



HAL
open science

**DECRIRE LES PRATIQUES ORTHOGRAPHIQUES
DES PERSONNES EN DIFFICULTE AVEC L'ECRIT
A PARTIR D'UNE DICTEE (IVQ 2011) : ASPECTS
METHODOLOGIQUES ET ETUDE DE LA
VARIATION OBSERVEE SUR UN MOT ET UN
PSEUDO-MOT**

Carole Blondel

► **To cite this version:**

Carole Blondel. DECRIRE LES PRATIQUES ORTHOGRAPHIQUES DES PERSONNES EN DIFFICULTE AVEC L'ECRIT A PARTIR D'UNE DICTEE (IVQ 2011) : ASPECTS METHODOLOGIQUES ET ETUDE DE LA VARIATION OBSERVEE SUR UN MOT ET UN PSEUDO-MOT. Colloque international des Etudiants chercheurs en Didactique des langues et en Linguistique, Lidilem, Jun 2014, Grenoble, France. hal-01251981

HAL Id: hal-01251981

<https://hal.science/hal-01251981>

Submitted on 7 Jan 2016

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.

**DECRIRE LES PRATIQUES ORTHOGRAPHIQUES DES PERSONNES EN
DIFFICULTE AVEC L'ECRIT A PARTIR D'UNE DICTEE (IVQ 2011) :
ASPECTS METHODOLOGIQUES ET ETUDE DE LA VARIATION
OBSERVEE SUR UN MOT ET UN PSEUDO-MOT**

Carole BLONDEL
carole.blondel@u-grenoble3.fr
LIDILEM, Université Grenoble Alpes
F-38040 Grenoble
France

Abstract : This study of orthographic practices, concerning people with difficulties in writing, is part of a doctoral research project, funded by the Rhône-Alpes region, in which the results of a spelling test, proposed to the most disadvantaged writing respondents in the IVQ survey conducted by INSEE in 2011, is analysed. After presenting the survey protocol, we will present the variation in the corpus of the Rhône-Alpes region, from the analysis of a word and a no-word which have close linguistic characteristics. We observe that, despite the heterogeneity of the results, the writers seem to have a relatively effective phonemic awareness and the frequency effect seems to have an impact on the spelling success.

1. Evaluer le niveau de littératie chez les adultes

L'évaluation du niveau de littératie chez les adultes n'est pas une préoccupation récente. Depuis le milieu des années 90, trois enquêtes internationales ont été proposées par l'OCDE¹ : IALS², ALL³ et PIAAC⁴, dont les tests visaient principalement à évaluer les compétences en lecture et en compréhension.

Les premiers résultats d'IALS qui ont été médiatisés pour la France ont entraîné de nombreuses polémiques sur la fiabilité de l'enquête : « Trois français sur quatre se situeraient à un niveau tout juste suffisant pour obtenir un emploi semi-qualifié » (Bastyns, 2012 : 26), mais bien que très controversée, l'enquête IALS a permis de prendre conscience en France, de certaines faiblesses dans le domaine de l'évaluation des adultes, notamment en matière de méthodologie (Besse et Guérin-Pace, 2002 : 18).

C'est dans ce contexte que la France a chargé l'INSEE de développer une mesure de la littératie des adultes, qui prenne en compte tous les niveaux de compétences et notamment les faibles niveaux pour estimer l'importance de l'analphabétisme ou de l'illettrisme.

¹ OCDE : Organisation de Coopération et de Développement Economiques

² IALS : « International Adult Literacy Survey », vingt pays dont la France entre 1994 et 1998

³ ALL : « Adult Literacy and Life skills survey », onze pays (pas la France) entre 2003 et 2008

⁴ PIAAC : « Programme pour l'évaluation internationale des compétences des adultes », 23 pays (dont la France) entre 2008 et 2013

Décrire les pratiques orthographiques des personnes en difficulté avec l'écrit à partir d'une 2 dictée (IVQ 2011) : aspects méthodologiques et étude de la variation observée sur un mot et un pseudo-mot

1.1. Evaluer l'usage de l'écrit dans la vie quotidienne : l'enquête IVQ (INSEE, 2002, 2004, 2011)

L'enquête sur les usages de l'Information dans la Vie Quotidienne (IVQ) a été conçue pour mesurer la capacité de la population adulte (âgée de 16 à 65 ans et résidant en France métropolitaine) à communiquer dans les domaines fondamentaux de l'écrit : savoir déchiffrer, savoir écrire, comprendre ce qui est lu, savoir compter. Son objectif est à la fois quantitatif et qualitatif : il s'agit d'évaluer le niveau de compétence et d'établir une typologie des différentes formes de compétences.

Une première enquête à visée méthodologique réalisée en 2002 sur un nombre restreint de personnes a permis de mettre au point le protocole proposé en 2004 puis en 2011 (auprès de 13750 personnes), en tenant compte des biais révélés dans les résultats de l'enquête IALS.

1.1.1. Le protocole de l'enquête : évaluer tous les niveaux de compétence

L'enquête débute par une série d'exercices d'orientation axés sur les compétences en lecture et en compréhension orale, qui permettent d'orienter l'enquêté vers un parcours différencié : en fonction de son taux de réussite, celui-ci est dirigé vers des exercices complexes (module haut), ou vers des exercices simples (module bas) centrés sur le repérage et l'évaluation des personnes qui présentent des difficultés dans les domaines fondamentaux de l'écrit. Un module « intermédiaire » permet d'affiner le niveau de compétences en re-testant les personnes qui se situent aux limites du seuil de compétences suite à la première série d'exercices « le diagnostic étant désormais plus fiable pour les personnes situées autour du seuil de compétence minimal retenu » (Murat, 2005 :4). Tous les enquêtés ont par ailleurs 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 enquêtés sont finalement répartis en deux catégories : « Pas de difficulté dans les domaines fondamentaux de l'écrit » ou « En difficulté dans les domaines fondamentaux de l'écrit ». 16% des enquêtés ont ainsi été dirigés vers le module bas, parce qu'ils ont réussi moins de 8 questions sur 10 dans l'une des compétences testées. L'étude de leur profil⁵ a permis à l'ANLCI de déterminer que 7% des enquêtés se trouve en situation d'illettrisme, ce qui représente environ 3 millions de personnes vivant en France métropolitaine.

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. Les observations menées par le laboratoire PsyEF auprès de personnes en difficulté ont mis en relief l'impact de la fragilité psychologique des enquêtés dans les résultats en estimant à 15% le taux d'erreurs d'une même personne en fonction du contexte de l'entretien (Besse, 2002). Alors que l'on connaît les résistances des publics concernés à toute situation d'évaluation, le nombre de « non-réponses » s'est pourtant avéré faible : pour la région Rhône-Alpes, seuls 2 enquêtés sur les 107 interrogés n'ont pas produit de texte, probablement à cause d'une réelle incapacité à

⁵ « [...] l'illettrisme qualifie la situation de personnes de plus de 16 ans qui, bien qu'ayant été scolarisées, ne parviennent pas à lire et à comprendre un texte portant sur des situations de leur vie quotidienne, et/ou ne parviennent pas à écrire pour transmettre des informations simples » (ANLCI, 2013)

produire des mots écrits (« Trop difficile ne savait pas écrire donc a refusé la dictée », indique par exemple un enquêteur).

Le protocole de l'enquête semble donc avoir plutôt réussi à mettre en confiance les personnes interrogées, ce qui renforce la fiabilité des données recueillies.

1.2. *L'épreuve de production écrite : présentation de la dictée*

L'épreuve de production écrite (la seule proposée dans l'enquête IVQ), est intégrée au module bas : il s'agit d'une liste de course⁶ que les enquêtés ont reproduit sous la dictée, de manière manuscrite.

Pharmacie
Anti-rhume <i>MICATOL</i> ou <i>SIRAPE</i>
Alcool à 90° <i>DUX</i>
Epicerie
Tomates de pays
Confiture
Cerises
Sel <i>GOBAR</i>
Trois fromages qui ne sentent pas fort
Librairie
Le fiancé solennel de <i>JOC PEVANOR</i>

Tableau 1 : Dictée proposée en 2011 – les pseudo-mots sont présentés en italique

La liste est composée de trente-neuf items (en comptant les mots outils, les chiffres, ...). Les 18 mots et pseudo-mots qui ont fait l'objet d'une correction ont été choisis pour leur longueur (de trois à sept phonèmes) et leur caractéristique : des mots phonographiquement réguliers – *anti*, *épicerie*, *tomates*, *confiture*, *cerises*, *sel* –, des mots irréguliers – *pharmacie*, *rhume*, *alcool*, *pays*, *fiancé*, *solennel* – qui permettent de rendre compte des connaissances lexicales des enquêtés et des pseudo-mots⁷ – *micatol*, *sirape*, *dux*, *gobar*, *joc*, *pévanor* – utilisés pour déterminer le niveau de conscience phonographique sur 19 phonèmes (13 consonnes et 6 voyelles). La conscience morphologique a été jugée d'après la réalisation de l'accord en nombre d'un nom – *fromages* – et d'un verbe – *sentent* –, mots qui n'ont pas fait l'objet d'une correction orthographique. Cette épreuve ne permet néanmoins d'observer qu'un nombre restreint de phénomènes linguistiques : lettres muettes internes (*ex. rhume*), finales muettes (*ex. épicerie*), accord d'un nom (*fromages*) et d'un verbe (*sentent*), marques de nombre facultatives (*ex. cerises*) ; l'objectif premier étant de mesurer leur capacité à utiliser l'écrit pour l'usage quotidien.

Nous disposons, grâce à cette enquête, d'un corpus inédit de 2000 dictées produites par des adultes en difficulté à l'écrit, qui nous permettent de déterminer leurs capacités orthographiques avec pour enjeu de se questionner sur des degrés d'illettrisme.

⁶ Conçue par l'équipe de Jean-Marie Besse, laboratoire PsyEf (Lyon 3)

⁷ Nous appelons « pseudo-mot » les mots qui ont été inventés par les concepteurs du test ; nous notons qu'il ne s'agit ici que de noms propres représentant des noms de médicament et les nom-prénom d'un auteur d'ouvrage.

Décrire les pratiques orthographiques des personnes en difficulté avec l'écrit à partir d'une 4 dictée (IVQ 2011) : aspects méthodologiques et étude de la variation observée sur un mot et un pseudo-mot

2. Exploiter les données d'une dictée réalisée par des personnes en difficulté : la nécessité d'une analyse particulière

L'exploitation de ce corpus pose les bases d'une réflexion méthodologique sur la manière de traiter de telles données. Une évaluation classique qui consisterait à repérer les graphies incorrectes et à établir des scores de réussite, ne nous permettrait pas de rendre compte des capacités de ce public à communiquer par écrit.

2.1. Présentation des critères et des attentes de l'INSEE

Pour évaluer les performances à l'écrit, l'INSEE a réalisé des scores de réussite des items de la dictée (à partir d'une analyse quantitative des données). Ce traitement des données (et le fichier tel qu'il nous a été fourni) a permis de mesurer les compétences orthographiques des enquêtés et de calculer le taux de personnes en situation d'illettrisme.

▪ Traitement informatique des items et calcul des scores de réussite

Pour établir des statistiques, l'INSEE a saisi les réponses des enquêtés en attribuant pour chaque item⁸, un code « Faux », « Absent » ou « Juste » permettant l'attribution de points.

C'est la conformité à la norme orthographique pour les mots et les attentes graphophonologiques des concepteurs du test pour les pseudo-mots qui ont été jugés. 1 point a été accordé par réponse correcte et 2 points ont été ajoutés si les marques obligatoires de pluriel (accord d'un verbe et d'un nom) étaient présentes. Un total de 22 points détermine ainsi le niveau de l'enquêté sur l'épreuve de production écrite.

Les enquêtés qui ont réussi moins de 8 mots sur les 20 qui ont été corrigés ont été considérés en grande difficulté dans ce domaine.

L'objectif de l'INSEE étant quantitatif, certaines variations présentes sur les items n'ont pas fait l'objet d'un relevé et le fichier informatique des graphies, lorsqu'il est comparé avec la version papier des dictées, montre une perte d'informations orthographiques.

▪ Impossibilité d'observer la variation orthographique et phonologique

Dans le fichier des données, seule une variante (parmi toutes les variantes acceptables) a été relevée :

La marque de pluriel pour *- cerise(s) -* et *- tomate(s) -* n'était pas considérée comme obligatoire par les concepteurs du test, ces items ayant été dictés sans déterminant. La présence du « s » était néanmoins attendue, ce qui a engendré une uniformisation des réponses des enquêtés : les graphies ont systématiquement été enregistrées avec la marque du pluriel :

Ex. Version papier de la dictée : « cerise » vs graphie enregistrée dans le fichier : « cerises »

Cet exemple ne remet pas en question le respect de la norme orthographique, puisque les deux graphies sont acceptables, dans ce contexte. En revanche, de manière erronée, le même traitement a été appliqué à l'adverbe (invariable) « fort » :

Ex. Version papier de la dictée : « fort » vs graphie enregistrée dans le fichier : « forts »

Pour le traitement statistique du taux de réussite sur les pseudo-mots : une seule forme a également été relevée dans le fichier (lorsque la graphie proposée par l'enquêté était graphophonologiquement correcte) :

⁸ L'INSEE a fait le choix de ne corriger que 20 items parmi les 39 qui ont été dictés, certains mots comme « trois » par exemple ou comme les mots outils ou ont été jugés trop faciles.

Ex. Version papier de la dictée : « *syrape* » vs graphie enregistrée dans le fichier : « *sirape* »

L'accentuation n'a pas non plus été relevée dans le fichier informatique : l'item « *fiancé* » apparaît donc systématiquement sans accent, si bien qu'à partir du fichier, on ne peut pas savoir si l'enquêté a utilisé le phonème / e /.

Cette perte d'information sur les mots « cerise(s) », « tomate(s) » et « fort » et les pseudo-mots ainsi que sur l'accentuation de « fiancé » ne permet pas, pour certains items, d'observer la variation orthographique des formes proposées ni la restitution de la forme phonologique.

- **Les imputations et les graphies illisibles**

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, qui s'est appuyé sur les données individuelles collectées lors de la passation du module biographique.

Ces différents exemples ont pour conséquence 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 établir des scores de correspondance graphophonologique ni pour rendre compte précisément des capacités à l'écrit de ce public. Un travail de vérification et de correction des graphies a donc été réalisé pour disposer d'un fichier qui corresponde aux graphies proposées par les enquêtés avec notamment l'attribution d'un code « illisible » pour les graphies dont nous avons considéré la reconnaissance des lettres trop aléatoire.

2.2. *Juger de l'efficacité de la communication écrite des personnes en difficulté : quels critères d'analyse ?*

Pour mesurer la capacité d'écriture, notre analyse ne peut se limiter au taux de réussite sur la vingtaine d'items : elle doit analyser précisément les erreurs et utiliser toutes les données qui peuvent être extraites dans, et en dehors des textes.

2.2.1. *Le traitement des items*

Pour rendre compte précisément de la manière dont ce public parvient à restituer un message, nous avons choisi de traiter le corpus sous différents aspects : le respect de la norme orthographique avec ou sans respect des marques morphologiques flexionnelles (pour les mots) et le respect de la forme phonologique (pour les mots et les pseudo-mots). Une analyse des choix graphiques des enquêtés (sur les pseudo-mots et sur les formes erronées) en lien avec les travaux réalisés en linguistique sur l'analyse orthographique (Catach, 1995) sera effectuée pour comprendre le fonctionnement des enquêtés face à la production d'écrit. Pour rendre compte le plus précisément possible des difficultés des scripteurs, une description linguistique est prévue à la fois sur la syllabe et sur le phonème.

2.2.2. *L'analyse des métadonnées : matérialité de l'écriture et rapport à l'écriture*

En plus du traitement des graphies, nous avons souhaité décrire toutes les informations produites par les scripteurs sur la copie selon une approche sémiotique de l'écriture. Cette analyse permettra de se questionner sur le rapport à l'écriture des enquêtés, abordée dans sa matérialité mais également sur la posture du lecteur qui reçoit le texte. Cette analyse sera également l'occasion d'aborder la question du traitement graphique des items et des choix qui ont dû être fait lorsque nous avons numérisé les textes.

Toutes les données présentes autour des textes, que nous avons appelées « métadonnées », n'ont pas été prises en compte par l'INSEE. Pour les recenser et les comparer, nous avons alors constitué une grille d'analyse composée de 4 champs d'investigation. Nous nous sommes d'abord intéressés à l'organisation matérielle de la copie, qui représente la première

Décrire les pratiques orthographiques des personnes en difficulté avec l'écrit à partir d'une 6 dictée (IVQ 2011) : aspects méthodologiques et étude de la variation observée sur un mot et un pseudo-mot

approche du lecteur avec le texte. Nous avons observé ici la présentation choisie par l'enquêté (sous forme de liste, items écrits les uns à la suite des autres) et l'effort d'organisation et de lisibilité qui peut être repéré (matérialisation des titres, utilisation de la ponctuation, marques d'autocorrection). Nous avons ensuite souhaité rendre compte du geste graphique des enquêtés et de leur rapport à la matérialité de l'écriture ; une grande variation sur cette question apparaît en effet à la lecture de ce corpus. Dans cette partie, nous avons détaillé la ou les forme(s) d'écriture(s) utilisée(s) (écriture cursive, écriture script), la présence ou non de majuscules, le type de majuscules utilisées (majuscules cursives ou lettres capitales) et les marques d'accentuation particulières, qui ne correspondent pas à notre système orthographique. Il nous a également semblé intéressant de rendre compte de la manière dont les enquêtés ont répondu à la consigne. Il s'agissait, dans cette épreuve, de reproduire le texte dicté. Hors, certains enquêtés ont ajouté ou enlevé des items, modifié l'ordre de la liste. Un dernier champ « commentaire » permet enfin de détailler certaines informations qui ne pourront pas faire l'objet d'un traitement quantitatif mais qui méritent d'être signalées : le détail des items qui ont été ajoutés, les commentaires des enquêteurs, etc.

Certaines informations sont renseignées à partir d'un jugement subjectif, comme la qualification d'« écriture très difficile » qui a été déterminée en s'appuyant sur la lisibilité des graphies ou sur la manière de former les lettres (geste qui paraît hésitant, lignes tremblantes, lettres irrégulières ou relevant d'autres alphabets). Ces informations devraient nous permettre d'obtenir un éclairage différent sur le rapport à l'écrit et à l'écriture de ce public en difficulté.

La saisie de ces métadonnées, la mise en relief des phénomènes pertinents ainsi qu'à terme, la corrélation des items (et des choix opérés au niveau de la syllabe et du phonème) avec les données biographiques de l'enquête, nous ont amené à la conception d'une base de données qui nous a semblé être un outil particulièrement adapté à l'exploitation de ce corpus.

2.2.3. *Une base de données pour étudier les graphies erronées*

Le traitement du corpus est effectué en collaboration avec le département de Traitement Automatique des Langues de l'Université Stendhal (Grenoble 3). Ce travail a abouti à la conception d'une base de données⁹ qui nous permet de saisir les items de la dictée ainsi que les métadonnées dont nous disposons, à partir de la version papier des dictées. Ces données pourront être reliées aux données du module biographique, afin de procéder à un traitement statistique très fin pour observer le profil des personnes en difficulté et éventuellement corréler certains types de difficultés avec des déterminants précis d'ordre social.

En cours de développement, cet outil nous permettra d'obtenir le détail des variantes pour chaque item avec leurs occurrences, ce qui nous aidera à mettre en évidence les régularités statistiques. D'autres fonctionnalités sont en cours d'expérimentation, notamment pour obtenir une analyse sur la syllabe et sur le phonème.

⁹ Réalisée par C. Ponton, Maître de conférence en TAL et EIAH), à qui nous adressons nos remerciements, pour ce travail

3. Présentation de la variation orthographique des scribes rhônalpins en difficulté : l'exemple du mot « cerise(s) » et du pseudo-mot « sirape »

Une première étude réalisée de manière systématique sur le corpus de dictées disponible pour la région Rhône-Alpes permettra de sélectionner les critères d'analyses pertinents qui seront étendus au corpus national des 2000 dictées.

Nous disposons de 106 dictées réalisées par les 108 rhônalpins qui ont été dirigés vers le module de l'ANLCl.

Pour présenter la variation présente dans le corpus Rhône-Alpes, nous avons choisi d'analyser le mot « cerise(s) » et le pseudo-mot « sirape », qui comportent 6 graphèmes pour 5 phonèmes ainsi qu'un « e » muet final. Nous notons que dans ce cas, le « e » final n'est pas complètement muet puisqu'il n'est pas phonogramme, mais modifie la valeur phonologique d'un autre graphème (« s » pour « cerise(s) » et « p » pour « sirape »).

3.1. Détail de l'analyse et résultats pour l'item « cerise(s) »

104 scribes, parmi les 106 qui ont rédigé la dictée, ont produit le mot « cerise(s) ».

Le tableau ci-dessous détaille les variantes avec leurs occurrences et les présente selon 3 critères : les productions attendues (orthographiquement correctes), les productions qui permettent de reconnaître le mot (correspondance phonie-graphie correcte) et les autres variantes, qui ne correspondent pas au système orthographique du français ou qui ne permettent pas de restituer la forme phonologique du mot. Nous avons également isolé les graphies qui comportent une marque de pluriel : nous avons trouvé intéressant de regarder si les scribes ont prêté de l'attention à cette information, qui nécessitait de procéder à une inférence. Nous n'avons pas considéré le « s » final des formes « seris » et « ceris » comme une marque de pluriel, car ces variantes semblent davantage correspondre à un traitement phonétique de la forme entendue. Par ailleurs la graphie « is » qui correspond en français au phonème / i / ne permet pas de restituer les phonèmes / i / et / z /. Nous avons donc considéré que les formes qui n'avaient pas de « e » muet, ne pouvaient pas être considérées comme graphophonologiquement correctes.

	Orthographiquement correctes	Phonologiquement correctes	Autres variantes
Avec « s »	cerises (28) <i>= 40% parmi les formes ortho. correctes</i>	se rises - serises <i>= 33% parmi les formes phono. correctes</i>	cerices (4) <i>= 15% parmi les autres variantes</i>
Sans « s »	cerise (43)	çerise (2) serise (2)	cerice (7) - seris (2) caurise - cericie - cérique - criz cerisse - cirise - curize - sorise - ceris sourice - sourise - sugüz - ceriz - seriz
Total	71 graphies (68 %)	6 graphies (6 %)	27 graphies (26 %)

Tableau 2 : Détail des variantes et occurrences pour l'item "cerise(s)"

En neutralisant la présence de majuscules, nous avons relevé 23 variantes.

La majorité des graphies (71) sont orthographiquement correctes, ce qui représente 68% des dictées. Parmi les formes qui n'étaient pas attendues, seules 6 graphies sont phonologiquement correctes (soit 6%), les 27 autres (26% des graphies) sont des variantes graphophonologiquement incorrectes.

Décrire les pratiques orthographiques des personnes en difficulté avec l'écrit à partir d'une 8 dictée (IVQ 2011) : aspects méthodologiques et étude de la variation observée sur un mot et un pseudo-mot

34 occurrences comportent la marque du pluriel (soit dans 33 % des dictées). Cette marque est davantage présente sur les items qui sont correctement orthographiés (40%) et sur ceux qui ont restitué la forme phonologique (33%). Ce sont les formes graphophonologiquement incorrectes qui se démarquent : seules 15% d'entre elles comportent une marque de pluriel. Il semble que mieux la graphie respecte la forme phonologique du mot et plus la marque du pluriel est présente : la prise en compte de la morphologie par l'utilisation d'une inférence semble donc augmenter avec les compétences graphophonologiques.

Parmi les 27 autres variantes, nous pouvons observer des formes phonologiquement très proches : 18 variantes n'ont qu'un seul phonème qui ne correspond pas à la forme phonologique attendue (parmi la restitution des 5 phonèmes).

Par ailleurs, une seule variante peut-être considérée comme très éloignée de la forme attendue : « sugüz ».

Lorsque nous observons la variante « caurise », nous constatons que le traitement par phonème ou par graphème n'est pas pertinent : en effet, si le graphème « c » est correct, la graphie « au » sélectionne la valeur phonologique / k / et rend la 1^{ère} syllabe / ko / phonologiquement incorrecte. Cet exemple nous montre la nécessité de détailler les choix graphiques des enquêtés par syllabe afin d'observer pour chacune d'elle, la distribution des phonèmes et des graphèmes et la conformité graphophonologique de l'item proposé avec la forme attendue. Ce travail nous permettra d'observer, en contexte, les phonèmes et les graphèmes qui ont posé le plus de difficultés aux enquêtés.

3.2. *Détail de l'analyse et résultats pour le pseudo-mot « sirape »*

Le pseudo-mot « sirape » comporte 101 occurrences réparties en 29 variantes (5 enquêtés ne l'ont pas transcrit).

Nous distinguons pour cette étude, les variantes qui restituent la forme phonologique de l'item, de celles qui ne la restituent pas. Comme pour le mot « cerise(s) », nous avons considéré que les formes qui ne comportaient pas de « e » final, ne pouvaient pas être considérées comme graphophonologiquement correctes.

Phonologiquement correctes	Phonologiquement incorrectes
sirape (22) cirape (11) cyrape (2) sirapeu - çirape - cirappe - sirappe - syraper (1) Problème de segmentation : ousirape	sirap (33) cirap (8) ciraque (2) syrap - cyrap - cerape - cirabe - cirope - cirat – cyrak sirame - sirapa - siraphe - siraple - sirope - sqpa – sixrape (1) Problème de segmentation : ocirap - ousirap - ucirap
41 graphies (41%)	60 graphies (59%)

Tableau 3 : Détail des variantes et occurrences pour le pseudo-mot "sirape"

Nous avons relevé 4 occurrences qui n'ont pas été segmentées avec l'item précédent « ou »¹⁰.

La majorité des graphies proposées (59%) sont phonologiquement incorrectes (en neutralisant le problème de segmentation qui concerne 4% des graphies) et sont réparties en 16 variantes. Nous remarquons une forme récurrente : « sirap », qui concerne 58% des

¹⁰ L'item « sirape » fait partie de la phrase : « Anti-rhume micatol ou sirape. »

formes phonologiquement incorrectes proposées. Cette graphie (avec un « e » final) est également celle qui a été la plus utilisée (56%) parmi les formes phonologiquement correctes.

Le pseudo-mot « sirape » a été présenté comme la marque d'un anti-rhume. Phonologiquement et graphiquement, cet item est très proche du mot « sirop ». Cette proximité pourrait expliquer les variantes « sirap » (34 occurrences) ou « cirap » (10 occurrences) ainsi que l'absence du « e » final muet. Parmi les formes graphophonologiquement erronées, deux variantes se rapprochent également du mot « sirop » : « cirope » - « sirope ». Pour orthographier ce pseudo-mot, il semble que certaines personnes se soient inspirées d'un mot connu.

Nous pouvons également constater une complexification orthographique qui n'a pas été constatée avec le mot « cerise(s) » : 6 graphies comportent un « y » et 2 graphies ont un redoublement de consonne « pp ». Il pourrait s'agir d'un choix de complexification volontaire lié à la présence d'un pseudo-mot associé au domaine médical.

Pour la variante « sixrape », proposée par un enquêté, il aurait été intéressant d'interroger la personne pour savoir s'il s'agit d'une recherche de sens lexical (six râpes ?).

Par ailleurs, c'est la restitution du dernier phonème / p / qui a engendré le plus de variation et certains choix sont très éloignés du phonème attendu : / m /, / b /, / t /, / f /, ce qui n'est pas le cas pour le mot « cerise(s) », puisque c'est le 2^{ème} phonème / ə / qui a le plus généré le plus de variation : / o /, / e /, / i /, / y /, / u /, outre les difficultés avec / s / (c, ç, s) et / z / (s et z).

4. Conclusion

Le taux de réussite graphophonologique sur ces deux items nous montre que parmi les personnes en difficulté avec l'écrit, certains semblent avoir une conscience phonémique relativement efficace. Les difficultés constatées chez certains scripteurs ont concernés des phonèmes différents (/ ə / pour « cerises » et / p / pour « sirape »). La position du phonème dans l'item ne semble pas avoir d'influence sur sa réussite et certains phonèmes semblent poser davantage de difficultés, ce qu'il s'agira de vérifier sur d'autres phonèmes et sur des items plus longs.

En revanche, l'exemple de la confusion des phonèmes / s / et / z / dans l'item « cerise(s) » montre que certains scripteurs n'ont pas conscience des règles de correspondance phonie-graphie ou que leurs compétences sont fragilisées lorsqu'un mot présente, dans sa composition, des graphèmes phonologiquement proches.

La proportion de variantes qui restituent la forme phonologique du mot « cerise(s) » est plus élevée (74%) que pour le pseudo-mot « sirape » (41%). Il semble donc que la connaissance de l'item dicté représente un facteur de réussite supplémentaire à l'écrit et que l'effet de fréquence ait un impact sur la réussite. Cet effet de la fréquence pourra également être testé à partir de l'analyse des choix graphémiques des enquêtés pour voir si ce sont les correspondances phonie-graphie les plus probables en français qui sont les plus utilisées.

Par ailleurs, la restitution du « e » final muet, bien que davantage réussie pour le mot (seules 7 graphies ne comportent de « e » final muet) que pour le pseudo-mot tend à montrer que cette particularité du français semble plutôt bien intégrée chez ces scripteurs.

Cette première étude nous offre des perspectives d'analyses intéressantes à étendre à tous les items de la dictée et au corpus national.

Parallèlement à ce travail, un recueil de dictées produites par des personnes qui ont participé à l'enquête IVQ en 2011 mais qui n'avaient pas été repérées en difficulté et qui n'ont donc pas passé l'épreuve de production écrite, est en cours. Cette enquête nous offrira un éclairage intéressant pour notre étude, en nous permettant de comparer les productions de deux publics

Décrire les pratiques orthographiques des personnes en difficulté avec l'écrit à partir d'une dictée (IVQ 2011) : aspects méthodologiques et étude de la variation observée sur un mot et un pseudo-mot

aux compétences différentes. Ces nouvelles données nous aideront également à juger l'efficacité des items choisis dans l'épreuve de dictée pour évaluer les capacités à l'écrit d'adultes en difficulté.

Références

- ANLCI, <http://www.anlci.gouv.fr/>, (nd). Consulté en juin 2014
- 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*, n°185, p. 26-39.
- BESSE, J.-M., GUÉRIN-PACE, F. (2002). Une évaluation des compétences sur l'écrit : l'enquête « Information et vie quotidienne », *Économie et Humanisme*, n° 363, numéro spécial *Vaincre l'illettrisme*, pp. 17-21.
- CATACH, N., en collaboration avec GRUAZ, C., DUPREZ, D. (1995). *L'orthographe française, traité théorique et pratique*. Paris : Nathan.
- MURAT F. (2005). La compétence des adultes à l'écrit, en calcul et en compréhension orale, *Insee Première*, n° 1044.