « bricolage ») ont été choisis pour leur complémentarité avec ceux déjà testés (autres cas d'accord, consonnes doubles). Ces items pourront selon leur pertinence être repris dans de futures enquêtes.
3. de compléter le module biographique en interrogeant les personnes sur leurs pratiques personnelles de communication électronique (lecture et écriture), via les réseaux sociaux, les téléphones portables (SMS), les chats, le courrier électronique, les forums, etc.
4. de recueillir en parallèle des représentations de l'orthographe, de ses difficultés, de son apprentissage, de son importance sociale.

Exposé du projet

Le projet *Compétences orthographiques des adultes et contextes sociaux d'usage* s'inscrit dans un projet financé par la région Rhône-Alpes qui nous a accordé une allocation doctorale pour 2013-2016 (titre : *Capacités d'écriture des personnes en difficulté avec l'écrit : description linguistique et contextes sociaux d'usage*).

L'objectif de ce projet est de mieux comprendre le fonctionnement de l'orthographe du français pratiquée par les personnes en difficulté avec l'écrit dans une société où la maîtrise de l'écriture est plus que jamais nécessaire, notamment en situation de travail. Alors que la relation entre exclusion sociale et difficulté en littéracie est bien documentée, la recherche a en effet prêté peu d'attention aux capacités d'écriture des personnes dites « illettrées ».

Le travail consistera à décrire finement, d'un point de vue linguistique, les écritures ordinaires produites par les scripteurs les plus en difficulté avec l'écrit. Il s'agira aussi de mettre en relation leurs choix graphiques avec les résultats des tests de lecture et les réponses au module biographique du questionnaire (parcours scolaire, parcours professionnel, habitudes de lecture et d'écriture, langues parlées ou écrites, origine géographique, etc.). Un autre volet consistera à développer des stratégies d'intervention adaptées.

Toutefois, la tâche de production de mots qui nous intéresse – la seule qui permette d'observer des compétences orthographiques – est incluse seulement dans les

exercices simples. En d'autres termes, les données recueillies concernent seulement les personnes les plus en difficulté à l'écrit, soit 16 % des personnes interrogées au niveau national et 13 % des personnes interrogées en Rhône-Alpes.

Nous souhaiterions donc profiter de l'enquête IVQ pour étendre ce corpus. Il nous semble en effet important, pour évaluer les compétences en orthographe des personnes les plus en difficulté, de les comparer avec celles des personnes qui ont réussi les exercices d'orientation. En leur faisant passer l'épreuve de production écrite, nous aurons une idée des compétences et pratiques courantes en la matière. Cela nous permettra d'évaluer les personnes en difficulté non pas par rapport à la norme académique mais par rapport à des usages attestés de l'orthographe dans la population « ordinaire ». Le premier objectif de cette enquête complémentaire est donc de mieux caractériser, du point de vue qualitatif, les pratiques orthographiques des personnes en difficulté à l'écrit en Rhône-Alpes.

Du point de vue de l'analyse des données, de la même façon que nous le ferons à partir du corpus de productions écrites déjà recueilli en 2011, nous souhaitons non seulement exploiter les scores obtenus à l'épreuve et à chaque item, mais aussi décrire et analyser la diversité des réponses fournies : la diversité des bonnes réponses correctes pour les non-mots, la diversité des réponses non conformes à la norme pour les autres items. Nous commencerons par isoler les erreurs portant sur la transcription phonographique et celles portant sur la part sémiographique de l'orthographe (écrire *toumate* ou *taumate* par exemple ne nous dit pas la même chose des compétences en écriture et en orthographe) ; nous affinerons ensuite, en fonction du corpus, les catégories d'analyse. Cette grille d'analyse sera élaborée en collaboration avec l'équipe de Clara Mortamet à Rouen.

Mais nous souhaitons également profiter de cette enquête complémentaire pour recueillir de nouvelles données, en augmentant l'épreuve de production écrite, et en réinterrogeant les personnes sur quelques points relatifs aux pratiques de l'écrit, notamment électronique, et aux représentations de l'orthographe.

En ce qui concerne l'épreuve de production écrite (PE), il nous a semblé important de compléter la liste de départ pour évaluer davantage de points d'orthographe. Il s'agit en particulier :

- d'accords en nombre : il n'y avait qu'un élément de la liste comportant un accord verbal dans l'épreuve PE de l'enquête IVQ, et trois pluriels nominaux dont deux sans déterminant (*trois fromages, cerises, tomates*). Nous avons souhaité tester d'autres cas, tels qu'avec un adjectif qualificatif en -é (*parfumé*), avec une forme verbale homophone au singulier et au pluriel (*dégraisse/ dégraisent*), des pluriels dans des compléments du nom (*quatre paquets de vis plates, une boîte d'allumettes*), les déterminants *quelques* et *quatre* ; nous avons donc ajouté un pluriel verbal et six nominaux ;
- d'accords en genre, qui étaient absents dans l'épreuve IVQ 2011 (*vis plates*) ;
- de gémées dans différents cas de figure (*allumettes, arracher, arrosage*) ;
- de morphogrammes lexicaux (*bois, pot* qui s'ajoutent à *pays* et *fort*, mais aussi *clou* qui n'en a pas) ;
- des mots qui nous semblent intéressants du point de vue phonographique (digrammes et trigrammes communs et cas particuliers) : *ciseau, pince, dégraisent, boîte, peinture, bois, clous, arracher, vis, tuyau* ;
- des mots intéressants aussi pour leurs finales : *paquet, produit, normal, peinture*.

Ces items ont par ailleurs été choisis pour leur référence à des domaines à priori plus masculins : en ajoutant un rayon « bricolage » à la liste de courses établie, nous faisons l'hypothèse de réduire d'éventuels écarts d'implication dans l'exercice entre hommes et femmes.

Nous veillerons, dans l'analyse des données, à tirer toutes les conclusions utiles quant à la pertinence de chaque item interrogé au regard de l'ensemble (redondances, items trop difficiles ou trop faciles, etc.), mais aussi d'évaluer globalement la pertinence de l'épreuve pour caractériser les personnes les plus en difficulté. Cela nourrira des réflexions quant à l'élaboration d'enquêtes ultérieures.

Enfin, nous souhaitons profiter de ces enquêtes pour interroger des personnes sur leurs pratiques d'écriture au quotidien, en particulier leurs pratiques de communication électronique. Le module biographique n'interroge en effet que les usages de l'informatique au travail (BC32-BC36), mais pas les usages privés. Il ne permet pas non plus d'examiner dans le type de textes écrits : rapports, courriers officiels, mails, forums, *chat*, textos. Nous ajouterons aussi des questions concernant les doutes, les difficultés rencontrées et les moyens que mettent en œuvre les personnes pour les résoudre. Ces questions ont en effet été abordées seulement dans le module « débrouille » du questionnaire biographique (BG 1-3 et BG7-9), que n'ont pas passé les personnes qui ont réussi les tests d'orientation.

Enfin, nous souhaitons interroger les personnes, en les enregistrant cette fois si cela s'avère possible, sur leurs représentations de l'orthographe à travers la question de son apprentissage, de sa valeur sociale, des difficultés qu'elle pose.

Là encore, l'examen des réponses à l'ensemble de ces questions supplémentaires permettra d'éclairer de futures enquêtes.

Si l'on ajoute à ces enquêtes celles prévues par Clara Mortamet dans la région Haute Normandie (laboratoire Dysola, financement Région Haute Normandie, enquêtes menées également via CDA), nous pourrions apporter un éclairage important sur les productions écrites de l'enquête IVQ. L'intérêt en outre du travail concerté et parallèle de nos deux équipes serait de faire passer en plus de la production écrite du protocole IVQ les mêmes épreuves orthographiques complémentaires et les mêmes questionnaires.

Critères de sélection des enquêtés

Nous souhaitons réinterroger les personnes résidant dans la région Rhône-Alpes qui n'ont pas passé le module ANLCI, qui ont donc été orientées vers le module Haut, et ne présentent pas de difficulté manifeste dans les domaines fondamentaux de l'écrit. Il s'agit donc des variables **IMHA=1** et **IMHB=1**.

Afin de réduire le coût de chaque réinterrogation, et donc d'augmenter leur nombre, nous souhaitons privilégier les personnes résidant en agglomération (agglomération de Lyon et Grenoble).

Il s'agit donc des codes suivants :

REG = 82

DEP=69 et 38 en priorité

TUU2010= 5, 6, 7 (agglomérations de Lyon et de Grenoble)

Nous demandons au minimum 200 adresses.

Comme nous craignons de « perdre » un nombre important d'informateurs et donc de ne pas bénéficier pleinement du soutien offert par la DGLFLF si nous n'avons pas assez d'adresses, nous souhaiterions disposer de la totalité des fiches possibles pour la Région Rhône-Alpes (362 ayant accepté d'être réinterrogé moins 39 ayant déjà passé le module IVQ, soit 323 fiches disponibles).

Calendrier de réalisation de l'enquête qualitative

Les financements étant déjà obtenus pour la passation par une entreprise spécialisée, nous envisageons de les faire réaliser le plus tôt possible, dès le premier semestre 2014.

Dès la réception des données, les analyses seront menées par la doctorante, Carole Blondel, dont la thèse portant sur l'orthographe des personnes en difficulté avec l'écrit est financée par la région Rhône-Alpes, sous la supervision de Catherine Brissaud et Fanny Rinck.

Les cadres d'analyses et les priorités seront fixés en partenariat avec Clara Mortamet du laboratoire Dysola (Rouen), qui mènera le même questionnaire dans la région Haute-Normandie.

Nous espérons pouvoir dégager les premiers résultats d'analyse à la fin de l'année 2014.

Description des moyens humains et financiers disponibles pour la réalisation de l'enquête

Ce projet sera financé 1/ par la DGLFLF, qui nous a accordé 10 000 euros pour la passation de cette enquête complémentaire ; 2/ par la région Rhône-Alpes, qui finance l'allocation de thèse du doctorant qui traitera les données.

L'enquête portera sur des personnes résidant dans la région Rhône-Alpes et sera assurée par une entreprise spécialisée dans la réalisation d'enquêtes de ce type et qui connaît déjà le protocole IVQ. L'entreprise Catherine Delannoy et associés (CDA) propose, pour notre budget de 10 000 euros, de réaliser entre 150 et 200 réinterrogations.

Les transcriptions et l'analyse seront effectuées par la doctorante.

Grille d'entretien semi-directif détaillée

Production écrite : dictée ANLCI + items supplémentaires (en gras)

(rayon) Pharmacie

Anti-rhume

Micatole

Ou sirape

Alcool à 90° DUXE

(rayon) Epicerie

Tomates de pays

Confiture

Cerises

Sel Gobar
Trois fromages qui ne sentent pas fort
(rayon) Librairie
Le fiancé solennel de JOC PEVANORE

(rayon) Bricolage
deux produits parfumés qui dégraissent bien
quatre paquets de vis plates
un ciseau à bois normal
une boîte d'allumettes
une pince pour arracher les clous
un pot de peinture
quelques mètres de tuyau d'arrosage

Questions complémentaires (entretien oral)

[écriture à la maison]

Q1 - Écrivez-vous sur papier ?

Q2 - Quoi ? (y compris les listes (de courses), la copie de textes, les courriers divers, les messages laissés sur la table, etc.)

[ordinateur et internet à la maison]

Q3 - Avez-vous un ou plusieurs ordinateurs à la maison ? (précisez le nombre)

Q4 - Si oui l'utilisez-vous :

1. Très souvent (plusieurs fois par jour) /
2. Souvent (1 fois par jour) /
3. De temps en temps (1 ou 2 fois par semaine) /
4. Rarement (moins d'1 fois par semaine) /
5. Jamais

Q5 - Sur internet, vous consultez ou vous écrivez ?

1. Vous consultez
2. Vous écrivez
3. Les deux

Q6 - Pour quoi écrivez-vous ? (mail, forums, réseaux sociaux...)

Q7 - À quelle fréquence écrivez-vous ?

1. Très souvent (plusieurs fois par jour) /
2. Souvent (1 fois par jour) /
3. De temps en temps (1 ou 2 fois par semaine) /
4. Rarement (moins d'1 fois par semaine) /
5. Jamais

Q8 - Utilisez-vous un ordinateur au travail ?

Q9 - Si oui à quelle fréquence ?

1. Très souvent (plusieurs fois par jour) /
2. Souvent (1 fois par jour) /
3. De temps en temps (1 ou 2 fois par semaine) /
4. Rarement (moins d'1 fois par semaine) /
5. Jamais

Q10 - Pour quelle utilisation ? (logiciels (lesquels ?), messagerie, etc.) ?

[téléphone portable et SMS]

Q11 - Avez-vous un téléphone portable ?

Q12 - Utilisez-vous les SMS (textos) ?

Q13 - Si oui, à quelle fréquence ?

1. Très souvent (plusieurs fois par jour) /
2. Souvent (1 fois par jour) /
3. De temps en temps (1 ou 2 fois par semaine) /
4. Rarement (moins d'1 fois par semaine) /

5. Jamais

Q14 – Utilisez-vous l'écriture intuitive (ou T9) ? les abréviations ? le langage *texto* ?

[Orthographe]

Q15 - Avez-vous des doutes, des difficultés, des hésitations sur la façon d'écrire un mot ?

1. oui, très souvent /
2. oui, assez souvent /
3. non, pas très souvent /
4. non, jamais

Q16 - Cherchez-vous à écrire sans faute ?

1. oui, toujours
2. Ca dépend de ce que j'écris (ou à qui j'écris)
3. non, jamais

Q17 - Est-ce qu'il y a des points particuliers sur lesquels vous hésitez souvent ?

Q18 - Si oui, lesquels ?

Q19 - Cherchez-vous de l'aide ?

Q20 - Si oui, quels outils utilisez-vous? (proposer une liste : dictionnaire, demande à une autre personne, correcteur automatique des logiciels, autre (précisez))

Q21 - Est-ce facile pour vous de ne pas faire de fautes ?

[apprendre l'orthographe] *début enregistrement*

1/ Quels sont les trois premiers mots qui vous viennent à l'esprit pour qualifier l'orthographe ?

2/ Qu'est-ce que vous pensez de l'orthographe du français ?

3/ Vous vous souvenez de votre apprentissage de l'orthographe ? Pour vous, c'est facile d'apprendre l'orthographe ?

4/ Avez-vous eu l'occasion d'aider des personnes à apprendre l'orthographe (aide aux devoirs par exemple, ou conseil à des amis, etc.) qui ? des enfants pour l'école ? des adultes ?

5/ Que pensez-vous des personnes qui font des erreurs d'orthographe ?

6/ Quand vous écrivez, est-ce qu'il y a d'autres choses que l'orthographe qui sont difficiles pour vous ?

Annexe 5 : Avis d'examen du Comité du label de la Statistique Publique



COMITÉ DU LABEL
DE LA STATISTIQUE PUBLIQUE

Paris, le 7 janvier 2014 – N°12 /Label/L201

AVIS D'EXAMEN

Post-enquêtes qualitatives associées à l'enquête Information et Vie Quotidienne 2011

Service producteur : Insee – Direction des statistiques démographiques et sociales (Dsds)
Réunion du Comité du label de la statistique publique du 28 novembre 2013 (commission Ménages)

L'Insee a collecté en 2011 l'enquête Information et Vie Quotidienne (IVQ) en France métropolitaine. L'enquête s'intéresse spécifiquement à la mesure des compétences des adultes en compréhension orale, à l'écrit et en calcul, et au repérage des personnes en difficulté dans ces trois domaines.

Pour son édition 2011, l'enquête proposait de nouveaux modules d'interrogation (module Haut B et module commun) et une refonte partielle du module biographique, destinés à fournir des scores continus de compétence à l'écrit et à mieux cerner les facteurs explicatifs et les conséquences de la plus ou moins bonne maîtrise des compétences dans les trois domaines évalués.

Près de 14 000 personnes ont été interrogées pour l'enquête dont environ 2 000 ont été orientées vers le modules ANLCI. La possibilité de réaliser des post-enquêtes qualitatives est prévue dans un délai maximum de trois années après la fin de la collecte (fin janvier 2015). Elles porteront chacune sur un échantillon restreint à une centaine de personnes ayant donné leur accord pour être à nouveau contactées.

Les objectifs des post-enquêtes qualitatives associées à l'enquête IVQ 2011 auront une visée méthodologique ou permettront d'approfondir un thème abordé dans le questionnaire quantitatif. Elles peuvent par exemple évaluer la capacité des questions à renseigner sur les phénomènes étudiés, souvent complexes et difficiles à saisir dans un questionnaire fermé.

Les projets peuvent viser l'élaboration de propositions argumentées de méthodes d'enregistrement alternatives. Ils peuvent encore s'intéresser à une population particulière qui ne fait en général pas l'objet d'enquête spécifique, mais qui est couverte par l'enquête (jeunes adultes, personnes en situation d'illettrisme...). Enfin, les projets tireront bénéfice de l'approche qualitative qui permet un approfondissement des thématiques et s'attacheront à en présenter les enjeux et le type de connaissances attendues.

Trois dossiers de candidatures ont répondu aux critères exprimés dans l'appel à projet :

Projet n°1

Rattachement institutionnel du projet : Laboratoire de rattachement : identités et territoires des populations (INED, UR12), Institut national des études démographiques.

Objet : parcours, difficultés face à l'écrit, et pratiques linguistiques pour les personnes socialisées dans une langue issue de l'immigration ou régionale.

Nombre d'entretiens prévus : 30 à 50 entretiens.

Localisation : régions Nord, Normandie, Ile-de-France, Alsace et Rhône-Alpes.

RÉPUBLIQUE FRANÇAISE

Secrétariat général du Cnis : Timbre D030 - 18, Bd. A. Pinard 75675 PARIS Cedex 14 - Tél. : 01 41 17 52 62 - Fax : 01 41 17 55 41 - secretariat-general@cnis.fr - www.cnis.fr

Projet n°2

Rattachement institutionnel du projet : LIDILEM, EA 809 – Université Stendhal Grenoble III, Direction de la recherche

Objet : Compétences orthographiques des adultes et contextes sociaux d'usage

Nombre d'entretiens prévus : minimum 200 adresses

Localisation : Rhône-Alpes

Projet n°3

Rattachement institutionnel du projet : Laboratoire Dysola (Dynamiques sociales et langagières), Université de Rouen

Objet : compétences de base en orthographe, pratiques et représentations de l'orthographe en Normandie

Nombre d'entretiens prévus : 150 et 200 entretiens

Localisation : Haute-Normandie

Le Comité du label rappelle au service les procédures à mener :

- la déclaration à la CNIL,
- l'information auprès du CNIS au printemps 2014,
- le passage des prestataires devant le Comité du secret (après vérification auprès de l'Unité des affaires juridiques et contentieuses que cette procédure est bien nécessaire)
- la convention avec les chercheurs et les différents acteurs juridiques.

Il est important, pour la qualité des opérations de post-enquêtes, que celles-ci puissent se réaliser dans les meilleurs délais et, par suite, que ces formalités soient accomplies sans retard.

Accès et résultats aux chercheurs

Le Comité souhaite que les chercheurs puissent avoir accès aux données issues des post-enquêtes via la Banque de données sur les enquêtes qualitatives gérée au sein du Centre Quetelet. Par ailleurs, il faudra veiller à ce que les chercheurs aient également accès aux enregistrements de l'enquête source, sous réserve que ceux-ci aient été anonymisés.

Livraison des fichiers

Au moment de la livraison des fichiers aux chercheurs, les notices explicatives devront rappeler les précautions d'utilisation à respecter, notamment celles relatives au fait qu'il s'agit d'opérations qualitatives et non statistiques.

Calendrier

Compte-tenu du fait que les post-enquêtes auront lieu plus tardivement que l'enquête principale, il sera utile qu'une question préliminaire soit posée aux enquêtés pour savoir si leur situation a évolué depuis la réalisation de l'enquête principale, en particulier pour vérifier si ces personnes se sont formées pour améliorer leurs compétences (par exemple, dans la connaissance de la langue française).

Par ailleurs, les lettres-avis devront faire le lien avec l'enquête principale.

Livraison de l'échantillon

Il est suggéré, pour le projet n°2, d'élargir le champ géographique aux régions limitrophes à celles initialement prévues, de façon à disposer d'un vivier plus important et de procéder à la livraison de l'échantillon en deux temps (avec un échantillon de réserve). Les conditions de déclenchement de la livraison de l'échantillon devront être précisées dans la convention.

Le Comité du label attire l'attention sur les projets qui ont un champ géographique commun ; pour ceux-ci, il faut assurer une disjonction au sein des fiches-adresses.

Enfin, le service doit informer les directions régionales de l'Insee qu'elles devront assurer la livraison des fiches-adresses (FA) à une échéance fixée pour que la MOA puisse les mettre à disposition des chercheurs.

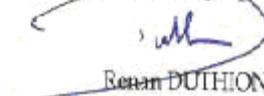
RÉPUBLIQUE FRANÇAISE

Secrétariat général du Cnis : Timbre D030 - 18, Bd. A. Pinard 75675 PARIS Cedex 14 – Tél. : 01 41 17 52 62 – Fax : 01 41 17 55 41 – secretariat-general@cnis.fr – www.cnis.fr

Le Comité du label de la statistique publique attribue un avis d'examen favorable aux post-enquêtes associées à l'enquête Information et Vie Quotidienne 2011.

Cet avis d'examen est valide pour les années 2014 et 2015

Le Président du Comité du label
de la statistique publique



Renaud DUTHION

RÉPUBLIQUE FRANÇAISE

Secrétariat général du Cris : Timbre D030 - 18, Bd. A. Pinard 75675 PARIS Cedex 14 - Tél. : 01 41 17 52 62 - Fax : 01 41 17 55 41 - secretariat-general@cris.fr - www.cris.fr

Annexe 6 : Avis d'examen du Conseil National du label de l'Information Statistique



**DIRECTION DE LA METHODOLOGIE ET DE LA
COORDINATION STATISTIQUE ET INTERNATIONALE**

Dossier suivi par :
Corinne Requier
Tél. : 01 41 17 54 15
Mél : corinne.requier@insee.fr

Responsable de commission :
Dominique Rouaut
Tél. : +33 (0)1 41 17 38 59
Fax : +33 (0)1 41 17 55 41
Mél : dominique.rouaut@insee.fr

Le Directeur de la méthodologie et de la coordination
statistique et internationale

à

Direction des statistiques démographiques et sociales
Division Emploi

A l'attention de Sylvie Le Minez

Paris, le 21 janvier 2014
N°797 / Visa / DG75-D001

Objet : Avis d'examen et numéro de visa

Je vous prie de bien vouloir trouver ci-joint l'avis d'examen relatif aux post-enquêtes qualitatives associées à l'enquête « information et vie quotidienne 2011 (IVQ) » qui a été examinée par le comité du label lors de sa réunion du 28 novembre 2013.

L'enquête a reçu le visa n°2014X074EC du Ministre de l'économie et des finances, valable pour les années 2014 et 2015.

Vous voudrez bien adresser par mél à secretariat-general@cnis.fr, le questionnaire revêtu du numéro de visa et tous les documents définitifs remis aux enquêtés.

Par ailleurs, nous vous rappelons que toute modification apportée à l'enquête (dont le nom) ne peut se faire qu'avec l'accord préalable du comité du label.

Le Directeur de la méthodologie et de la
coordination statistique et internationale

Philippe CUNEO

P.J. : Avis d'examen
Mentions à faire figurer sur le questionnaire

ENQUÊTE AYANT RECU UN AVIS D'EXAMEN DU COMITE DU LABEL

—
**Post-enquêtes qualitatives associées à
l'enquête « information et vie quotidienne (IVQ) »**

La mention suivante doit être inscrite en 1ère page du questionnaire et présentée encadrée.

Vu l'avis d'examen favorable du Conseil National de l'Information Statistique, les post-enquêtes qualitatives associées à l'enquête principale « information et vie quotidienne (IVQ) » qui a été reconnue d'intérêt général et de qualité statistique avec le caractère obligatoire.

Visa n°2014X074EC du Ministre de l'économie et des finances valable pour l'année 2014 et 2015.

En application de la loi n°51-711 du 7 juin 1951 modifiée, les réponses à ce questionnaire sont protégées par le secret statistique et destinées à (nom du service).

La loi n°78-17 du 6 janvier 1978 modifiée, relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites à la présente enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Ce droit peut être exercé auprès de



Annexe 7 : Courrier envoyé aux enquêtés « haut »



Service Statistique
Division Enquêtes Ménages

Dossier suivi par :
Gérard PICCIOLI
Tél. : 04 78 63 23 81
Mél : dr69-division-des-enquetes-menages@insee.fr

Lyon, le 12 septembre 2014
N° 652/DR69-SES/FL/TL

Objet : Étude complémentaire sur l'usage de l'information dans la vie quotidienne

Madame, Monsieur,

L'Institut National de la Statistique et des Études Économiques (INSEE) a réalisé en 2011 et 2012 une importante étude sur l'usage de l'information dans la vie quotidienne en France.

L'objectif était de recueillir des données précises sur le niveau de maîtrise des compétences de base en calcul, en compréhension orale et en compréhension écrite de la population adulte et sur l'utilisation de l'information dans la vie quotidienne (notices d'appareils ménagers, journaux, formulaires administratifs, etc.). Cette information, de plus en plus abondante et complexe, est-elle toujours bien comprise ? Faut-il revoir le système de formation pour en faciliter l'accès ?

Lors de votre participation à cette enquête vous aviez donné votre accord pour qu'un chercheur puisse reprendre contact avec vous. Une équipe de recherche de l'université Stendhal (Grenoble) entreprend une étude complémentaire sur l'utilisation de l'écrit dans la vie de tous les jours et souhaiterait recueillir, en complément de vos précédentes réponses, vos pratiques habituelles. Cet entretien durerait environ 30 minutes.

Madame X, enquêtrice, prendra contact avec vous. Elle sera munie d'une carte officielle de l'Insee.

Afin de faciliter la prise de rendez-vous, pourriez-vous transmettre votre numéro de téléphone ou précisant bien votre nom :

- par téléphone, au ...
- ou par courrier électronique à l'adresse suivante :

Votre participation à cette enquête est importante car elle apportera des éclairages nouveaux sur les pratiques quotidiennes de l'écrit.

Soyez assurés que comme la loi en fait la plus stricte obligation, vos réponses resteront anonymes et confidentielles.

Je vous remercie de l'accueil réservé à l'enquêtrice, et vous prie d'agréer Madame, Monsieur, l'expression de ma considération distinguée.

Le Directeur Régional

Pascal OGER

Vu l'avis d'examen du Conseil national de l'information statistique, cette enquête, a reçu le visa n°2014X074EC du Ministre de l'Économie, de l'Industrie et de l'emploi valable pour l'année 2014 et 2015.
L'utilisation par l'Insee des renseignements individuels collectés sera conforme à la loi du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques. Les dispositions de l'article 6 interdisent notamment toute utilisation à des fins de contrôle fiscal.
La loi n° 78-17 du 6 janvier 1978 modifiée, relative à l'informatique, aux fichiers et aux libertés, s'applique aux réponses faites à la présente enquête. Elle garantit aux personnes concernées un droit d'accès et de rectification pour les données les concernant. Ce droit peut être exercé auprès de l'Insee.

Annexe 8 : Protocole de l'enquête complémentaire, consignes aux enquêteurs

Enquête complémentaire à IVQ 2011

Juin 2014

Consignes aux enquêteurs

Objectif de l'enquête « Information et Vie Quotidienne »

Pourquoi une enquête complémentaire qui fait suite à IVQ 2011 ?

Cette enquête, conduite sur la région Rhône-Alpes et la région Haute-Normandie, porte sur les compétences orthographiques de base de la population résidant dans ces deux régions. Il s'agit de tester, en production écrite, des personnes qui ont passé l'enquête IVQ en 2011 et qui étaient volontaires pour être réinterrogées. Elles ne sont pas, a priori, en grande difficulté à l'écrit. Les personnes sélectionnées passeront une épreuve de production écrite et répondront à un questionnaire oral.

Sélection de l'échantillon

Vous trouverez sur la fiche adresse les informations suivantes :

- Nom/prénom de l'occupant principal du logement. Ce peut être l'enquêté mais pas forcément.
- Adresse
- Informations sur le logement : année de construction, type, surface, ...
- Profession en 2011 de l'occupant principal
- La taille du ménage
- L'année de naissance

Demandez pour chaque logement quelle est la personne qui a été interrogée lors de l'enquête INSEE IVQ 2011. Pour vous aider à vérifier qu'il s'agit bien de la bonne personne, utilisez les informations que nous vous transmettons sur l'enquêté, à savoir le sexe et l'année de naissance.

Le déroulement de l'interview

Durée de l'enquête

Nous comptons une vingtaine de minutes pour effectuer l'enquête complète.

Des questions sur la durée vous seront très souvent posées : **il faut répondre que l'enquête va durer environ 20 minutes.**

Pour ne pas s'exposer à un refus immédiat, il est préférable de ne pas décrire la forme très particulière des questions lorsque vous êtes en train de négocier l'accord pour l'entretien.

Cela ne veut pas dire que vous allez prendre au piège l'enquêté. S'il accepte l'entretien, il pourra très bien se raviser lorsque vous lui aurez expliqué les détails de son déroulement. De même, si l'enquêté demande plus de détails, il est hors de question de lui cacher la forme de l'enquête (une liste de courses à rédiger pour un livreur).

L'approche et la présentation générale de l'enquête

C'est un des moments-clés de l'enquête. Nous avons fait un effort pour bien fixer et standardiser cette phase. Aussi, nous avons préparé un texte, qui n'est pas prévu pour être répété textuellement. En revanche nous vous demandons d'insister sur chacune des idées, et notamment de bien prêter attention aux mots employés. Cette phase aura certainement lieu au moment de la première prise de contact, au téléphone ou de visu. Il faut donc présenter généralement l'enquête. Voici un modèle de phrase.

« C'est une enquête du laboratoire Lidilem/Dysola de l'Université de Grenoble/Rouen, en partenariat avec l'INSEE, qui fait suite à l'enquête sur l'Information dans la Vie Quotidienne qui vous a été proposée par l'INSEE en 2011. Cette enquête complémentaire vise à mieux connaître les usages de l'écrit des personnes interrogées, en particulier celles qui n'ont pas été repérées comme en difficulté lors de l'enquête réalisée en 2011. »

Précaution : Ne parlez jamais d'évaluation, d'épreuve, et surtout pas de dictée, c'est-à-dire de tout ce qui peut être relatif à une situation scolaire. Préférez le terme de questionnaire, comme pour une enquête classique.

Si vous avez besoin d'insister sur l'utilisation des résultats :

« Ils serviront à orienter plus efficacement l'action du système de formation des adultes (la formation tout au long de la vie prend de plus en plus d'importance) Ils serviront à rendre un peu plus concret les savoirs que l'on enseigne, à les adapter à la vie quotidienne bref rajouter du concret. »

Le matériel et la mise en scène

Pour l'épreuve de production écrite, il faut mettre à disposition de l'enquêté un stylo et une feuille de papier. Il faut recommander à l'enquêté de se mettre dans une position confortable et veiller à être soi-même dans une position suffisamment pratique pour débarrasser le matériel tranquillement. Le mieux est d'être sur une table.

Proposez une fiche avec un cadre pour éviter que les personnes écrivent trop près du bord de la feuille (dans le cas contraire, les photocopies risquent d'être illisibles).

Certains éléments sont à noter tels que les événements particuliers (ex : gêne, interruption etc.). Rédigez un commentaire expliquant la situation.

Précisez, le cas échéant, la lenteur d'écriture ou la difficulté manifeste des enquêtés lors de l'épreuve de production écrite.

Pour le questionnaire, commencez par les questions à saisir sur l'ordinateur (questions fermées pour lesquelles l'enquêté doit choisir une réponse parmi plusieurs propositions) puis passez aux questions suivantes (questions ouvertes) après avoir lancé l'enregistrement. Pour le questionnaire, il est possible que l'enquêté soit amené à regarder la question sur l'ordinateur : cela n'est pas gênant. Aucune information dévalorisante n'y figure.

Détails de l'enquête

Fiche d'identification de l'interviewé

Cette fiche est à remplir avant de commencer l'enquête. Les informations peuvent être fournies par l'interviewé ou avoir été pré-indiquées par l'INSEE : vérifier les informations dont on dispose et les corriger.

1^{ère} partie : Epreuve de production écrite

Il s'agit d'une liste de courses à écrire sur une feuille fournie par l'enquêteur.

2^{ème} partie : Le questionnaire

Le questionnaire est composé de questions fermées et de questions ouvertes. Toutes les réponses sont fournies oralement. Les premières questions sont fermées et doivent être saisies sur ordinateur. Les dernières questions sont ouvertes et doivent être enregistrées.

1) Épreuve de Production écrite

Protocole de la liste de courses

Avant de donner la feuille à l'enquêté, remplissez, au recto les cadres dans lesquels doivent être inscrits le nom, prénom et identifiant de l'enquêté, etc. Au début de cet exercice, vous distribuez la feuille (côté verso).

Donnez à l'enquêté la feuille B1 et un stylo et dites :

« À présent je vais vous demander d'écrire une liste de courses à remettre à un livreur de supermarché. Dans cette liste, il faut écrire le nom du rayon et les produits que l'on doit acheter. Vous allez entendre des mots que vous ne connaissez pas, ce sont des noms de marques que nous avons dû inventer car nous n'avons pas le droit de faire de la publicité. Écrivez-les comme vous pensez que cela doit s'écrire. »

Commencez à lire les mots que l'enquêté doit inscrire sur la feuille, en lui indiquant si besoin du doigt l'endroit où il doit écrire le mot que vous venez de lire. Lisez lentement et le plus distinctement possible en adaptant votre lecture au rythme d'écriture de l'enquêté. Répétez la phrase si nécessaire (il ne s'agit ni d'une épreuve de vitesse, ni de mémoire, mais

seulement d'analyse des compétences à écrire un message de façon compréhensible. ce n'est pas une épreuve de mémoire). Si l'enquêté semble ne pas bien comprendre les mots inventés, répétez-les-lui distinctement en articulant chaque syllabe et en lui rappelant que c'est une marque inventée. Vous pouvez aussi intervenir si l'enquêté oublie un mot par exemple « *je crois que vous n'avez pas écrit xxxx* ». Si l'enquêté répond que c'est parce qu'il ne sait pas l'écrire, vous pouvez lui demander d'essayer malgré tout, s'il refuse continuez sans commentaire. Vous pouvez intervenir avec bienveillance à chaque fois que vous remarquez que l'enquêté n'a pas bien compris ce qu'il avait à faire mais vous ne devez surtout pas l'aider à orthographier ou à accorder un mot. Si l'enquêté vous demande la bonne réponse, ne la lui fournissez pas. Dites-lui que vous ne la connaissez pas ou que c'est interdit. De manière générale, ne manifestez pas de réprobation ou d'acquiescement, ni par la voix ni par le geste ou l'attitude.

A la fin de l'épreuve, vous ramassez la liste de courses, inscrivez vos références d'enquêteurs et les autres informations demandées au recto. Les bons de commande seront déposés chez votre responsable.

Si l'enquêté en fin de passation vous demande s'il a réussi, vous lui répondrez qu'il a bien coopéré mais que vous ne connaissez pas son score, que de toute façon cela n'a pas de conséquence sur son travail, qu'il n'y aura pas de note, mais que c'était très important pour l'enquête qu'il réponde de son mieux. S'il insiste rassurez-le à la mesure de sa performance (néanmoins ne dites pas « *c'était très bien* » à quelqu'un qui n'a pas répondu à une seule question etc.).

Texte de l'épreuve de production écrite à lire à l'enquêté

Pharmacie

- Anti-rhume Micatol ou Sirape (*Micatol et Sirape, ce sont deux marques de produit contre le rhume. Ecrivez-les comme vous le pensez*)

- Alcool à 90° Duxe (*Duxe c'est la marque de l'alcool à 90°. Ecrivez-la comme vous le pensez*)

Épicerie

- Tomates de pays

- Confiture

- Cerises

- Sel Gobar (*Gobar, c'est la marque du sel. Ecrivez-la comme vous le pensez*)

- Trois fromages qui ne sentent pas fort (*Si la personne demande pour 'trois' en chiffres ou en lettres, répondez « comme vous voulez ». Oui, écrivez tout, « trois fromages qui ne sentent pas fort »*)

Librairie

- Le fiancé solennel de Joc Pévanor (*le titre du livre c'est « le fiancé solennel », l'auteur s'appelle « Joc Pévanor » vous devez écrire le titre et le nom de l'auteur*)

Bricolage

Deux produits parfumés qui dégraissent bien (Si la personne demande pour 'deux' en chiffres ou en lettres, répondez « comme vous voulez ». Oui, écrivez tout, « deux produits parfumés qui dégraissent bien »)

Quatre paquets de vis plates (Si la personne demande pour 'quatre' en chiffres ou en lettres, répondez « comme vous voulez »).

Un ciseau à bois normal

Une boîte d'allumettes

Une pince pour arracher les clous

Un pot de peinture

Quelques mètres de tuyau d'arrosage

2) Questionnaire

Le questionnaire comprend deux parties. L'une est composée de questions fermées (l'enquêté doit choisir parmi plusieurs modalités), l'autre de questions ouvertes. Les questions doivent être posées dans l'ordre. Cette partie est obligatoire.

Questions fermées : il suffit de cocher les réponses que l'enquêté vous a faites. Si la réponse est « autres », inscrivez la (ou les) réponse(s) dans la case prévue à cet effet (il peut y avoir plusieurs réponses).

Questions ouvertes : les questions ouvertes sont enregistrées. Veillez à la qualité de l'enregistrement : pas de télévision, radio, musique etc.

Détail des questions

1) Questions fermées

L'écriture à la maison

Q1 : Ecrivez-vous sur papier ?

- Non
- Oui

Q2 : Qu'écrivez-vous ?

- Des listes (ex. de courses) oui, non
- Des textes (y compris la copie de texte) oui, non
- Des courriers divers oui, non
- Des messages (ex. mots laissés sur la table) oui, non
- Autre (précisez)

Q3 : Est-ce que vous avez un ordinateur à la maison ?

- Non
- Oui,

Q3.1 : Plusieurs ? (y compris ordinateurs portables)

Q4 : A quelle fréquence utilisez-vous votre ordinateur à la maison :

- Très souvent (plusieurs fois par jour)
- Souvent (1 fois par jour)
- De temps en temps (1 ou 2 fois par semaine)
- Rarement (moins d'1 fois par semaine)
- Jamais

Q4.b : Avez-vous une tablette tactile ?

- Non
- Oui

Q4.c : Plusieurs ?

Q5 : Utilisez-vous Internet à la maison ?

- Non
- Oui

Q5.1 : Est-ce que vous consultez Internet ? Est-ce que vous écrivez ? Les 2 ?

- Vous consultez
- Vous écrivez
- Les 2
-

Q6 : Si vous écrivez, est-ce :

- Des mails ? oui, non
- sur les réseaux sociaux (Facebook, Twitter, LinkedIn, etc.) oui, non
- sur des Forums de discussions oui, non
- sur des Chats oui, non

Q7 : Si vous écrivez toujours, à quelle fréquence écrivez-vous sur Internet ?

- Très souvent (plusieurs fois par jour)
- Souvent (1 fois par jour)
- De temps en temps (1 ou 2 fois par semaine)
- Rarement (moins d'1 fois par semaine)

Ordinateur au travail

Q8 : Utilisez-vous un ordinateur au travail ?

- Non
- Oui

Q9 : Si oui, à quelle fréquence l'utilisez-vous :

- Très souvent (plusieurs fois par jour)
- Souvent (1 fois par jour)
- De temps en temps (1 ou 2 fois par semaine)
- Rarement (moins d'1 fois par semaine)

Q10 : Pour quelle utilisation ?

- Utilisation de logiciels oui, non, lesquels ?
- Messagerie électronique oui, non
- Navigation sur Internet oui, non
- Textes (rapports, compte-rendu, notes internes, etc.) oui, non
- Autres (précisez)

Orthographe :

Q11 : Vous arrive-t-il d'avoir des doutes, des difficultés, des hésitations sur la façon d'écrire un mot ?

- Oui, très souvent
- Oui, assez souvent
- Non, pas très souvent
- Non, jamais

Q12 : Cherchez-vous à écrire sans faute ?

- Oui, toujours
- Ca dépend de ce que j'écris ou à qui j'écris
- Non, jamais

Q13 : Est-ce qu'il y a des points particuliers d'orthographe sur lesquels vous hésitez souvent ?

- non
- oui,

Q14 : Si oui, lesquels ?

Laisser la personne déclarer spontanément ses points faibles, les lister.

Q15 : Vous arrive-t-il d'aller chercher de l'aide ?

- Non
- Oui

Q16 : Quelle type d'aide utilisez-vous :

- Demande d'aide à une personne

- Correcteur orthographique
- Manuels (bled etc.)
- Dictionnaires
- Autres (précisez)

Q17 : Finalement, est-ce facile pour vous de ne pas faire de fautes ?

- oui
- non

Téléphone portable

Q18 : Avez-vous un téléphone portable ?

- Non
- Oui :

Q19 : Ecrivez-vous des SMS (textos) ?

- Non
- Oui

Q20 : Si oui, à quelle fréquence écrivez-vous des SMS ?

- Très souvent (plusieurs fois par jour)
- Souvent (1 fois par jour)
- De temps en temps (1 ou 2 fois par semaine)
- Rarement (moins d'1 fois par semaine)

Q21 : Utilisez-vous l'écriture intuitive (T9) ? :

- Non
- Oui

Q21.1 : Ecrivez-vous avec des abréviations ?

- Non
- Oui

Q21.2 : Ecrivez-vous en langage texto ?

- Non
- Oui

Fin de la saisie et début de l'enregistrement

2) Questions ouvertes

Expliquez que vous allez lire les questions et que l'enquêté a le temps d'y répondre.

Q1 : Quels sont les trois premiers mots qui vous viennent à l'esprit si je vous dis « orthographe » ?

Annexe 9 : Transcription de l'entretien de l'enquêté « haut » n° 187

Transcription fichier audio - n° 187

Enquêteur 9

Commentaires : enquêté qui indique avoir travaillé au Ministère de la Défense, pour la formation en français des jeunes recrues illettrées, puis qui a été responsable du groupe de lutte contre l'illettrisme, avec Alain Bentolila.

Temps de l'entretien : 6mn28

Q1 : Quels sont les trois premiers mots qui vous viennent à l'esprit si je vous dis « orthographe » ?

- Etymologie
- Grammaire
- Vocabulaire

Q2 : Qu'est-ce que vous pensez de l'orthographe du français ?

« Que c'est une très belle langue qui malheureusement est oubliée, qui est mal appliquée et c'est sûrement la plus belle langue au monde, parce qu'elle a une telle déclinaison de possibilités, comment dire, pour définir un mot, pour définir un objet, on va avoir plusieurs possibilités aussi bien de nom commun que d'adjectif qualificatif, c'est ça qui est extraordinaire, c'est la plus belle langue au monde. »

Q3 : Est-ce que vous vous souvenez de votre apprentissage de l'orthographe ?

« Ah bien entendu, l'apprentissage de l'orthographe c'était au quai Jules Ferry, école maternelle, école primaire et qui était complété par mon grand-père tous les jeudis et puis après c'était tous les mercredis après-midi. J'allais chez lui et il sortait de vieux ouvrages et puis on faisait des dictées, on faisait des grammaires, voilà, c'est comme ça que j'ai appris à conjuguer et que j'ai appris à analyser les mots, hein, adjectif qualificatif épithète de ... masculin, féminin, etc., c'était fabuleux. »

- Vous avez peut être un bon ou un mauvais souvenir ?

« Ah c'est des excellents souvenirs, jusqu'en sixième j'étais le premier de la classe en calcul et en français et puis comme j'aimais beaucoup l'histoire et la géographie, ça allait bien. »

- Vous rappelez-vous d'exercices ou de leçons que vous faisiez ? De manuels que vous utilisiez ?

Question non posée

Q4 : Selon vous, est-ce facile d'apprendre l'orthographe du français ?

« Quand il est bien dispensé, oui. Il faut avoir devant soi des gens qui sont passionnés or aujourd'hui, l'éducation nationale s'est écartée de son chemin. On n'a plus affaire qu'à des syndiqués, je le sais

parce que les trois quarts de ma famille sont des enseignants avec qui on se fâche régulièrement parce qu'ils se sont détournés de leur métier, de leur foi, de leur profession. Aujourd'hui ça ne correspond plus à rien, c'est de la politique. On râle après le ministère qui fait des réformes mais personne ne s'efforce de faire son travail correctement, avec patience. Donc on pourrait parfaitement le faire, c'est très simple, je l'ai fait avec mes enfants, je suis pas enseignant, je l'ai fait avec mes enfants. L'aînée écrit parfaitement bien le français, aujourd'hui elle est courtier en assurances et l'autre qui continue ses études, pareil. Je veux dire, bon, la plus jeune a plus de mal mais on s'y efforce. Avec un petit peu d'esprit, avec des petits trucs que les enseignants des années 1920 avaient, ben voilà on arrive à faire. »

- *Qu'est-ce qui est facile/ difficile ?*

Question non posée

Q5 : Avez-vous eu l'occasion d'aider des personnes à apprendre l'orthographe ?

« Ah oui, ah ben là ! Au ministère de la défense quand j'étais dans le groupe de lutte contre l'illettrisme avec Mr Bentolila, ils étaient plusieurs, ils étaient trois. Oui j'ai eu des groupes, alors au départ j'étais formateur avant de devenir responsable de zone au ministère de la défense. On rassemblait des élèves, des jeunes soldats dans une salle et puis on s'efforçait de leur expliquer les règles de grammaire, même le français courant et puis on faisait aussi du calcul parce que les deux n'alliaient pas l'un sans l'autre en définitive et on avait...c'était très intéressant parce qu'on voyait les résultats et au terme de leur service militaire, ces jeunes qui n'avaient aucune formation aucun diplôme, ils sortaient avec un certificat de pratique professionnelle, ils sortaient avec le permis de conduire et ils sortaient avec un petit papier qu'on leur délivrait qui était signé du groupe de lutte contre l'illettrisme du ministère de la défense, et bien ils avaient un petit document qui disait voilà, vous avez un certificat pratique de l'apprentissage de la langue. Donc ils savaient au moins écrire une phrase, ce qu'ils ne savaient pas faire correctement avant, qui avait du sens. »

[Enquêteur : et au travail aujourd'hui est-ce que ça vous arrive ?]

« Ah oui [rires], comme je vous le disais tout à l'heure, j'ai des gens qui me demandent toujours, sur le plan professionnel j'en ai beaucoup, une cinquantaine. J'en ai deux, je leur ai dit, avant d'envoyer quoi que ce soit, un courrier ou un mail, vous me l'envoyez pour que je le corrige parce que les trois quarts du temps, d'abord ça n'a pas de sens ! Vous imaginez à 50 ans ! Ou alors pire, il y'a une faute d'orthographe à tous les mots. »

- *Avec des membres de votre famille, vos enfants (petits enfants), des voisins, des amis ?*

Question non posée

- *Au travail, ça vous est arrivé ?*

Question non posée

Q6 : Que pensez-vous des personnes qui font des erreurs d'orthographe ?

« Alors il y a deux choses, deux cas. Les premiers cas ce sont les gens qui ne s'appliquent pas, c'est de l'étourderie, c'est de la nonchalance. Ils cherchent pas à évoluer ou à grandir donc ils s'en fiche, voilà c'est de la nonchalance. Et le deuxième cas, c'est un problème de formation donc il y'a ceux qui ont été oubliés par l'éducation nationale. Il faut des résultats, alors on laisse, on laisse aller et puis ils sortent du cursus scolaire au fil du temps et ça devient un peu n'importe quoi. Alors pour moi y'a deux cas. »

Q7 : Quand vous écrivez, est-ce qu'il y a d'autres choses que l'orthographe qui sont difficiles pour

« Vous pouvez préciser ? »

[Enquêteur : par exemple la tournure des phrases ?]

« Ah non, au contraire. Non au contraire, alors parfois il faudrait que je simplifie, mais non, non non je vois pas. »

[Enquêteur : la ponctuation ?]

« Non au contraire, la ponctuation elle va comme on parle c'est-à-dire que si vous sortez une phrase, tout d'abord il faut faire des phrases courtes dans lesquelles il faut mettre un point, mais après la virgule par exemple, les gens ne savent pas où mettre la virgule, mais la virgule se met à la reprise d'une idée, sur une même phrase, ou un complément d'idées. Quand on met un complément d'objet direct sans en avoir mis un préalablement, on va en mettre une, et un complément d'objet indirect a fortiori et ainsi de suite, non je vois pas. »

Annexe 10 : Dictée produite par l'enquêté « bas » n° 1960

1

Annexe 10

FARFACE

ANTIE REHE MIKATOUZ ET SIRAP

ALCOLA GO DE YE

IPSORE

TOMAT DE PEIE

CONVOR

SAL COPIE

3 FROMAJ KE NSANT PAS FOR

LIBRI RE

LE FIYANSE SOULANAL DAS JOUK PEVANER

93 ON 1113 804

FA 001113 de 38 sse a ou

(ANLNA)

FBRIMAS

OKIRM

MKTW ou SIRP

AKKL 90 DKS

PISRA

TOMT DE PVI

CFTIA

SORISI

SOL COPRA

3 FOMJA qui NE PA FORA

LIBARI

de fandi de LIN DE JOUKI PFINRA

93 01 1503 766

RÉSUMÉ

Dans cette thèse, nous nous intéressons aux compétences orthographiques de scripteurs âgés de 16 à 65 ans et résidant en France métropolitaine, qui ont été considérés en difficulté dans le domaine de l'écrit dans l'enquête Information Vie Quotidienne (IVQ) (INSEE, 2011). Le matériau de notre recherche se compose de deux dictées qui reflètent une situation de scription ordinaire (la rédaction d'une liste de courses), qui comportent des mots, des pseudo-mots et des syntagmes nominaux. Elles sont issues d'une première enquête d'IVQ et d'une enquête complémentaire que nous avons conduite en 2014. Notre approche s'inscrit dans le cadre de la littéracie et vise à mettre en évidence non pas seulement les lacunes de ces publics mais également leurs savoirs et savoir-faire. Notre hypothèse est que l'on peut envisager les compétences orthographiques de scripteurs même en grande difficulté autrement que sous l'angle d'une opposition entre forme correcte et forme erronée, et qu'une analyse plus fine est possible en mobilisant les apports de la linguistique. Nous montrons en particulier comment les scripteurs s'approprient la composante phonographique de l'orthographe : bien que certains soient jugés « illettrés » selon l'enquête IVQ, ils ont développé des connaissances sur le système orthographique du français, à la fois dans ses régularités et ses spécificités. Notre recherche se veut ainsi un préalable servant à mieux outiller les actions de formation par la prise en compte des difficultés des scripteurs de manière plus ciblée et par la mise en place de diagnostics qui peuvent prendre la forme d'une évaluation de type dictée, mais en travaillant sur le choix des items en regard des résultats de notre étude sur les niveaux de compétence à l'écrit des scripteurs en difficulté.

Mots clés : compétences orthographiques, adultes en difficulté avec l'écrit, littéracie, phonographie, langue française.

ABSTRACT

In this thesis, we are interested in the spelling abilities of writers aged 16 to 65 living in metropolitan France, who were considered to be in difficulty with writing according to the IVQ survey (INSEE, 2011). The material of our research consists of two dictations that reflect a situation of ordinary scription (the writing of a shopping list), which include words, pseudo-words and noun phrases. They come from a first IVQ survey and a complementary survey that we conducted in 2014. Our approach is part of the literacy framework and aims to highlight not only the weaknesses of these audiences but also their knowledge and know-how. Our hypothesis is that the spelling abilities of writers even in great difficulty, may be considered from a point of view that differs from the opposition between right and wrong forms, and that a more refined analysis is possible by mobilizing the contributions of linguistics. We show in particular how writers appropriate the phonographic component of spelling: although some are considered "illiterate" according to the IVQ survey, they have developed knowledge about the French spelling system, both in its regularities and its specificities. Our research thus provides a helpful tool for training by taking into account the difficulties of writers in a more targeted way, and by setting up diagnostics that can take the form of a dictated evaluation type, but by working on the choice of items against the results of our study on the writing proficiency levels of writers in difficulty.

Keywords: Spelling abilities, adults with trouble with writing, literacy, phonography, French language.