

ACADEMIE ORLEANS-TOURS
UNIVERSITE FRANCOIS RABELAIS
UFR DE MEDECINE
ECOLE D'ORTHOPHONIE
TOURS

Mémoire en vue de l'obtention du certificat de capacité d'orthophoniste

**Réduction d'une épreuve de répétition de non-mots
à destination des enfants bilingues**

Présenté par :

Zinaïda PETKOVIC épouse TAMIATTO

Membres du jury :

- M. Christophe dos Santos, phonologue et directeur de mémoire
- Mme Sophie Pecqueur, orthophoniste
- Mme Anne-Gaëlle Piller, orthophoniste
- Mme Rasha Zebib, psycholinguiste

Année Universitaire 2013-2014

ACADEMIE ORLEANS-TOURS
UNIVERSITE FRANCOIS RABELAIS
UFR DE MEDECINE
ECOLE D'ORTHOPHONIE
TOURS

Mémoire en vue de l'obtention du certificat de capacité d'orthophoniste

**Réduction d'une épreuve de répétition de non-mots
à destination des enfants bilingues**

Présenté par :

Zinaïda PETKOVIC épouse TAMIATTO

Membres du jury :

- M. Christophe dos Santos, phonologue et directeur de mémoire
- Mme Sophie Pecqueur, orthophoniste
- Mme Anne-Gaëlle Piller, orthophoniste
- Mme Rasha Zebib, psycholinguiste

Année Universitaire 2013-2014

Remerciements

Parce que ces quatre années auront été parmi les plus intéressantes de toutes mes vies, j'aurais voulu pouvoir hisser les mots qui suivent à la hauteur de ce que m'ont apporté ceux qui m'ont accompagné dans cette aventure... Ils savent tous mieux que personne l'importance des mots, aussi me pardonneront-ils une touche de grandiloquence...

J'adresse mes remerciements à...

Christophe dos Santos, mon directeur de mémoire, « Maître Jedi » de la phonologie, passionné-passionnant, pour sa disponibilité, sa rigueur - qui n'a jamais empêché la bonne humeur, ses relectures minutieuses et toujours pertinentes, sa façon magique de faire disparaître la pression pour ne recueillir que le meilleur du travail. Merci de m'avoir fait goûter à la recherche...

Mmes Anne-Gaëlle Piller, Sophie Pecqueur et Rasha Zebib pour l'honneur qu'elles me font de participer au jury de ce mémoire.

Sandrine Ferré, pédagogue hors-pair, pour m'avoir ouvert les portes de la phonologie et surtout donné l'envie d'y entrer...

Toute l'équipe du labo, en particulier Laurie Tuller, Rasha Zebib, Eléonore Morin, Laetitia de Almeida, Philippe Prevost pour leur chaleureux accueil, leur disponibilité et leur implication.

Sébastien Béland, pour avoir répondu à mon appel au secours si rapidement et si gentiment, et pour avoir mis son expertise au service de mon travail, alors que j'étais perdue dans les méandres psychométriques et statistiques... Vive les québécois, so far...so good !

Mes maîtres de stage : Frédérique Arreckx, Armelle Béranger, Elisabeth Beucet, Perrine Chalmet, Pascale Dansart, Célia Henry, Anne-Gaëlle Piller, Aurore Siras, pour m'avoir montré la voie.

L'école d'orthophonie de Tours, son directeur M. Emmanuel Lescanne, sa directrice pédagogique Mme Cécile Monjauze, pour m'avoir donné la chance d'accéder à ce métier à plus de quarante ans, et tous les professeurs qui ont supporté mes incessantes questions.

Mon travail est dédié à...

Aurore Siras, pour avoir su partager... Je voudrais lui faire ici une déclaration d'amitié, pour m'avoir permis d'hésiter, douter, me tromper, tout en me montrant subtilement les chemins possibles, avec l'intelligence et la bienveillance qui font d'elle une orthophoniste exceptionnelle. Merci également aux « filles » du cabinet, qui m'ont accueillie, soutenue, conseillée : Flore Chalanson, Laurence Danquigny, Florence Barbier ... et Fabienne Rivière pour les discussions-repas du jeudi midi.

Frédérique « Féérique » Arreckx, qui a, la première, personnifié l'image que j'aimais me faire d'une orthophoniste, et avec laquelle j'ai adoré travailler.

Anne-Gaëlle Piller, parce que quand je serai grande, je voudrais être Anne-Gaëlle Piller, pour la justesse de son sens clinique et diagnostique, sa curiosité professionnelle, son ouverture d'esprit, son humanité - au sens le plus noble du terme - la grande classe, quoi.

Tous les enfants et leur famille, pour l'accueil, le temps accordé, les cafés – et ceci sans contrepartie, pour faire avancer l'étude des enfants bilingues et permettre de donner naissance à des outils permettant de mieux comprendre leur développement et mieux les aider.

Mes copines de promo - Clémentine, Lulu, Jen, Macha, Anabelle, Clémence - pour m'avoir adoptée malgré les années qui nous séparent, et m'avoir permis de ne pas en sentir le poids... Pull-up !

Manue, pour l'échange de vies, et la fierté que me donne l'idée même d'être ton amie.

Vincent et Selma, qui font en sorte que le soleil se lève tous les matins...

TABLE DES MATIERES

LISTE DES TABLEAUX	1
LISTE DES GRAPHIQUES	2
LISTE DES FIGURES	3
LISTE DES ABRÉVIATIONS	4
1 Introduction	5
2 Problématique	7
<i>2.1 Les tests de Répétition de Non-mots</i>	7
2.1.1 Intérêt	7
2.1.2 La notion de complexité en phonologie	9
2.1.3 Répétition de non-mots et complexité phonologique.....	13
2.1.4 Le bilinguisme est-il un facteur influençant les performances aux tests ?.....	14
2.1.5 Répétition de non-mots et bilinguisme.....	15
<i>2.2 Les tests, instruments de mesure</i>	17
2.2.1 Qu'est-ce qu'un test ?.....	17
2.2.2 Méthodologie de la mesure : standardisation et normalisation	18
2.2.3 Test, épreuve, tâche ... ?	18
2.2.4 Différents types de mesure : tests critériés vs tests normatifs.....	19
2.2.5 Caractéristiques psychométriques : mesurer la mesure.....	20
<i>2.3 Processus d'élaboration d'un test</i>	23
2.3.1 Les étapes de construction d'un test.....	23
2.3.2 Pourquoi construire plus grand pour finalement réduire ?	24
2.3.3 Comment réduire ?	25
<i>2.4 Théorie classique du score vrai : analyse des caractéristiques des items et estimation de la fidélité</i>	26
2.4.1 Indice de difficulté	26
2.4.2 Indice de discrimination	27

2.4.3	Écart-type	28
2.4.4	Corrélation item/total ($r_{item/total}$).....	29
2.4.5	Le coefficient alpha de Cronbach : une estimation de la fidélité	29
2.5	<i>Les Modèles de Réponse à l'item : une nouvelle théorie de la mesure</i>	30
2.5.1	Une condition d'application des modèles : le nombre de sujets	31
2.6	<i>Objectifs de l'étude</i>	32
3	Matériel et Méthodes	34
3.1	<i>Présentation de LITMUS-NWR</i>	34
3.1.1	Pourquoi une nouvelle épreuve de répétition de non-mots ?	34
3.1.2	Élaboration de l'épreuve	34
3.1.3	Structure des non-mots de l'épreuve	36
3.1.4	Le cas des items contrôle	38
3.1.5	Procédure de passation de l'épreuve	39
3.1.6	Transcription et codage de l'épreuve	39
3.2	<i>Présentation de l'ensemble du protocole</i>	41
3.3	<i>Passations</i>	42
3.4	<i>Présentation de la population</i>	42
3.4.1	Recrutement et critères d'inclusion.....	43
4	Résultats	45
4.1	<i>Calcul de l'indice p de difficulté</i>	45
4.1.1	Nombre d'items en fonction de l'indice de difficulté	46
4.1.2	Indice de difficulté : comparaison DT/TSL	47
4.1.3	Indice de difficulté en fonction de la complexité.....	51
4.1.4	Indice de difficulté en fonction du type d'items : LI vs LD.....	52
4.1.5	Items candidats à l'élimination par l'indice de difficulté.....	53
4.2	<i>Calcul de l'indice D de discrimination</i>	55
4.2.1	Détermination des groupes « fort » et « faible »	55
4.2.2	Items candidats à l'élimination par l'indice D	57
4.2.3	Comparaison indice D et indice p	58

4.2.4	Lien entre l'indice p et l'indice D	58
4.3	<i>Calcul de l'écart-type</i>	60
4.3.1	Items candidats à l'élimination par l'écart-type (ET)	60
4.3.2	Écart-type des items avec indice p trop élevé ou trop bas et indice D faible.....	61
4.3.3	Écart-type des items avec indice p trop élevé et indice D correct.....	61
4.4	<i>Calcul de la corrélation r item/total</i>	62
4.5	<i>Premier essai de réduction de l'épreuve et premières interrogations</i>	63
4.6	<i>Second essai de réduction de l'épreuve</i>	67
4.6.1	Choix des items très complexes à retirer.....	68
4.6.2	Choix des items de faible complexité à ajouter.....	69
4.7	<i>Épreuve réduite LITMUS-NWR 50</i>	71
4.8	<i>Estimation de la cohérence interne : calcul du coefficient alpha de Cronbach</i>	72
5	Discussion	74
5.1	<i>Analyse des items de LITMUS-NWR : une analyse du construit</i>	74
5.1.1	Complexité phonologique et TSL	74
5.1.2	Discrimination de l'épreuve	75
5.2	<i>Réduction de LITMUS-NWR : signification statistique et pertinence clinique</i>	76
5.3	<i>Limites et perspectives</i>	77
5.3.1	Population.....	77
5.3.2	Construction des items : cadre de recherche vs utilisation clinique.....	78
5.3.3	Conditions de passation.....	79
5.3.4	Transcription des non-mots répétés.....	80
5.3.5	Temps de passation	81
5.3.6	Ordre des non-mots dans l'épreuve LITMUS-NWR 50	81
6	Conclusion	84
7	BIBLIOGRAPHIE	86
	ANNEXES	I

LISTE DES TABLEAUX

Tableau 1 : Liste des non-mots de LITMUS-NWR en fonction de leur complexité	37
Tableau 2 : Liste des items contrôle.....	38
Tableau 3 : Exemple de codage pour une substitution (/klil/ → /plil/).....	40
Tableau 4 : Exemple de codage pour une métathèse + addition (/kupalfi/ → /kuplafli/)	40
Tableau 5 : Exemple de codage pour une élision (/flaplu/ → /flapu/)	40
Tableau 6 : Épreuves du protocole de l'étude pour l'évaluation en français	41
Tableau 7 : Population de l'étude.....	42
Tableau 8 : Répartition DT/TSL en fonction de l'âge	43
Tableau 9 : Nombre d'items en fonction de leur indice de difficulté	46
Tableau 10 : indice <i>p</i> pour les groupes DT et TSL	47
Tableau 11 : Nombre d'items en fonction de leur indice de difficulté pour les groupes DT et TSL....	49
Tableau 12 : Moyenne des indices <i>p</i> selon le type d'items.....	52
Tableau 13 : Moyenne des indices <i>p</i> selon le type d'items par groupe DT/TSL	52
Tableau 14 : Les 22 items les plus faciles selon l'indice <i>p</i> + « klisp ».....	54
Tableau 15 : Groupes « fort » et « faible »	56
Tableau 16 : Classification des indices de discrimination (Ebel, 1972)	57
Tableau 17 : Classement des items selon d'indice D.....	57
Tableau 18 : Indice <i>p</i> des items les moins discriminants	58
Tableau 19 : Items avec moyenne extrême et écart-type faible	60
Tableau 20 : Écart-type des items avec <i>p</i> et D insuffisants.....	61
Tableau 21 : Écart-type des items avec <i>p</i> trop élevé et D correct.....	61
Tableau 22 : Items candidats à l'élimination	63
Tableau 23 : LITMUS-NWR 50-β : premier essai de réduction de l'épreuve à 50 items	65
Tableau 24 : Comparaison des épreuves en terme de type et de complexité des items	65
Tableau 25 : Indices des items de faible complexité monosyllabiques et des items très complexes....	68
Tableau 26 : Indices moins performants des items très complexes à éliminer	69
Tableau 27 : Indice <i>p</i> pour les items monosyllabiques (sauf kip).....	70
Tableau 28 : LITMUS-NWR 50 : réduction de l'épreuve à 50 items.....	71
Tableau 29 : Comparaison des 3 épreuves.....	72
Tableau 30 : Ordre possible des items de LITMUS-NWR 50	83

LISTE DES GRAPHIQUES

Graphique 1 : Répartition des indices p pour les groupes DT et TSL.....	48
Graphique 2 : Comparaison des moyennes des indices p pour les groupes DT et TSL.....	48
Graphique 3 : Exemple de comparaison de l'indice p DT/TSL sur 18 items.....	49
Graphique 4 : Items dont l'indice p varie peu selon les groupes (DT/TSL)	50
Graphique 5 : Moyenne des indices p en fonction de la complexité.....	51
Graphique 6 : Comparaison des moyennes des indices p par groupe en fonction du type d'items.....	53

LISTE DES FIGURES

Figure 1 : Schéma général de la syllabe	10
Figure 2 : Structure syllabique du mot « patte » [pat].....	11
Figure 3 : Types d'attaques en français.....	11
Figure 4 : /paR/ vs /pat/ selon le modèle rythmique d'Angoujard	12

LISTE DES ABREVIATIONS

ALL	Ensemble des sujets (résultats)
API	Alphabet Phonétique International
Bi-DT	Bilingue au développement typique
Bi-TSL	Bilingue avec troubles spécifiques du langage
BILO	Bilans Informatisés de Langage Oral
C	Consonne
COST	European Cooperation in Science and Technology
DT	Développement Typique
ET	Écart-type
LD	Langue-Dépendant
LI	Langue-Indépendant
LITMUS-NWR	Language Impairment Testing in Multilingual Settings – Nonword Repetition
Mo-DT	Monolingue au développement typique
Mo-TSL	Monolingue avec troubles spécifiques du langage
MRI	Modèles de Réponse à l'Item
N-EEL	Nouvelles épreuves pour l'examen du langage
TAK	Taaltoets Alle Kinderen
TSL	Troubles Spécifiques du Langage
V	Voyelle

« Les tests valent ce que valent les idées
qui ont présidé à leur construction »

(Huteau et Lautrey, 2006)

1 Introduction

Actuellement, les mouvements migratoires aboutissant à des changements démographiques dans bon nombre de pays (OCDE, 2013) accroissent significativement le nombre d'enfants grandissant dans un contexte d'acquisition bilingue. Dans cette perspective, de réelles difficultés se posent aux cliniciens chargés de dépister et diagnostiquer les troubles du langage, notamment en France. À notre connaissance, seules deux batteries de tests pour enfants bilingues sont disponibles, en Allemagne¹ et aux Pays-Bas.² En effet, il n'existe pas d'outils spécifiques en français d'évaluation du langage dans un contexte de bilinguisme, alors même que l'on trouve des similarités entre les manifestations langagières des enfants monolingues avec trouble du langage et celles des enfants bilingues (Paradis, 2010). Les orthophonistes sont rarement en mesure d'effectuer des évaluations dans la langue maternelle des enfants, ce qui conduit à des risques de sous-diagnostic ou de sur-diagnostic.

Cette question a fait naître un nouveau champ de recherches dans le domaine linguistique : l'étude des enfants bilingues avec trouble du langage et la recherche de marqueurs diagnostiques de ces troubles. C'est dans cette optique que l'Action Cost ISO804 « Language Impairments in a Multilingual Society : Linguistic Patterns and the Road to Assessment » a été mise en place. Son but est de coordonner les recherches visant à décrire les habiletés linguistiques et cognitives des enfants bilingues avec trouble du langage et ainsi améliorer leur évaluation.

Dans ce cadre de recherche, les linguistes de l'équipe 1 de l'Unité Inserm 930 de Tours ont créé une épreuve expérimentale de répétition de non-mots, LITMUS-NWR³, composée de 71 items. Dans cette épreuve, on demande aux enfants de répéter des formes

¹ Batterie LiSe-DaZ (*Linguistische Sprachstandserhebung – Deutsch als Zweitsprache*) (Schulz et Tracy, 2011)

² Batterie TAK (Taaltoets Alle Kinderen) (Verhoeven et Vermeer, 2002)

³ Language Impairment Testing in Multilingual Settings – Nonword Repetition

phonologiques qui ne constituent pas des mots réels du langage. Plusieurs études ont montré qu'une faible performance dans une tâche répétition de non-mots pouvait être un marqueur de trouble du langage (Dollaghan et Campbell, 1998 ; Coady et Evans, 2008 ; Ferré et al., 2012). Dans le cas d'enfants bilingues, cette tâche possède un certain nombre d'avantages. Par définition, les non-mots ne faisant pas partie d'un lexique, leur répétition n'est pas affectée par l'exposition à une langue. L'épreuve LITMUS-NWR a ainsi été conçue pour permettre une évaluation de la complexité phonologique pour des enfants bilingues. Le caractère discriminant de cette épreuve ayant été mis en évidence par des études antérieures (Lelièvre, 2013, Grimm et al., 2014), la présente étude a pour but de réduire l'épreuve pour qu'elle soit plus facilement utilisable en clinique.

La réduction d'un outil de mesure clinique implique une méthodologie psychométrique basée sur des modèles mathématiques et statistiques. Elle comprend une phase d'étude des caractéristiques statistiques des items appelée *analyse des items*. En choisissant le modèle d'analyse le plus adapté aux types de données dont nous disposons, cette étape permettra de proposer une version écourtée du test par l'élimination des items les moins discriminants. L'épreuve a été administrée à 85 enfants âgés de 5;2 ans à 8;10 ans, bilingues et monolingues, répartis en deux groupes : 34 enfants avec trouble spécifique du langage (TSL) suivis en orthophonie, et 51 enfants au développement typique (DT).

Nous présenterons tout d'abord les enjeux d'une tâche de répétition de non-mots évaluant la complexité syllabique dans un contexte de bilinguisme. Puis, après nous être penchés sur les caractéristiques psychométriques d'un instrument de mesure, nous préciserons les processus d'élaboration d'un tel outil, ainsi que les modèles d'analyses d'items permettant sa réduction. Nous exposerons ensuite notre démarche méthodologique et les résultats de notre analyse d'items pour l'épreuve LITMUS-NWR, aboutissant à sa réduction. Ces résultats seront enfin discutés afin de mieux cerner l'influence des variables entrant en jeu dans les performances à ce type de tâche, pour proposer aux enfants et aux orthophonistes un outil adapté à leurs besoins.

2 Problématique

2.1 *Les tests de Répétition de Non-mots*

2.1.1 Intérêt

La tâche de répétition de non-mots est un outil clinique dont le but est la détection de troubles du langage chez l'enfant. Plusieurs études ont démontré son efficacité, montrant que certains niveaux de performance en répétition de non-mots étaient de puissants prédicteurs permettant notamment de différencier les enfants suivis pour trouble du langage des enfants non suivis (Gathercole et al., 1994).

Stokes et Klee (2009) ont administré une épreuve de répétition de non-mots développée pour l'occasion à plusieurs centaines de très jeunes enfants, accompagnée de tests normés de langage et d'un questionnaire parental. Les résultats à l'épreuve permettaient de distinguer les enfants présentant un retard de parole de ceux au développement typique. Dollaghan et Campbell (1998) ont étudié les performances à une tâche courte de répétition de non-mots, de 60 enfants entre 6;0 et 9;9 ans, suivis pour trouble du langage (« enrolled in language intervention ») et non suivis (« developing language normally »). Les résultats étaient ensuite comparés à ceux d'un test de langage normé⁴. Cette étude révèle tout d'abord que les enfants avec trouble et ceux au développement typique ont des performances différentes en répétition de non-mots. D'autre part, une comparaison du coefficient de vraisemblance (« likelihood ratio ») de cette tâche avec celui d'un test classique (sur une population de 85 enfants) montre que la répétition de non-mots discrimine les enfants suivis mieux qu'un test de langage classique. Weismer et al. (2000), dans une étude longitudinale, ont examiné les performances de 581 enfants d'âge scolaire sur cette même tâche de répétition de non-mots. Les résultats confirment que les performances des enfants suivis pour trouble du langage sont déficitaires, comparées à celles des enfants contrôle. Ils notent cependant que les difficultés dans la répétition de non-mots ne concernent pas uniquement les

⁴ TOLD-2, Hammill & Newcomer 1998 : batterie en anglais comprenant plusieurs subtests « classiques », aboutissant à un SLQ : « quotient de langage parlé »

enfants avec TSL si l'on s'en tient à la définition traditionnelle de cette catégorie : les enfants avec trouble du langage non spécifique montrent des performances similaires à celles des enfants avec TSL. Ils suggèrent que la multiplicité des causes pouvant entrer en jeu dans les déficits en répétition de non-mots (à l'intérieur même du groupe des enfants avec TSL autant que dans d'autres types de troubles du langage) rendent cette tâche utile au dépistage de trouble du langage, mais non suffisante pour établir la spécificité du trouble. D'autres épreuves doivent donc lui être associées pour établir un profil langagier le plus juste possible.

En effet, les processus linguistiques sous-jacents dans la tâche de répétition de non-mots sont nombreux. Selon Coady et Evans (2008), ils concernent la perception de la parole, l'encodage (segmentation du signal en unités de parole qui seront stockées en mémoire) et l'assemblage phonologiques (planification motrice), ainsi que l'articulation. La diversité de ces habiletés rend difficile la détermination de ce que mesure exactement cette épreuve. Sur ce point, Coady et Evans (2008) ajoutent qu'il n'existe pas de réel consensus. D'après leur revue de la littérature concernant l'usage de la tâche de répétition de non-mots depuis les années 1970, celle-ci a été utilisée pour mesurer la mémoire de travail, la mémoire phonologique, les processus d'accès au lexique, l'assemblage et la segmentation phonologiques, ainsi que la planification motrice. Ces compétences se développant simultanément chez l'enfant, l'interprétation précise des résultats peut sembler problématique (Thordardottir et al., 2011). Il n'en reste pas moins que les habiletés nécessaires à la répétition de non-mots font de cette tâche un outil puissant de description du langage de l'enfant, sensible aux troubles du langage : quelle que soit la mesure analysée, on rapporte toujours de plus faibles performances chez les enfants avec TSL que chez les enfants au DT (Gathercole et al., 1994 ; Weismer et al., 2000 ; Marshall et al., 2002 ; Coady et al., 2010 ; Ferré et al., 2012).

C'est en manipulant les paramètres choisis pour la construction des non-mots (longueur des items, structure syllabique, etc.) que la tâche permet d'explorer les différentes compétences entrant en jeu. Même si on ne peut qu'admettre qu'une mémoire de travail déficitaire puisse affecter la répétition (Gathercole et al., 1994), il est possible de minimiser l'influence de ce facteur : par exemple en contrôlant la longueur des non-mots, en analysant les résultats en fonction du nombre de syllabes, indépendamment d'autres tests contrôlant la mémoire de travail (empan de chiffres). Ainsi, on peut évoquer l'influence d'autres facteurs,

sur le versant plus typiquement phonologique, notamment celui de la complexité phonologique.

2.1.2 La notion de complexité en phonologie

La complexité phonologique est un paramètre linguistique intervenant dans la description de l'acquisition du langage ainsi que dans différentes théories phonologiques des troubles de développement du langage. Selon Parisse et Maillart (2006), la complexité phonologique serait même « le seul paramètre linguistique d'ordre phonologique qui permette de classer les performances en phonologie » (2006 : 141). Sa définition exacte est encore discutée. Néanmoins, Ferré et al (2012) suggèrent que l'étude du développement pathologique du langage pourrait aider à identifier les zones du langage où intervient la complexité. Les structures complexes seraient ainsi acquises plus tardivement dans le développement phonologique et constitueraient des zones d'erreurs et d'évitement. La complexité phonologique interviendrait en français à au moins deux niveaux structurels du langage : au niveau segmental et au niveau syllabique. Dans d'autres langues comme l'anglais, l'accent tonique serait également une variable influençant la complexité (Ferré et al., 2012).

La notion de complexité met en jeu la notion de marque. Une structure marquée est considérée comme étant plus complexe qu'une structure non marquée, c'est-à-dire qu'elle contient, par exemple, « plus de segments ou de traits phonologiques » (dos Santos, 2007 : 73). Plusieurs segments consonantiques mis côte à côte sont appelés groupes consonantiques (ou clusters). Les groupes consonantiques peuvent apparaître en position initiale, médiane ou finale d'un mot. Certains segments apparaissent de façon plus fréquente dans ces groupes complexes. Ce sont les consonnes sonantes [ʁ] et [l], la fricative alvéolaire non-voisée [s], et les semi-voyelles [j], [w] et [ɥ]. Ils seront considérés comme étant complexes, ainsi que les nasales [m], [n], [ɲ], [ɳ] (Ferré, 2006).

La notion de complexité syllabique dépend de la théorie qui sous-tend la définition de la syllabe (Ferré et al. 2012). On peut vraisemblablement considérer qu'une syllabe de type CCVC⁵ ou CVCC est plus complexe qu'une syllabe de type CV – parce qu'elle est constituée

⁵ C pour toute consonne, V pour toute voyelle

de plus d'éléments, et parce qu'il y a plus de consonnes au début ou à fin de la syllabe (Maddieson, 2006 in Ferré et al. 2102). Cependant, la syllabation d'un groupe consonantique complexe, et donc la position de ses éléments dans les frontières du mot, dépendent de la théorie de la syllabe choisie pour la définir. Selon Meynadier (2001), la syllabe est une unité linguistique dont la taille se situe entre le mot et le segment. Dans toutes les langues, elle semble relever de l'intuition et apparaît ainsi avoir une existence psychologique. Cependant, les difficultés commencent lorsqu'il s'agit de délimiter ses frontières : selon les locuteurs, la frontière syllabique d'un mot comme « prestige » sera ressentie comme étant avant le /s/ ([pʁɛ.stiʒ]) ou avant le /t/ ([pʁɛs.tiʒ]).

Chaque langue possède ses propres contraintes phonotactiques, c'est-à-dire qu'elle exerce des restrictions sur la manière dont les phonèmes sont distribués dans la chaîne segmentale, et donc dans la syllabe, à laquelle on doit se référer pour les établir (Meynadier, 2001). Traditionnellement (modèle de la syllabe de Kaye et Lowenstamm, 1984), la syllabe est une structure arborescente hiérarchisée et stable, constituée de quatre éléments : l'attaque, la rime, le noyau et la coda. Les deux constituants de base sont l'attaque (A) et la rime (R). La rime peut se diviser en un noyau (N) et une coda (Co) (cf. figures 1 et 2). L'attaque, la rime et le noyau sont des constituants obligatoires. Lorsqu'il n'y a pas d'élément phonologique venant remplir ces positions, l'attaque est dite « vide », mais elle est représentée. La coda est facultative : lorsqu'il n'y a pas de segment en position de coda, on ne la représente pas. Tous les constituants peuvent brancher, c'est-à-dire se subdiviser en plusieurs éléments (dominer plusieurs segments). En français, l'attaque et la coda sont des constituants de type consonantique, le noyau est vocalique. La syllabe non marquée sera donc, selon ce modèle, de forme CV (Brousseau et Nikiema, 2001).

Figure 1 : Schéma général de la syllabe

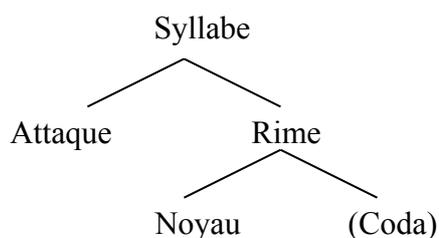
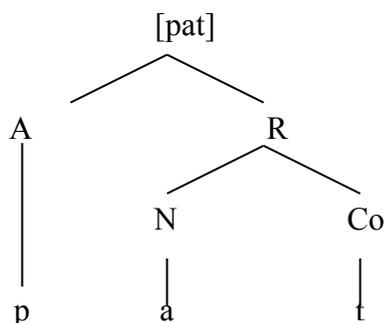
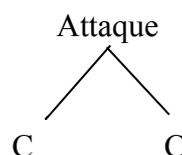


Figure 2 : Structure syllabique du mot « patte » [pat]

Dans les langues romanes, et notamment en français, deux types d'attaque sont possibles : attaque simple ou attaque branchante (cf. figure 3). Une attaque branchante est constituée de deux éléments, comme dans la syllabe CCV « bra » [br̥a]. Dans toutes les langues du monde, il existe plus de types de consonnes possibles en tête qu'en deuxième position d'attaque. Par exemple en français, les occlusives ne peuvent pas se trouver en position dépendante, cette position étant le plus souvent occupée par une liquide ([ʁ], [l]) (dos Santos, 2007).

Figure 3 : Types d'attaques en françaisa. Attaque simpleb. Attaque branchante

On considère qu'une structure syllabique est complexe si elle est constituée d'au moins un constituant branchant (Ferré, 2006). En français, les groupes consonantiques se retrouvent en attaque, et en coda. Les sources de complexité sont donc nombreuses. Cependant, selon le modèle syllabique choisi, l'analyse de la position des groupes consonantiques dans la syllabe diffère. Ainsi, ils peuvent être considérés comme faisant partie de la fin d'une syllabe ou du début de la syllabe suivante : par exemple, selon le modèle de

syllabe, le mot « souffle » [sufl] serait constitué d'une seule syllabe ou de deux, le groupe consonantique [fl] constituant alors l'attaque branchante d'une nouvelle syllabe (Harris & Gussmann, 1998, in Ferré et al. 2012).

Dans le modèle rythmique d'Angoujard (1997), une séquence de position est fondée sur une alternance et une répétition d'un modèle rythmique de base. La structure sonore s'organise sur un rythme binaire fondamental de type CV, avec une alternance entre un creux de sonorité et un sommet. Trois positions sont alors possibles : la position 1 est associée à un segment non vocalique (creux de sonorité = attaque), en position 2 se trouve un segment vocalique - ou une sonante, sous conditions (pic de sonorité = noyau), la position 3 est plus ou moins équivalente à une coda (creux de sonorité facultatif). À cette troisième position peut être seulement associé un segment de sonorité égale ou inférieure à celle du segment associé au sommet (donc une voyelle longue, par exemple en italien), dont la sonorité doit être au moins égale à celle des sonantes. Cela veut donc dire qu'une occlusive ne peut se trouver en position 3. L'analyse de la position de la consonne finale, selon qu'elle est sonore ou occlusive, n'est donc pas la même (le [ʁ] de /par/ est en position 3 – équivalant à une coda, le [t] de /pat/ en position 1 - attaque de la syllabe suivante) (cf. figure 4).

Figure 4 : /par/ vs /pat/ selon le modèle rythmique d'Angoujard (1997, in Ferré et al. 2012)

1	2	3		1	2	1	2
	x			x		x	x
x	x	x		x	x	x	x
p	a	ʁ		p	a	t	

Selon ce modèle, une consonne finale peut être considérée comme une attaque suivie d'un noyau vide. En raison de leur sonorité, seules les sonantes, les glides et la fricative alvéolaire non-voisée /s/ peuvent se trouver en position 3, c'est-à-dire en deuxième partie d'un constituant branchant. Par exemple, lorsqu'un enfant produit [pa:ti] pour [paʁti], il aurait, selon ce modèle, acquis la notion que cette position requiert un segment de sonorité égale à celle du segment qui lui est associé (il produit une voyelle longue), mais il n'aurait pas encore acquis la notion de baisse de sonorité dans cette troisième position (il ne produit pas le [ʁ], qui est une sonante). Comme nous l'avons vu, les sonantes, les glides et la fricative alvéolaire non-voisée /s/ sont des constituants des groupes consonantiques en français, qui représentent donc des zones de complexité. Le statut de la fricative alvéolaire non-voisée /s/

est particulier en français. C'est la seule consonne qui puisse se retrouver en première position d'une attaque triple (comme dans « stressé » [stʁese]). De plus, on la retrouve (contrairement par exemple à l'anglais) branchant avec une consonne initiale (comme dans « psychologie » [psikoloji]).

On pourrait alors suggérer une hiérarchie syllabique en fonction de la complexité. L'hypothèse qui en découlerait est que l'ordre des acquisitions chez les enfants suivrait cette hiérarchie (Levelt et al., 2000 ; Ferré, 2006) : les syllabes les moins complexes (et donc les moins marquées) seraient acquises en premier. Les tâches de répétition de non-mots permettraient d'explorer la complexité phonologique en analysant précisément le traitement phonologique des sujets.

Dans tous les cas, on considère que le système phonologique des enfants, lors du développement normal du langage, va se stabiliser aux alentours de l'âge de 5 ans (McLeod et al., 2001 ; Bowen, 2007). Une étude récente (Rvachew et al., 2013), portant sur la normalisation d'un test de dépistage de phonologie sur une population d'enfants québécois, indique même que le développement phonologique des enfants parlant français ne serait pas complet avant l'âge de 7 ans.

2.1.3 Répétition de non-mots et complexité phonologique

Gallon et al. (2007) ont administré un test de répétitions de non-mots élaboré par van der Lely et Harris⁶ (van der Lely et Harris, 1999, in Marshall et al., 2002) à un groupe d'adolescents (12-20 ans) avec TSL et à un groupe d'enfants (4-8 ans) au DT. Ce test prenait pour la première fois en compte les complexités métriques et syllabiques. Les résultats ont révélé que, le nombre de syllabes du non-mot restant le même, les performances des enfants avec TSL diminuent quand la complexité phonologique augmente. Des résultats similaires ont été rapportés dans l'étude de Ferré et al. (2012), utilisant une adaptation française du test de van der Lely et Harris⁷. En faisant varier les éléments de complexité structurale (nombre de segments dans un groupe consonantique complexe, position des segments dans le groupe, nombre de séquences consonantiques, etc.), on peut mettre en évidence le rôle de la

⁶ TOPhS (Test Of Phonological Structures)

⁷ SSS test (Syllabic Structure & Segments)

complexité phonologique dans les productions d'enfants avec trouble du langage et ainsi étudier la nature de leur déficit phonologique.

Outre le fait que la tâche de répétition de non-mots est prédictive de trouble du langage, elle permet également de contourner un certain nombre de facteurs influençant les performances et présents dans d'autres types de tests : la fréquence et la familiarité des mots du lexique, l'âge d'acquisition, les contextes socio-économiques et culturels (Dollaghan et Campbell, 1998 ; Weismer et al., 2000 ; Coady et Evans, 2008 ; Gutierrez-Clellen et Simon-Cerejido, 2010). L'épreuve de répétition de non-mots LITMUS-NWR étant conçue pour être utilisée dans un contexte de bilinguisme, nous étudierons plus spécifiquement cet aspect du problème : en quoi une tâche de répétition peut-elle permettre d'éviter le biais supposé du bilinguisme ?

2.1.4 Le bilinguisme est-il un facteur influençant les performances aux tests ?

Quelle que soit l'hypothèse sur les origines des troubles phonologiques chez les enfants avec TSL (déficit du traitement temporel de l'information auditive, du traitement phonologique initial, de la mémoire phonologique à court terme, hypothèse de la complexité grammaticale computationnelle etc.) et quels que soient les difficultés et les processus phonologiques présents dans ces troubles, la question de savoir si un contexte de bilinguisme minore ou majore les difficultés d'un enfant avec TSL reste discutée (Paradis, 2010 ; Grimm et al., 2014).

En pratique, bon nombre de difficultés subsistent pour le clinicien dans la détection des troubles du langage en situation de bilinguisme. Les problèmes de sur-/sous-diagnostic de ces troubles sont liés au fait qu'il n'existe pas de tests spécifiques à ce contexte. Or, des similitudes entre certaines caractéristiques du développement des enfants bilingues et celles d'enfant monolingues avec TSL ont été mises en évidence (Paradis, 2010 ; Rothweiler et al., 2012) : les études soulignent des difficultés sur la production de pronoms clitics objets, sur les marqueurs de temps et d'accord sujet/verbe (Paradis et Crago, 2003 ; Paradis, 2005). Plus instinctivement, certains cliniciens (et certains parents) peuvent considérer que des productions atypiques pouvant sous-tendre un trouble du langage ne seraient dues qu'à une acquisition bilingue. À moins de connaître parfaitement le développement linguistique

bilingue, par ailleurs encore discuté, il est en effet très difficile de faire la part des choses entre le trouble de langage réel et le résultat de l'interférence entre deux langues, notamment chez les enfants les plus jeunes.

Afin de minimiser ces risques de sur-/sous-diagnostic, une solution sur laquelle s'accordent de nombreux auteurs serait d'évaluer les enfants dans leurs deux langues (Yavas, 2004 ; Gutierrez-Clellen et Simon-Cerejido, 2010 ; Thordardottir et al., 2011 ; Thordardottir et al., 2012), ou au moins dans la langue dite dominante. Or, actuellement, l'idée même d'une telle évaluation est très difficile à concevoir, certes à cause du manque d'outils diagnostiques, mais également à cause d'un évident manque de temps et de moyens. Le questionnaire parental pourrait être une mesure permettant de prendre en considération la diversité des facteurs qui influencent le développement bilingue dans l'évaluation. Des études (Paradis et al., 2010 ; Saumon, 2012) mettent ainsi en évidence les capacités discriminatoires d'un tel questionnaire, qui est un outil précieux pour le clinicien. Il n'en reste pas moins que pour poser un diagnostic précis de trouble du langage, le questionnaire devrait être accompagné d'une évaluation normée, ou au moins standardisée.

Idéalement, le clinicien accueillant un enfant bilingue devrait donc avoir à sa disposition plusieurs outils : le questionnaire parental, des tests de langage spécifiquement conçus pour ce contexte, ainsi que la possibilité d'évaluer l'enfant dans sa langue dominante. Nous avons vu que les épreuves de répétition de non-mots participaient à la mise en place d'un diagnostic de trouble du langage. Qu'en est-il en situation de bilinguisme ?

2.1.5 Répétition de non-mots et bilinguisme

La tâche de répétition de non-mots a prouvé son utilité pour aider à discriminer les enfants avec TSL des enfants au DT. Nous avons vu que cette épreuve permettait de minimiser les biais culturels et dialectaux. Elle permet également de contourner les facteurs tels que la familiarité, la fréquence, l'âge d'acquisition, en minimisant les apports des connaissances lexicales. Il est à noter que pour obtenir des résultats interprétables dans ce sens, un non-mot ne doit pas ressembler à un mot du lexique. Gathercole et al. (1994) ont mis en évidence dans leur étude un effet de similitude (« wordlikeness effect ») : les non-mots ressemblant à de vrais mots sont de manière générale mieux répétés. Frisch et al. (2000)

montrent que les non-mots dont les constituants ont une haute probabilité de similitude sont mieux reconnus par les participants de leur étude, suggérant qu'ils utilisent pour cela leurs connaissances des gabarits lexicaux fréquents.

Afin de savoir si une épreuve de répétition de non-mots peut aider à poser un diagnostic de trouble du langage chez des enfants bilingues, il faut d'abord savoir si le bilinguisme a une influence sur les tests de répétition de non-mots pour des enfants au DT. Sur cette question, encore peu d'études ont mis en regard les performances d'enfants bilingues et monolingues sur le même test. Armon-Lotem et Chiat (2012) ont comparé les performances de 75 enfants bilingues russe-hébreu avec celles de 22 enfants monolingues hébreu et 26 enfants monolingues russe. Les résultats montrent que les performances du groupe d'enfants bilingues sont similaires à celles des monolingues sur ce test. Selon cette étude, le bilinguisme n'aurait donc pas d'effet sur les scores de répétition de non-mots pour des enfants au DT, suggérant ainsi qu'un test de répétition de non-mots aurait le potentiel d'identifier un trouble du langage chez des enfants bilingues (Thordardottir et. al, 2012 ; Lelièvre , 2013). Gutierrez-Clellen et Simon-Cereijido (2010) ont fait passer une tâche de répétition de non-mots à 144 enfants bilingues espagnol-anglais, dont 49 avec trouble du langage, afin de tester leur mémoire de travail verbale. Ils présentaient des listes de non-mots adaptées à chacune des langues, en termes de contraintes phonotactiques. Les résultats montrent des différences significatives entre les enfants avec trouble du langage et les enfants au DT, quelle que soit la langue. Par ailleurs, les auteurs plaident pour une évaluation bilingue, précisant que tester les enfants uniquement sur leur langue dominante pourrait mener à des erreurs de diagnostic.

Une épreuve de répétition de non-mots pourrait donc être un outil efficace pour aider les cliniciens à poser un diagnostic de trouble du langage, notamment dans un contexte de bilinguisme. C'est dans ce but que l'épreuve LITMUS-NWR a été construite. Notre étude porte sur la réduction de cette épreuve. Il nous semble alors nécessaire de situer plus précisément cette étape dans le long processus d'élaboration d'un outil permettant le test. Le choix des items définitifs ne peut se faire sans comprendre d'abord ce qu'est un instrument de mesure, quelles doivent être ses caractéristiques, comment et selon quels modèles il se construit.

2.2 *Les tests, instruments de mesure*

2.2.1 Qu'est-ce qu'un test ?

C'est à la fin du XIX^{ème} siècle, dans le domaine de la psychologie expérimentale, que l'on voit apparaître pour la première fois le terme de « test mental ». Depuis, cette notion a été étendue à d'autres champs d'application (éducation, industrie, médecine, orthophonie etc.). Malheureusement, la multiplication du nombre de tests a rendu difficiles les tentatives de définition claire et précise de ce que représente un test. Les Recommandations Internationales sur l'Utilisation des Tests (Vrignaud et al., 2003 : 13) précisent que « [t]oute tentative pour fournir une définition précise d'un test ou du testing en tant que processus échouera vraisemblablement », à cause notamment de la diversité des procédures utilisées.

Parmi toutes les définitions existantes, la plus régulièrement citée dans la littérature est celle de Pichot (1997 : 5) :

On appelle test mental une situation expérimentale standardisée servant de stimulus à un comportement. Ce comportement est évalué par une comparaison statistique avec celui d'autres individus placés dans la même situation, permettant ainsi de classer le sujet examiné, soit quantitativement, soit typologiquement.

Il faut préciser ici que ce terme de « test mental » est à prendre au sens large de connaissance des processus mentaux, par opposition aux tests « élémentaires » - par exemple les anciens tests anthropométriques où les individus n'étaient pas des objets d'étude mais, comme le note Martin (1997 : 109) « fournissaient le terrain expérimental » comme « représentants de l'espèce humaine ». Dans ce cadre, les tests de langage utilisés en orthophonie sont considérés comme des tests mentaux.

Il existe de nombreux types de tests, que Chartier et Loarer (2008 : 40) proposent de classer en fonction de quatre critères :

- le domaine évalué (intelligence, personnalité, créativité, connaissance, etc.),
- le format (papier/crayon, performance, test informatisé, etc.),
- le mode de passation (individuel ou collectif),
- le type de réponse (questions ouvertes ou fermées, différents types de notation des scores).

Quels que soient les critères choisis, le test va donc mesurer un comportement induit par une situation expérimentale : on passe d'un comportement à un score. Mais c'est bien la comparaison statistique qui va donner tout son sens à la mesure.

2.2.2 Méthodologie de la mesure : standardisation et normalisation

Les outils de mesure que sont les tests entrent dans une démarche clinique. Ils proposent des solutions à des questions qui impliquent un regard sur les processus en œuvre dans les épreuves, et pas uniquement sur les sujets qui y sont soumis. On mesure une capacité, un niveau de développement. En ce sens, les tests sont des outils cliniques, favorisant une aide au diagnostic, par la mise en correspondance des réponses des sujets les uns par rapport aux autres. Cette dimension clinique implique une méthodologie. Cette dernière s'appuie sur une réflexion basée sur le sens que l'on peut donner à la notion de « mesure ».

A partir de la définition de Pichot, deux notions fondamentales peuvent être isolées : la standardisation et la normalisation. **La standardisation** consiste à contrôler la situation de passation, c'est-à-dire à présenter à tous les sujets exactement le même test, dans les mêmes conditions, avec la même méthode de calcul de score. **La normalisation** (aussi appelée étalonnage) permet l'interprétation du score. On classe ainsi des notes standard pour obtenir un tableau de correspondance entre la note du sujet au test et la variable observée. Cette note sera comparée à un groupe de référence.

2.2.3 Test, épreuve, tâche ... ?

En anglais comme en français, les termes test (« test » en anglais), tâche (« task », « testing task »), sont utilisés par les chercheurs et les cliniciens. Le français ajoute à cette liste le terme d'*épreuve*. Ces mots différents recouvrent également des réalités différentes. A l'instar de Rondal et al. (2003), nous parlerons de *test* lorsqu'il s'agit d'une procédure d'évaluation standardisée et normalisée. Lorsque la procédure est standardisée et non normalisée, on parle d'« épreuve d'évaluation » (Rondal et al., 2003 : 384). Dans ce cas, l'anglais utilise le terme de tâche (« task »), que l'on retrouve également en français. Il est à noter cependant que le terme d'« épreuve » est également utilisé pour qualifier une partie

d'un test comportant plusieurs tâches. Il serait alors plus précis de parler dans ce cas de « subtest ».

La procédure d'évaluation LITMUS-NWR, au cœur de cette étude, est en cours de construction, et ne peut donc pas encore être appelée « test ». C'est pourquoi nous utiliserons préférentiellement les termes d'épreuve ou de tâche pour la qualifier.

2.2.4 Différents types de mesure : tests critériés vs tests normatifs

Depuis la célèbre profession de foi du psychologue Thorndike (1926 : 38) « If anything exists, it exists in some amount. If it exists in some amount, it can be measured ⁸ », la recherche en psychométrie a permis de préciser la notion de mesure. En faisant la distinction entre la mesure se référant à un critère (« criterion-referenced measurement ») et celle se référant à une norme (« norm-referenced measurement »), Glaser (1963) a permis de définir deux types de tests : les tests critériés et les tests normatifs.

Les tests critériés jugent les performances des individus en fonction d'un niveau d'acquisition des objectifs, alors que les résultats des tests normatifs s'interprètent à la lumière de normes, fondées sur les différences entre sujets. Cette distinction entraîne des différences fondamentales dans la conception des tests, notamment en termes de type d'échantillonnage : les tests critériés utilisent un échantillonnage d'items, alors que les tests normatifs s'appuient sur un échantillonnage de personnes. Laveault et Grégoire (1997 : 296) précisent donc que lors de la construction d'un test critérié, les items sont sélectionnés en amont, afin de définir un « univers de contenu ». Or, dans le cas d'une mesure normative, il est difficile de déterminer à l'avance quels items vont être les plus discriminants : il est alors nécessaire de mettre le test à l'essai, sur un échantillon représentatif de la population.

Les tests de langage utilisés en orthophonie posent, à l'étape même de leur conception, un certain nombre de problèmes auxquels la psychométrie, en tant que science étudiant les techniques de mesure, essaie d'apporter quelque réponse. Ceci implique également qu'elle s'intéresse aux techniques de *validation* de ces mesures. Parce que les tests sont des outils

⁸ « Si il existe quelque chose, il en existe une quantité. S'il en existe une quantité, elle peut être mesurée. » (notre traduction)

d'aide au diagnostic et à la prise de décision impliquant des individus, il est absolument nécessaire de pouvoir mesurer la mesure : ce sont les qualités métrologiques, appelées également caractéristiques psychométriques.

2.2.5 Caractéristiques psychométriques : mesurer la mesure

Selon Chartier et Loarer (2008 : 43), « les théories et les méthodes psychométriques permettent d'élaborer des tests dans l'objectif de garantir une objectivité de la mesure », notamment dans leurs phases d'élaboration. Pour être considérés comme des outils d'évaluation valables, les tests doivent donc répondre à des qualités psychométriques, au nombre de trois : la validité, la fidélité et la sensibilité.

- **Validité : le point crucial de l'évaluation d'un outil**

La validité est le degré avec lequel le test mesure ce qu'il a pour but de mesurer. Une définition psychométrique de la validité est donnée par Selltitz et al. (1977), ajoutant l'individu au cœur de la notion. En effet, ils notent que le degré de validité d'un test se mesure à partir d'observations de scores d'individus. Lorsqu'on cherche à mesurer une compétence, les différences de scores doivent traduire des différences réelles entre les sujets par rapport à cette compétence, et non pas simplement refléter des erreurs. On peut alors imaginer différentes manières d'établir la validité d'un test. Il a donc fallu faire la distinction entre plusieurs types de validité (on en recense plus de 30), dont trois se sont imposés dans la littérature (Chartier et Loarer, 2008) : validité de contenu (interne), validité prédictive (externe ou critérielle), validité de construit. Chaque type de validité mesure une caractéristique différente.

La *validité de contenu* (ou validité interne) concerne, comme le notent Dickes et al. (1994 : 49) « la pertinence et la représentativité des items de l'instrument de mesure ». Elle suppose la définition d'un univers d'items et la structuration d'un domaine : il s'agit d'organiser les items et de savoir d'où viennent les variations des réponses afin de pratiquer un échantillonnage le plus pertinent possible. Par exemple, dans le cas de l'épreuve LITMUS-NWR, il faut faire attention à ce que les non-mots choisis forment un échantillon représentatif de tous les non-mots qu'on aurait pu utiliser pour tester la complexité phonologique dans un

contexte de bilinguisme. Il faut également veiller au fait que la répétition de ces non-mots ne fasse pas appel à d'autres types de compétences, que l'on ne souhaiterait pas évaluer ici, comme la mémoire de travail (cf. section 3.2.2). La validité de contenu peut être estimée par des experts du domaine en question. On peut également calculer un coefficient de corrélation (coefficient de Bravais-Pearson) entre les résultats fournis par l'épreuve et les résultats d'un autre test, normé et standardisé, dont la validité de contenu est reconnue. Cette mesure ne sera pas effectuée dans la présente étude, car nous ne disposons pas en France de test de référence pour les enfants bilingues.

La *validité critérielle*, faisant référence à un critère externe, est peu utilisée dans les tests de langage ; elle concerne surtout les tests de diagnostic psychiatrique et d'orientation scolaire : la mesure d'une caractéristique est valide si elle permet de connaître avec le moins d'erreur possible la mesure d'une autre caractéristique (Dickes et al., 1994).

La *validité de construit* (ou validité conceptuelle, théorique ou hypothético-déductive) vise à « montrer l'existence du construit psychologique mesuré » (Dickes et al., 1997 : 49) : il s'agit de mettre le test en regard d'un modèle théorique de référence, mais également de comparer les résultats des sujets avec d'autres tests comparables. Rondal et al. (2003) notent qu'un test qui ne serait pas validé par une théorie doit être considéré comme étant faible, car la base sur laquelle il devrait être utilisé n'est pas clairement posée. Ce type de validité recouvre plusieurs aspects, dont un qui nous semble pertinent dans le cadre de notre étude : il s'agit de donner des explications statistiques aux variations que l'on observe dans la mesure des items.

Il nous faut également souligner que même s'il est encore largement employé dans les manuels accompagnant les tests, le terme de « validité » est actuellement abandonné par les chercheurs en psychométrie qui se penchent sur le problème de la mesure, au profit de celui de *validation* : « on ne valide plus un instrument de mesure mais les mesures qu'il permet d'obtenir » (Dickes et al., 1994 : 49), c'est-à-dire que l'on prend en compte non seulement l'instrument, mais également les sujets auxquels il s'applique et le contexte dans lequel il est employé.

- **Fidélité (ou fiabilité)**

Un instrument de mesure est fidèle si l'on peut reproduire son résultat. Le test administré à un sujet plusieurs fois de suite, ou par des examinateurs différents, doit donner les mêmes résultats (en excluant les éventuels effets d'apprentissage). Autrement dit, les

résultats doivent être constants : c'est la *stabilité* du test, mesuré par un coefficient (par exemple, corrélation de Pearson ou corrélation par rang de Spearman).

L'homogénéité du test mesure quant à elle la *fidélité interne* (ou consistance interne) de l'instrument de mesure : on s'assure donc que tous les items d'une épreuve mesurent la même variable. La consistance interne repose sur le postulat que le test ne mesure qu'un seul concept. Si plusieurs dimensions sont contenues dans le test, la fidélité interne sera calculée pour chacun de ces sous concepts. Le coefficient alpha de Cronbach (Cronbach, 1951) permet de la mesurer (le calcul de ce coefficient faisant partie intégrante de l'analyse des items, il sera détaillé en section 2.4.5). La fidélité du test conditionne sa validité.

Dans le cadre de notre objectif, qui est la réduction du nombre d'items de l'épreuve LITMUS-NWR, il est à noter que « plus une épreuve comporte d'items, plus elle sera représentative de la dimension à évaluer, donc plus sa fidélité augmente. En effet, la consistance de la mesure s'améliore avec le nombre d'items » (Chartier et Loarer, 2008 : 50). Cependant, un test trop long risque de ne pas être utilisé en clinique. Il s'agira donc d'arriver à un compromis entre les contraintes de fidélité et celles liées à la durée de passation.

- **Sensibilité**

La sensibilité d'un test « concerne son pouvoir discriminatif ou classificatoire » (Rondal et al., 2003 : 385). Plus un test est sensible, plus il permet de classer de manière nuancée les différences entre sujets (Bernaud 1998). La sensibilité s'évalue au moyen d'une analyse d'indicateur de dispersion comme l'écart-type. Le même compromis que pour la fidélité doit être fait : plus le test comporte d'items et plus il y a de chances qu'il soit sensible. Il est à noter que la sensibilité en tant que qualité métrologique ne doit pas être confondue avec les notions de sensibilité et de spécificité d'un outil, qui sont destinées à classer des sujets dans des catégories diagnostiques.

Rondal et al. (2003 : 385) résument ainsi les exigences auxquelles un instrument de mesure doit répondre pour être qualifié de test :

Une procédure ou une tâche peut donc être légitimement appelée test dans l'évaluation psychologique quand elle autorise la mesure, lorsqu'elle est valide, fiable et sensible, lorsqu'elle est standardisée du point de vue de son application et de sa correction, et lorsqu'on dispose d'une normalisation.

Plus pragmatiquement, Rondal et al. (2003) notent que d'autres caractéristiques doivent également entrer en jeu, notamment pour les tests de langage : la facilité et la rapidité d'administration. Ces dimensions sont peu étudiées en tant que telles dans les manuels de psychométrie, mais les argumentaires de vente des tests les utilisent pour convaincre le clinicien d'acheter l'instrument de mesure en question. Or, nous avons vu que la réduction du nombre d'items influence négativement les caractéristiques psychométriques d'un test. Rondal et al. (2003) considèrent alors que les concepteurs de tests se doivent de prouver que les items et les tâches proposées dans les tests sont réellement pertinents pour l'évaluation proposée.

Un test doit donc posséder des qualités psychométriques indispensables pour être utilisé en clinique. Ce mémoire ayant pour principal objet une étape de la construction d'un test de langage, il nous semble nécessaire de comprendre comment développer un test dans le respect de ces exigences.

2.3 Processus d'élaboration d'un test

2.3.1 Les étapes de construction d'un test

Selon Laveault et Grégoire (1997), le processus de construction d'un test doit suivre plusieurs étapes :

Étape 1

Déterminer les utilisations prévues du test : à quoi et à qui va-t-il servir ? Par exemple, on peut considérer que l'épreuve LITMUS-NWR cherche à discriminer des sujets dans une population pour laquelle l'épreuve est élaborée, et veut permettre de « comprendre le sens d'une performance » (Laveault et Grégoire 1997 : 79). LITMUS-NWR est par ailleurs une épreuve de détection de trouble, donc de dépistage.

Étape 2

Définir très précisément ce que l'on souhaite mesurer.

Étape 3

Créer les items :

- choisir le bon format,
- trouver le bon niveau de difficulté,

- choisir le nombre d'items, en fonction de la durée désirée du test et du niveau de fiabilité (un test long étant plus fiable qu'un test court).

Étape 4

Évaluer les items.

Étape 5

Déterminer les propriétés métriques du test final

La réduction d'un test appartient à l'étape 4 de la construction du test.

2.3.2 Pourquoi construire plus grand pour finalement réduire ?

Laveault et Grégoire (1997) considèrent qu'il est plus prudent de créer dans un premier temps un nombre d'items plus grand que celui qui sera utilisé dans le test final, ceci afin de ne choisir que les items possédant les caractéristiques voulues. Le test n'en sera que meilleur. C'est ce qu'on appelle « l'optimisation des caractéristiques d'un test » (Laveault et Grégoire, 1997 : 252). Les épreuves seront donc écourtées des items les moins pertinents. Habituellement, un surplus de 30 à 50% d'items est nécessaire (Laveault et Grégoire 1997 ; Dickes et al. 1994). L'épreuve LITMUS-NWR comprend actuellement 71 items. Le nombre d'items de départ répond aux exigences de recherche, mais ne correspond pas forcément aux contraintes d'une épreuve utilisée en clinique. La réduction devra se faire pour aboutir à un nombre d'items compris entre 40 et 50. Le surplus de départ est donc de 42% si on aboutit à une épreuve comprenant 50 items (c'est-à-dire à l'intérieur de la fourchette proposée par Laveault et Grégoire (1997) ou Dickes et al. (1994)), et de 77% pour une épreuve de 40 items. Dickes et al. (1994 : 166) ajoutent que l'objectif est ici de construire une épreuve fidèle mais également économique : « minimiser le nombre d'items tout en maximisant la fidélité ». Le nombre final d'items correspond au nombre nécessaire pour que la passation de l'épreuve n'excède pas cinq minutes. Actuellement, l'épreuve dure au maximum une dizaine de minutes. Cette longueur n'est pas excessive en soi. Cependant, d'autres facteurs entrent en ligne de compte :

- l'épreuve est destinée à être accompagnée d'autres tests, dans le cadre d'une batterie évaluant d'autres aspects du langage de l'enfant,
- le coût cognitif qu'elle peut représenter pour des enfants avec TSL entraîne une fatigabilité qui pourrait à elle seule être source d'erreurs,

- l'épreuve « parait » longue à la passation pour les enfants. Cette dimension subjective n'est pas à négliger. Elle est peut-être due au fait que les non-mots proposés se ressemblent, étant donné qu'ils ont été construits avec un faible nombre de consonnes différentes (cf. section 3.1.2).

2.3.3 Comment réduire ?

L'étape de l'analyse des items est incontournable lorsqu'on souhaite développer une épreuve ou un test (Laveault et Grégoire 1997). Pourtant, comme le souligne Grégoire (2005), les praticiens ignorent parfois l'existence de cette étape car les manuels des tests ne la mentionnent que rarement. Il s'agit de l'étude des caractéristiques statistiques des items. Son but est de sélectionner, parmi les items du test original, ceux qui figureront dans la version finale. Cette étape permet donc de repérer les items les moins pertinents, et donc de réduire le test aux items essentiels au but de départ fixé. Dickes et al. (1994 : 166) la définissent ainsi :

[N]ous nous situerons dans la perspective du chercheur qui élabore un instrument pour mesurer une caractéristique des sujets. Il élabore des items en rapport avec cette caractéristique et les applique à des sujets. Il doit, dans un premier temps choisir, parmi les items qu'il a créés, les plus pertinents pour cette mesure, il doit faire ce qu'on appelle une *analyse d'items*.

Bien sûr, le choix de l'ensemble des premiers items se fait dans le cadre d'une théorie. L'étape de l'analyse des items est donc fondamentale pour la théorie relative au domaine qui a servi à concevoir les items. L'analyse d'items amènera des arguments empiriques à la théorie. Elle « participe à la mise à l'épreuve de théories sur le domaine » (Dickes et al., 1994 : 167).

L'analyse des items s'effectue selon des modèles de mesure. L'objectif d'un modèle de mesure est d'encadrer l'élaboration d'un instrument de mesure, à partir d'un certain nombre de règles. Il existe un grand nombre de modèles statistiques permettant l'analyse psychométrique des items. Les concepteurs de tests en retiennent deux, dont nous tenterons d'exposer les grandes lignes : la *théorie classique du score vrai* et les *modèles de réponse à l'item* (Dickes et al., 1994 ; Grégoire, 2005 ; Chartier et Loarer, 2008).

C'est à partir de la théorie classique du score vrai qu'ont été présentées les analyses d'items, permettant notamment de réduire les tests. Cette théorie propose également un cadre permettant d'estimer la fidélité du test. Mais elle possède également des limites, repoussées en partie par une nouvelle théorie de la mesure : les modèles de réponse à l'item (MRI). Ces modèles, probabilistes, reprennent les bases d'analyse d'item, mais permettent d'obtenir des estimations « plus robustes » de la difficulté des items (Grégoire, 2005 : 44).

2.4 Théorie classique du score vrai : analyse des caractéristiques des items et estimation de la fidélité

L'analyse des items fait appel à l'utilisation de deux indices : l'indice de difficulté et l'indice de discrimination, auxquels peuvent s'ajouter le calcul de l'écart-type, celui de la corrélation item/total et celui de l'alpha de Cronbach (Dickes et al., 1994 ; Laveault et Grégoire, 1997 ; Grégoire, 2005).

2.4.1 Indice de difficulté

L'indice de difficulté d'un item, traditionnellement noté p , est la probabilité que l'item soit réussi, ce qui donne l'expression suivante :

$$p = \frac{\sum x}{n}$$

où $\sum x$ = somme des résultats obtenus à l'item et n = nombre de sujets.

Dans une épreuve comme LITMUS-NWR où tous les items sont notés de façon dichotomique, sur la même échelle (0/1), l'indice de difficulté correspond à la moyenne de l'item. Plus grande est la moyenne, plus l'item a été réussi. Par exemple, un indice de difficulté de 0,25 veut dire que 25% des sujets ont réussi l'item.

L'indice varie entre 0 et 1. Un item très difficile, c'est-à-dire échoué par la plupart des sujets, sera proche de 0. À l'inverse, un item très facile s'approchera de 1. On pourrait donc

dire que l'indice de difficulté est en fin de compte un indice de facilité. Il est important dans un test d'avoir des items échelonnés le plus régulièrement possible, pour qu'ils puissent discriminer les individus dans une séquence développementale donnée (Grégoire, 2005) : des items trop faciles ne donneraient aucun renseignement sur les enfants en fin de développement ; de même, des items trop difficiles ne renseigneraient pas sur les sujets en début de développement.

Lorsqu'un item est trop facile ou trop difficile, sa distribution est considérée comme étant asymétrique. Les items faciles permettent d'effectuer une première distinction entre sujets forts et sujets faibles. On peut considérer que si un sujet rate un item réussi par la plupart des autres, cet échec est plus significatif que s'il échoue à un item qui a fait chuter un grand nombre de ses pairs. Pour Dickes et al. (1994), les items ayant un indice de difficulté inférieur à 0,10 ou supérieur à 0,90 doivent être écartés. Cependant, dans l'épreuve LITMUS-NWR - qui s'adresse à des enfants et permet de dépister un éventuel trouble phonologique - il sera important de garder certains items faciles pour que l'enfant se sente valorisé et que sa motivation ne soit pas entamée.

Les items de difficulté moyenne ($p = 0,5$) ont un pouvoir de discrimination symétrique : ils sont donc importants surtout dans les tests qui différencient les sujets entre eux, quel que soit le score total qu'ils obtiennent. Si le test ne comporte pas d'items de difficulté moyenne, un sujet assez fort pour réussir des items faciles mais trop faible pour réussir des items difficiles ne sera pas correctement évalué. Il est donc essentiel de bien distribuer la difficulté des items.

2.4.2 Indice de discrimination

Lorsque le but de la mesure est de différencier des sujets en fonction d'un critère donné (dans le cadre de l'épreuve LITMUS-NWR, la capacité à répéter correctement des non-mots), les items doivent posséder un pouvoir de discrimination élevé : ils doivent permettre d'établir cette différence le plus finement possible. L'indice de discrimination d'un item (indice D) est la différence entre la proportion de réussite à cet item chez les sujets ayant obtenu des résultats élevés à l'ensemble de l'épreuve (groupe fort = p_+) et cette même

proportion chez ceux dont les résultats sont les plus faibles (groupe faible = p_-). La formule est donc :

$$D = p_+ - p_-$$

Kelley (1939, in Laveault et Grégoire, 1997) propose que le groupe fort se situe dans les 27% supérieurs (score total), et le groupe faible dans les 27% inférieurs. Les 46% restants constituent le groupe intermédiaire. Par exemple, dans un groupe de 15 sujets, on prendra les 4 résultats les plus élevés et les 4 les plus bas pour calculer p_+ et p_- ($0,27 \times 15 = 4,05$).

L'indice D se situe entre -1 et +1. Le pouvoir discriminatif de l'item est d'autant plus élevé que la valeur de D est proche de 1. Si D prend la valeur 0, cela signifie que le score obtenu à l'item par le groupe fort est le même que celui obtenu par le groupe faible. S'il prend une valeur négative, cela signifie que l'item a été réussi par plus de sujets ayant eu un score total faible. Ebel (1972) propose les repères suivants :

Item qui discrimine très bien	0,4 et plus
Item qui discrimine bien	0,3 à 0,39
Item qui discrimine peu	0,2 à 0,29
Item-limite, à améliorer	0,1 à 0,19
Item sans utilité réelle pour l'examen	Moins de 0,1

Lorsque les items discriminent peu, ou prennent une valeur négative, il sera nécessaire de les étudier afin de comprendre le résultat (Laveault et Grégoire, 1997).

2.4.3 Écart-type

Au calcul des indices de difficulté et de discrimination, il faut ajouter celui de l'écart type (ou déviation standard) qui correspond à la racine carrée de la variance. En effet, une évaluation doit pouvoir distinguer les sujets entre eux. L'écart-type de l'item permet de savoir « à quel point les résultats à cet item sont dispersés ou non » (Laveault et Grégoire, 1997 : 230), c'est-à-dire à quel point tous les scores sont éloignés ou rapprochés de la moyenne. L'écart-type des scores à chaque item doit donc être suffisant, parce qu'un item réussi ou échoué par tous ne nous fournit pas une information pertinente, permettant de distinguer les sujets les uns des autres. Les items à chaque extrémité de l'échelle (faciles et difficiles)

présentent les écarts types les plus faibles, les items de difficulté moyenne ont les écarts types les plus élevés. Il s'agira, dans le cadre de la réduction du test, d'examiner les moyennes et les écarts-types, et éliminer les items qui présentent des moyennes extrêmes et des écarts-types faibles. .

2.4.4 Corrélation item/total ($r_{item/total}$)

Il s'agit de la corrélation entre le score à chaque item et le score total à l'épreuve. C'est une mesure qui s'apparente à celle de l'indice de discrimination. Ce calcul nous donne une indication sur la relation entre l'item et l'ensemble de l'épreuve (Grégoire, 2005). Normalement, la probabilité de réussir un item, quel qu'il soit, est plus grande pour les sujets ayant obtenu un score total élevé à l'ensemble de l'épreuve que pour les sujets ayant obtenu un score plus faible. Cependant, certains items peuvent ne pas être corrélés avec le score total, témoignant du fait que les performances à ces items ne dépendent pas des performances aux autres. Ces items seront dit hétérogènes et devront en principe être éliminés, car il est vraisemblable qu'ils ne mesurent pas la même caractéristique que les autres (Grégoire, 2005). La corrélation entre un item et le score total doit être au moins égale à 0,30 (Laveault et Grégoire, 1997). Le calcul de cette corrélation doit être fait en retirant le score de l'item en question du score total, pour éviter une élévation artificielle du coefficient. Cette mesure se rapproche d'une mesure de cohérence interne, mais elle concerne les items pris isolément. Or, l'analyse des items doit également prendre en compte l'ensemble de l'épreuve.

2.4.5 Le coefficient alpha de Cronbach : une estimation de la fidélité

Le coefficient alpha de Cronbach évalue la cohérence de l'ensemble des items (Cronbach, 1951), c'est-à-dire à quel point les réponses fournies aux différents items mesurent bien ce que l'épreuve mesure effectivement dans son ensemble. Il traduit donc le degré d'homogénéité d'une épreuve, mesurant ainsi sa fidélité interne. Des items fortement corrélés entre eux indiquent qu'ils mesurent bien le même construit, c'est-à-dire que l'échelle possède une bonne cohérence interne. Autrement dit, le construit sera considéré comme étant unidimensionnel. Si des items qui constituent l'échelle mesurent un construit multidimensionnel, le coefficient alpha sera probablement bas. Il doit donc être calculé en

dernier, une fois que les autres indices ont permis de retirer les items les moins performants de l'épreuve, c'est-à-dire, en pratique pour notre étude, une fois que la réduction de l'épreuve aura été effectuée. Une estimation de ce coefficient peut être également effectuée en amont sur le test non réduit, afin de comparer sa valeur une fois la réduction effectuée. Le coefficient alpha peut prendre plusieurs valeurs, s'échelonnant entre 0 et 1. Les auteurs ne s'accordent pas sur la valeur minimale à obtenir pour que le test soit considéré comme étant fiable. Néanmoins, on note qu'une fidélité « acceptable » correspond généralement à un coefficient de 0,7 (Nunnally et Bernstein, 1994). Lorsque le coefficient est supérieur à 0,8, la fidélité interne du test est considérée comme étant très bonne.

Le coefficient alpha de Cronbach se calcule en appliquant la formule suivante, où n est le nombre total d'items, σ_T^2 la variance de l'instrument dans son ensemble (du score total), et σ_i^2 la variance de l'item i :

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_T^2} \right]$$

Comme le notent Chartier et Loarer (2008), la plupart des tests qui sont aujourd'hui utilisées en France sont basés sur la théorie classique. Dans cette approche, les scores sont calculés en prenant la moyenne des résultats aux items, c'est-à-dire que le score ne va dépendre que du nombre d'items qui seront réussis. Or, si l'épreuve est difficile, peu d'items seront réussis et on dira du sujet qu'il est « faible ». Si l'épreuve est facile, on considèrera à l'inverse qu'il est « fort ». Cependant, il serait plus juste de reconnaître à un sujet qu'il est compétent s'il réussit mieux des items difficiles que des items faciles. La contribution de chaque item à l'estimation des compétences d'un sujet devrait être pondérée par un critère qui tiendrait compte de caractéristiques fixes de l'item (Bertrand et Blais, 2004) : c'est ce que proposent les MRI.

2.5 Les Modèles de Réponse à l'item : une nouvelle théorie de la mesure

Si l'on prend en considération l'indice de difficulté de l'item (donné par le nombre de sujets qui le réussissent), on constate que sa valeur dépend autant du niveau du groupe que de la difficulté propre de l'item. Il est alors difficile de connaître l'effet de l'un ou de l'autre. Il

en va de même pour l'indice de discrimination : si le groupe est homogène, la variance du test sera faible et on conclura que l'item ne discrimine pas vraiment. Si le groupe a des compétences hétérogènes, la corrélation sera plus élevée. Le coefficient alpha de Cronbach, calculé en fonction de la corrélation, sera lui aussi affecté. L'évaluation des caractéristiques d'un item dans le cadre de la théorie classique donne donc des résultats qui dépendent toujours de l'échantillon de population sur lequel l'item a été testé. Or, un item jugé facile ou difficile dans un groupe peut ne plus l'être dans un groupe différent. Les MRI évaluent les caractéristiques des items indépendamment du groupe. Ce sont des modèles de type probabiliste. Ils sont fondés sur le postulat que la réponse d'un sujet à un item peut s'expliquer par deux facteurs (Bertrand et Blais, 2004 ; Chartier et Loarer, 2008) : la compétence du sujet, que l'on qualifie de trait latent (appelé *trait θ* , non directement observable) et les caractéristiques propres des items (difficulté et discrimination).

Lorsque le sujet fournit une réponse à l'item, celle-ci dépend donc de ces deux facteurs. Les MRI sont donc des modèles qui établissent une relation entre la compétence du sujet (le trait latent θ), et la probabilité pour ce sujet de réussir l'item. Sur le plan mathématique, cette relation est formalisée par une fonction et représentée par une courbe : la Courbe Caractéristique d'Item. Le modèle le plus simple et le plus fréquemment utilisé est appelé modèle de Rasch, d'après le mathématicien danois Georg Rasch (Dickes et al., 1994 ; Laveault et Grégoire, 1997 ; Bertrand et Blais, 2004 ; Grégoire, 2005).

2.5.1 Une condition d'application des modèles : le nombre de sujets

Les propositions de modélisation mathématique pour l'analyse statistique impliquent des conditions qui balisent les applications pour que le modèle soit pertinent. Un échantillon comportant un nombre suffisant de sujets est une des premières conditions de l'utilisation de ces modèles statistiques. Dans le cas du modèle de Rasch, il a été établi qu'un échantillon minimal de 100 à 200 sujets est nécessaire afin que l'utilisation du modèle soit pertinente : une étude récente (Chen et al., 2014) montre que l'analyse selon le modèle de Rasch basée sur des échantillonnages « réduits » (<100 sujets) pourrait amener à des conclusions opposées à celles basées sur un échantillon plus large de sujets. Les auteurs concluent sur une incitation à l'extrême prudence lors de l'utilisation de ce modèle lorsque l'échantillonnage n'est pas suffisant.

Notre étude se base sur un ensemble de 85 sujets, ce qui constitue une des plus importantes cohortes européennes étudiées sur la question de la phonologie en contexte de bilinguisme. Néanmoins, ce nombre étant inférieur à 100, nous ne pourrons pas utiliser le modèle de Rasch pour l'analyse des items et la réduction de l'épreuve LITMUS-NWR. En conséquence, nous utiliserons le modèle de la Théorie Classique.

2.6 Objectifs de l'étude

Le présent mémoire s'inscrit dans le cadre de l'Action COST ISO804 «Language Impairments in a Multilingual Society : Linguistic Patterns and the Road to Assessment». En coordonnant la recherche dans de nombreux pays et à travers différentes communautés migrantes, cette action a pour but de décrire les habiletés linguistiques et cognitives des enfants bilingues avec TSL. Ainsi, elle s'intéresse aux difficultés que pose le plurilinguisme dans l'établissement d'un diagnostic de trouble du langage, et propose l'élaboration d'outils spécifiques à ce contexte.

Notre étude consiste à réduire l'épreuve expérimentale de répétition de non-mots élaborée par les linguistes de l'équipe 1 de l'Unité Inserm 930 de Tours, le test LITMUS-NWR. Elle fait suite au mémoire de Lelièvre (2013) poursuivant les recherches de validation du test entamées par Abi-Aad et Attalah (2012) sur une partie des items, dans le contexte libanais. Lelièvre (2013) avait noté la faible valeur discriminatoire de certains items et la longueur excessive du test lorsqu'il est associé à d'autres épreuves d'évaluation du langage.

Les principaux objectifs de l'étude sont :

- l'analyse des caractéristiques métriques des items de l'épreuve LITMUS-NWR et de leur degré de cohérence (analyse d'items),
- la réduction de l'épreuve, grâce à l'élimination des items les moins discriminants, à partir de la théorie classique du score vrai,
- l'examen du construit de l'épreuve en fonction de son but principal, c'est-à-dire permettre de différencier les enfants avec TSL des enfants au DT.

L'objectif final de cette étude est de proposer une épreuve réduite visant à alléger le temps de passation tout en gardant les capacités discriminatoires de l'épreuve LITMUS-NWR. Ceci permettra de contribuer à proposer un outil fiable d'évaluation phonologique en contexte de bilinguisme.

3 Matériel et Méthodes

3.1 Présentation de LITMUS-NWR

3.1.1 Pourquoi une nouvelle épreuve de répétition de non-mots ?

Comme nous l'avons vu, la tâche de répétition de non-mots a démontré son efficacité en tant qu'outil clinique pour aider à repérer un possible trouble du langage chez l'enfant. Cependant, nous avons également noté qu'il était difficile de poser un diagnostic de trouble du langage pour des enfants en situation de bilinguisme, à moins de pouvoir tester leurs compétences dans les deux langues.

C'est dans ce contexte que l'épreuve LITMUS-NWR propose une liste de non-mots dont l'élaboration avait pour objectif d'être utilisable dans un grand nombre de langues du monde. Ainsi, que l'enfant soit mono-, bi- ou plurilingue, et quelle que soit sa langue dominante, l'épreuve LITMUS-NWR a été créée pour pouvoir tester la complexité syllabique et ainsi discriminer les enfants avec TSL des enfants au développement typique.

3.1.2 Élaboration de l'épreuve

L'épreuve est originellement constituée de 71 non-mots. Les non-mots varient en longueur de syllabes, nombre de segments (phonèmes), complexité des structures syllabiques, ainsi que sur quelques points de complexité phonologique. La longueur a été contrôlée pour limiter l'interférence avec les capacités en mémoire de travail :

- la longueur des syllabes varie entre une et trois,
- la longueur des segments varie entre trois et sept,
- trois structures syllabiques sont présentes : CV, CCV (avec groupe consonantique complexe obstruante + liquide), CVC.

L'épreuve est composée de deux catégories d'items : 30 items qualifiés de « langue-indépendant » (LI) et 41 items qualifiés de « langue-dépendant » (LD)⁹. Pour les bilingues français langue seconde, les items LD vont donc tester quelques points de complexité phonologique du français. Les items LI ont été créés de façon à neutraliser la complexité segmentale, en se basant sur des phonèmes répertoriés dans la plupart des langues du monde.

Items LI (10 items contrôle, 20 items test)

- Voyelles : /i/, /u/, /a/ : situées aux extrémités du triangle vocalique.
- Consonnes : /p/, /k/, /f/, /l/ : différents lieux et modes d'articulation.
- Point de complexité : attaque branchante valide (ex : /fluka/). Selon Maddieson (2006), 88% d'un échantillon de 515 langues ont au moins une attaque branchante composée d'une obstruante et d'une liquide. Toutes ces langues possèdent au moins les structures CVC#¹⁰ et CCV.

-

Items LD dans la version française (3 items contrôle, 38 items test)

- Voyelles identiques à celles des items LI
- Consonnes identiques à celles de items LI + /s/
- Points de complexité :
 - présence du phonème /s/ au sein de groupes consonantiques ou en position de coda. Ferré et al. (2011) notent que l'on trouve en français des séquences de plusieurs consonnes comportant la fricative alvéolaire [s] (ex : strié, esprit, muscle),
 - présence du phonème /l/ en position de coda, notamment interne, par opposition aux items LI où /l/ est en position branchante : /pilfu/ (LD) vs /plifu/ (LI). Cette position se retrouve notamment en français (ex : calfeutrer).

⁹ Notons cependant que ces items seraient plus justement qualifiés de « quasi-indépendants » à la langue, car il est bien entendu difficile de s'abstraire totalement de celle-ci.

¹⁰ # : frontière de mot

Les items LD comportent donc globalement plus de points de complexité que les items LI. L'ordre des items a, dans un premier temps, été établi de façon aléatoire, puis légèrement corrigé pour éviter des suites d'items phonologiquement trop proches, ce qui aurait pu constituer un biais. Les items sont pré-enregistrés par une orthophoniste sur support digital.

Les scores obtenus par les sujets sont notés de trois façons :

- nombre d'erreurs,
- score par item : 1 pour un item réussi, 0 pour un item échoué
- % score : pourcentage réussi au sein de l'item

Pour l'analyse des items, nous utiliserons uniquement la notation dichotomique du score par item. Les autres types de scores ne sont pas utiles pour ce qui est du calcul des différents indices.

3.1.3 Structure des non-mots de l'épreuve

L'épreuve LITMUS-NWR testant la complexité syllabique, les résultats de l'analyse des items et le choix des items définitifs seront analysés en fonction de la complexité des structures de chaque item, présentées dans le tableau 1. Ceci nous permettra de comparer les items avec leur degré de complexité. Ce tableau a été réalisé avec les données de la littérature concernant la complexité syllabique (cf. section 2.1.2). Certains points (notamment celui de savoir si un mot contenant une attaque branchante est plus ou moins complexe qu'un mot avec coda) sont encore sujets à débat.

Tableau 1 : Liste des non-mots de LITMUS-NWR en fonction de leur complexité

	Structures de faible complexité		Marqueurs de complexité
☐	CCV	kla, fli, plu	1 groupe consonantique (GC) (<i>items contrôle</i>)
	CVC	kip, paf, fuk	1 coda (<i>items contrôle</i>)
	CVCV	faku, pilu, kapi	(<i>items contrôle</i>)
	CVCVC	pukif, kafip	1 coda
	LVCV	lafi	(<i>item contrôle</i>)
	CVCVCV	kifapu, pufaki	3 syll
☐	CVCVs	kifus, fapus	1 coda
	CVCVL	fapul, kufal	1 coda
	sCV	spu	1 GC (/s/+C) (<i>item contrôle</i>)
	CVs	kis	1 coda (<i>item contrôle</i>)
	CVL	fal	1 coda (<i>item contrôle</i>)
	Structures de complexité moyenne		Marqueurs de complexité
☐	CVCCV	paklu, fupli	5 segments, 1 GC (/l/ intervocalique)
	CCVCV	plifu, fluka	1 GC avec /l/ (attaque)
☐	CVCs	fips, piks	4 seg (segments), 1 coda, 1 GC (C + /s/)
	CVsC	pusk, kusp	4 seg, 1 coda, 1 GC (/s/ + C)
	CCVC	klaf, fluk	1 GC, 1 coda
	CCVL	plal, klil	1 GC, 1 coda
	CCVs	flis, klis	1 GC, 1 coda
	sCCV	skla, spli	1 /s/ + GC
	sCVC	skap, spaf	1 GC (/s/+C), 1 coda
	sCVCV	skafu, spiku	1 GC (/s/+C)
	CVLCV	pilfu, filpa	1 coda (/l/ interne)
	CVsCV	kuspa, fiska	5 seg, 1 GC (/s/+C interne), 1 coda
	Structures très complexes		Marqueurs de complexité
☐	CCVCVC	flukif, klifak	6 seg, 1 GC, 1 coda
	CCVCCV	flaplu, plaklu	6 seg, 2 GC
	CVCVCVC	kapufik, pifakup	3 syll, 7 seg, 1 coda
	CVCCVCV	kufłapi, pikłafu	3 syll, 7 seg, 1 GC (intervocalique syllabe 2)
	CVCVCCV	kupifła, fikupła	3 syll, 7 seg, 1 GC (intervocalique syllabe 3)
	CCVCVCV	flipuka, klipafu	3 syll, 7 seg, 1 GC (attaque)
☐	CCVCs	płiks, klups	5 seg, 2 GC (C+/l/ attaque et C+/s/ final), 1 coda
	CCVsC	klisp, plusk	5 seg, 2 GC, (C+/l/ attaque et /s/+C final) 1 coda
	sCVCVCV	skapufi, spakifu	3 syll, 7 seg, 1 GC (/s/+C)
	CVCVLCV	kufłapi, kupłafi	3 syll, 7 seg, 1 coda (/l/ interne)
	CVCVsCV	pafuski, fikuspa	3 syll, 7 seg, 1 coda (/s/ interne)
	CVCVCVs	kifapus, pifukas	3 syllabes, 7 segments, 1 coda (/s/ final)
	CVCVCVL	fikapul, pakifal	3 syllabes, 7 segments, 1 coda (/l/ final)

3.1.4 Le cas des items contrôle

L'épreuve LITMUS-NWR a été conçue au départ comme une épreuve de dépistage. Les items contrôle (cf. tableau 2) ont donc été intégrés au test pour deux raisons :

- 1) vérifier que les enfants sont capables de produire toutes les consonnes du test (pour éliminer d'éventuelles erreurs d'articulation),
- 2) vérifier que les enfants sont capables de produire les séquences de consonnes présentes dans l'épreuve, qui sont des séquences non harmonisées (sans relation d'identité : [pilu] vs [pipi]).

Ces items ne présentent aucune complexité phonologique. À partir de 5 ans, ils devraient être correctement réalisés pour un enfant sans trouble phonologique important.

Tableau 2 : Liste des items contrôle

	items contrôle
LI	faku
	fli
	fuk
	kapi
	kip
	kla
	lafi
	paf
	pilu
	plu
LD	fal
	kis
	spu

3.1.5 Procédure de passation de l'épreuve

L'image d'un petit extraterrestre que nous avons nommé « Zoubidou » est présentée à l'enfant sur un fichier Powerpoint. Nous indiquons à l'enfant qu'il vient d'une autre planète et parle une langue très étrange. Pour apprendre la langue de Zoubidou, l'enfant doit répéter les mots qu'il prononce. Les non-mots ainsi présentés se déclenchent à chaque changement de diapositive. L'enfant est équipé d'un casque audio relié à l'ordinateur, afin de maintenir sa concentration et de s'assurer de la bonne audition du stimulus sonore. Les productions de l'enfant sont enregistrées, afin d'une part de privilégier la relation avec l'enfant, et d'autre part de pouvoir transcrire et coder a posteriori de manière plus sûre.

3.1.6 Transcription et codage de l'épreuve

Un fichier Excel permet de reporter les productions de l'enfant. Dans le tableau, chaque non-mot à répéter est inscrit et défini par rapport à l'ordre dans lequel il apparaît dans l'épreuve, sa structure syllabique, son nombre de syllabes, de clusters (groupes consonantiques), de codas, la position du segment /s/ s'il y a lieu, son type (LI ou LD) et son nombre de segments. Les non-mots produits incorrectement sont transcrits en API¹¹. Le codage est réalisé par segment. Le type d'erreur est indiqué pour chaque segment erroné : substitution, métathèse, addition, élision et lexicalisation. L'erreur est codée en fonction de sa position dans le segment et dans la syllabe (attaque simple, attaque branchante, etc.), en fonction de la forme cible. Des exemples de codage sont proposés dans les tableaux 3 à 5.

¹¹ API : Alphabet Phonétique International

Tableau 3 : Exemple de codage pour une substitution (/klil/ → /plil/)

Cible	k	l	i	l
Production	p	l	i	l
Codage	S			

Tableau 4 : Exemple de codage pour une métathèse + addition (/kupalfi/ → /kuplafli/)

Cible	k	u	p		a	l	f		i
Production	k	u	p	l	a		f	l	i
Codage	M							A	

Le codage de la métathèse (« M ») se fait sur la position prosodique que le segment occupait dans la forme cible.

Tableau 5 : Exemple de codage pour une élision (/flaplu/ → /flapu/)

Cible	f	l	a	p	l	u
Production	f	l	a	p		u
Codage						E

3.2 Présentation de l'ensemble du protocole

Outre l'épreuve de répétition de non-mots, le protocole de l'étude comprend des épreuves expérimentales et des épreuves standardisées, évaluant les domaines verbal et non-verbal en français. L'ensemble des épreuves est regroupé dans le tableau 6.

Tableau 6 : Épreuves du protocole de l'étude pour l'évaluation en français

Épreuves verbales	
<i>Expérimentales</i>	<i>Standardisées</i>
Répétition de phrases (LITMUS-SR-French)	Lexique en production (N-EEL) (Chevrie-Muller et Plaza, 2001)
Tâche d'exhaustivité	Lexique en réception (N-EEL)
Langage spontané induit	Morphosyntaxe en production (N-EEL)
Répétition de non-mots (LITMUS-NWR)	Morphosyntaxe en réception (N-EEL)
	Répétition de mots (BILO) (Khomsi et al., 2007)
+ PabiQ - Questionnaire parental (COST, 2011)	
Épreuves non-verbales (fonctions exécutives)	
<i>Expérimentales</i>	<i>Standardisées</i>
Card Sorting (Iluz-Cohen, 2009)	Matrices de Raven (Raven, 1964)
La souris cachée	Empan de chiffres du WISC-IV (Weschler, 2003)
Monsieur Cacahuète	

Les enfants bilingues ont également été évalués dans leur langue première (anglais, arabe, portugais et turc), avec des épreuves de tests standardisés propres à chaque langue. Les résultats des enfants aux différentes épreuves permettent à l'ensemble de l'équipe impliquée dans le projet de déterminer les participants à l'étude, selon les critères d'inclusion définis en section 3.4.1.

3.3 Passations

Le déroulement des passations en français a eu lieu sur deux séances, réparties sur deux jours. Elles ont été réalisées dans les écoles des enfants, à leur domicile, ou à l'université dans une salle mise à la disposition de l'unité INSERM U930. Les passations ont, dans la mesure du possible, été réalisées par les mêmes examinateurs. Les passations dans la langue première des enfants ont suivi le même protocole. L'enfant avait à sa disposition une fiche comportant des cases correspondant au nombre d'épreuves du jour, avec des autocollants à collectionner à la fin de chaque épreuve. À la fin des séances, l'enfant pouvait choisir entre plusieurs cadeaux pour le récompenser de son travail (bulles, carnets, livres, set de papèterie, etc.). Toutes les épreuves impliquant une production verbale ont été enregistrées (audio), afin de faciliter la transcription, le codage et la cotation, et de privilégier la relation entre l'examineur et l'enfant.

3.4 Présentation de la population

Notre analyse des items permettant la réduction de l'épreuve LITMUS-NWR se fera à partir des résultats de 85 enfants âgés de 5;2 à 8;10 ans, appartenant à l'ensemble des 4 groupes testés (cf. tableau 7) :

- **Mo-DT** : monolingues au développement typique
- **Mo-TSL** : monolingues avec trouble spécifique du langage, suivis en orthophonie
- **Bi-DT** : bilingues au développement typique
- **Bi-TSL** : bilingues avec trouble spécifique du langage, suivis en orthophonie

Tableau 7 : Population de l'étude

	Bilingues		Monolingues	
	Bi-TSL	Bi-DT	Mo-TSL	Mo-DT
Nb total	19	39	15	12
L1 anglais	3	12		
L1 arabe	9	19		
L1 portugais	7	7		
L1 turc	0	1		
Total Bi = 58 Total Mo= 27				
Total TSL= 34 Total DT = 51				
Total général = 85				

Un tableau présentant l'ensemble de la population, incluant les âges et l'appartenance au groupe est consultable en annexe 1.

Si nous calculons la proportion d'enfants au DT et d'enfants TSL en fonction des âges, nous observons que la répartition n'est pas équivalente (cf. tableau 8).

Tableau 8 : Répartition DT/TSL en fonction de l'âge

	5 ans	6 ans	7 ans	8 ans
Nombre	23	31	18	13
Mo-DT	6	6	0	0
Mo-TSL	0	3	7	5
Total Mo	6	9	7	5
Bi-DT	13	16	8	2
Bi-TSL	4	6	3	6
Total Bi	17	22	11	8
Total DT	19	22	8	2
%DT	83%	71%	44%	15%
Total TSL	4	9	10	11
%TSL	17%	29%	56%	85%

3.4.1 Recrutement et critères d'inclusion

Les enfants ont été recrutés par le biais d'annonces et d'avis de recherche diffusés dans les écoles d'Indre et Loire et des départements voisins. Un entretien téléphonique avec les orthophonistes de la région a également eu lieu afin de recruter des enfants suivis en orthophonie pour trouble du langage oral. Les enfants Mo-TSL ont été recrutés par le biais du Centre Référence Langage de l'Hôpital de Clocheville à Tours.

Les critères d'inclusion concernant tous les enfants, quel que soit leur groupe, sont :

- l'âge : entre 5;2 ans et 8;7 ans,
- des résultats dans la norme aux Matrices de Raven,
- pas de problèmes d'audition signalés au questionnaire parental.

À ces critères s'ajoutent ceux concernant plus précisément les enfants bilingues :

- le bilinguisme précoce (simultané ou consécutif),

- une scolarisation à l'école française et une exposition à la langue première dans le contexte familial (au moins un parent).

Les enfants au DT ne doivent en outre pas avoir d'antécédents dans l'histoire médicale récente qui pourraient faire suspecter un trouble du langage. Les enfants avec TSL doivent être suivis en orthophonie pour trouble du langage oral.

4 Résultats

L'étape d'analyse des items nécessite une analyse des résultats au fur et à mesure que les calculs sont effectués. La réduction de l'épreuve s'effectue pas à pas, en fonction des résultats qui surviennent à chaque étape de calcul des indices. En raison de cette particularité, cette partie ne pourra pas constituer une simple présentation des résultats obtenus. En effet, ceux-ci devront être analysés et discutés à mesure, afin d'aboutir à une réduction tenant compte d'un maximum de paramètres révélés par les calculs statistiques et les impératifs cliniques qui sous-tendent l'élaboration d'une épreuve.

4.1 Calcul de l'indice p de difficulté¹²

Comme nous l'avons vu, l'indice de difficulté (indice p) est la probabilité qu'un item soit réussi. Dans le cas de l'épreuve LITMUS-NWR dans sa totalité, que nous appellerons dorénavant LITMUS-NWR 71, on s'attend à ce que :

- les indices de difficulté des items soient plus élevés (items plus faciles) pour le groupe des enfants au DT que pour le groupe des enfants avec TSL,
- les items intégrant un degré de complexité syllabique moindre présentent un indice de difficulté plus élevé que celui des items plus complexes sur l'ensemble des groupes,
- les items LD, présentant plus de points de complexité que les items LI, soient plus difficiles (indice p plus faible) que ces derniers.

Nous devrions également observer une différence entre les enfants au développement typique plus jeunes et les enfants plus âgés. Cependant, l'épreuve étant destinée à faire partie d'une batterie évaluant l'ensemble des habiletés linguistiques, l'âge des enfants a été choisi comme un compromis entre l'acquisition de la syntaxe et l'acquisition de la phonologie - le développement de la syntaxe se terminant plus tardivement que celui de la phonologie. Un test sur des enfants de 4 ans permettrait vraisemblablement d'obtenir des résultats qui complèteraient les données déjà obtenues sur cette épreuve de phonologie et ainsi de mieux

¹² Tous nos calculs ont été effectués à l'aide du logiciel Excel.

répondre à la question du développement phonologique. Or, les données dont nous disposons sont limitées à la tranche 5;10 ans-8;2. pour les enfants au DT.

Le calcul de l'indice de difficulté nous permettra également de pointer les items dont les indices varient peu entre le groupe d'enfants au DT et celui d'enfants avec TSL. Cela signifie que les items en question sont aussi faciles ou aussi difficiles pour le groupe TSL que pour le groupe DT. Ces items ne sont pas intéressants pour une épreuve dont le but est de détecter un éventuel trouble phonologique. C'est donc parmi eux que les premiers choix d'élimination vont être faits.

4.1.1 Nombre d'items en fonction de l'indice de difficulté

Nous avons d'abord calculé l'indice de difficulté p pour l'ensemble des sujets (ALL), incluant tous les groupes, bilingues et monolingues (Bi-DT, Bi-TSL, Mo-DT, Mo-TSL). Les résultats sont présentés en annexe 2. Puis nous avons classé ces indices en fonction du degré de difficulté, et déterminé le nombre d'items dans chaque catégorie (cf. tableau 9).

Tableau 9 : Nombre d'items en fonction de leur indice de difficulté

	très facile	facile	moyen	difficile	très difficile
	$p \geq 0,95$	$0,70 \leq p < 0,95$	$0,50 \leq p < 0,70$	$0,50 < p \leq 0,10$	$p < 0,10$
Nb ALL	6	45	18	2	0

Nous observons tout d'abord que l'épreuve LITMUS-NWR ne comporte que 2 items difficiles ([klisp] et [kupifla]) et aucun item très difficile pour l'âge considéré. Etant donné que le nombre des items très faciles et faciles est élevé (72%), nous choisissons de déterminer les candidats à l'élimination en ajoutant une variable : le calcul des indices de difficulté pour chaque groupe (DT et TSL) et la comparaison de ces indices pour chaque item.

4.1.2 Indice de difficulté : comparaison DT/TSL

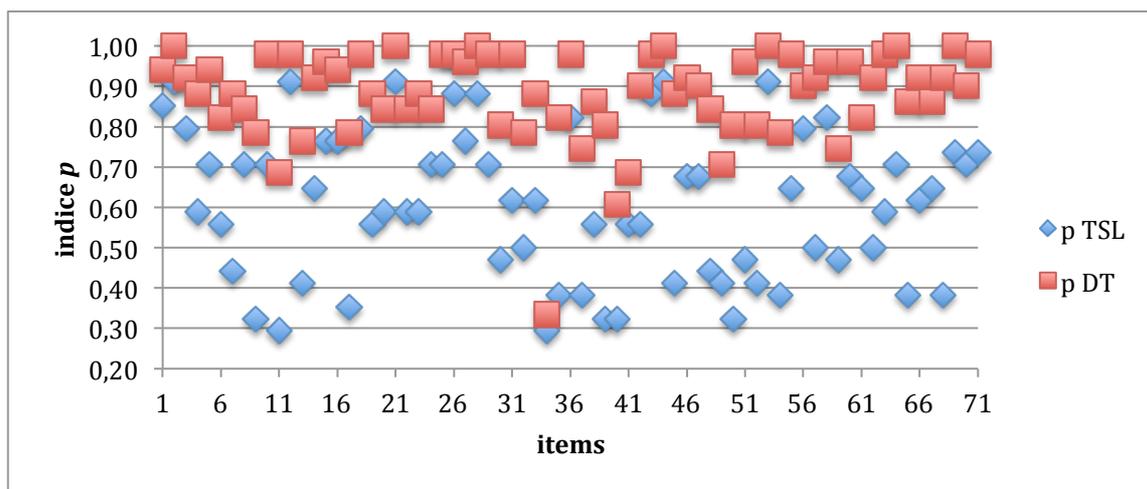
Lorsqu'on effectue le calcul de l'indice p en différenciant les groupes DT et TSL, on s'aperçoit que cet indice est globalement plus élevé pour le groupe DT (cf. tableau 10).

Tableau 10 : indice p pour les groupes DT et TSL

item	p DT	p TSL	item	p DT	p TSL	item	p DT	p TSL	item	p DT	p TSL
faku	0,94	0,85	fupli	0,88	0,56	kufalpi	0,75	0,38	plal	0,98	0,65
fal	1,00	0,91	kafip	0,84	0,59	kufilapi	0,86	0,56	plifu	0,90	0,79
fapul	0,92	0,79	kapi	1,00	0,91	kupalfi	0,80	0,32	pliks	0,92	0,50
fapus	0,88	0,59	kapufik	0,84	0,59	kupifla	0,61	0,32	plu	0,96	0,82
fikapul	0,94	0,71	kifapu	0,88	0,59	kusp	0,69	0,56	plusk	0,75	0,47
fikupla	0,82	0,56	kifapus	0,84	0,71	kuspa	0,90	0,56	pufaki	0,96	0,68
fikuspa	0,88	0,44	kifus	0,98	0,71	lafi	0,98	0,88	pukif	0,82	0,65
filpa	0,84	0,71	kip	0,98	0,88	paf	1,00	0,91	pusk	0,92	0,50
fips	0,78	0,32	kis	0,96	0,76	pafuski	0,88	0,41	skafu	0,98	0,59
fiska	0,98	0,71	kla	1,00	0,88	pakifal	0,92	0,68	skap	1,00	0,71
flaplu	0,69	0,29	klaf	0,98	0,71	paklu	0,90	0,68	skapufi	0,86	0,38
fli	0,98	0,91	klifak	0,80	0,47	pifakup	0,84	0,44	skla	0,92	0,62
flipuka	0,76	0,41	klil	0,98	0,62	pifukas	0,71	0,41	spaf	0,86	0,65
flis	0,92	0,65	klipafu	0,78	0,50	piklafu	0,80	0,32	spakifu	0,92	0,38
fluk	0,96	0,76	klis	0,88	0,62	piks	0,96	0,47	spiku	1,00	0,74
fluka	0,94	0,76	klisp	0,33	0,29	pilfu	0,80	0,41	spli	0,90	0,71
flukif	0,78	0,35	klups	0,82	0,38	pilu	1,00	0,91	spu	0,98	0,74
fuk	0,98	0,79	kufal	0,98	0,82	plaklu	0,78	0,38			

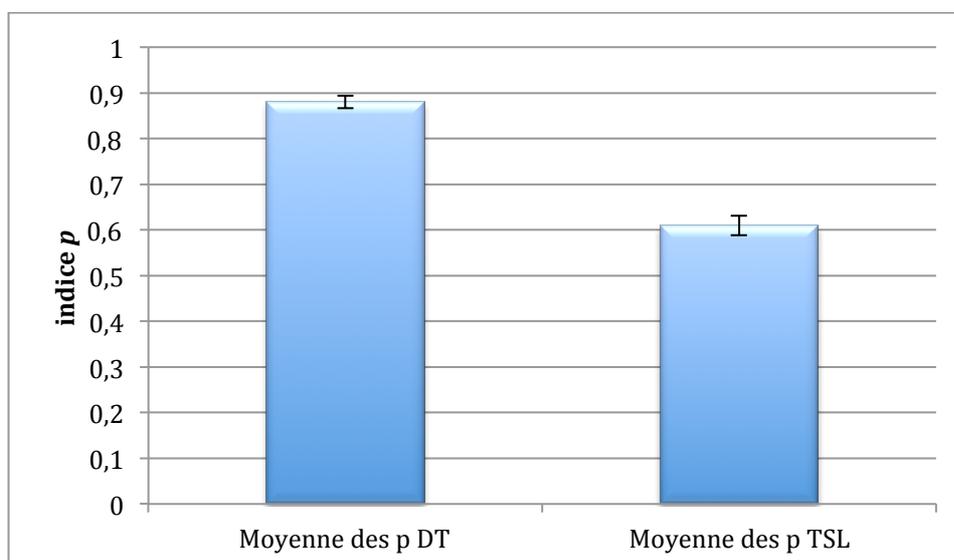
Le graphique 1 permet de visualiser cette différence : on remarque que les indices p des items pour le groupe DT sont tous (sauf l'item [klisp]) au-dessus de 0,60 (items de difficulté moyenne, faciles ou très faciles). Les indices p des items pour le groupe TSL sont répartis de façon plus hétérogène, mais un grand nombre d'entre eux ont des indices plus bas que ceux du groupe DT (items plus difficiles).

Graphique 1 : Répartition des indices p pour les groupes DT et TSL



Les moyennes des indices de difficulté ont été calculées (cf. graphique 2), et un test de Wilcoxon a été effectué, avec un résultat montrant que la différence entre les deux groupes est significative (p-value <0,05).

Graphique 2 : Comparaison des moyennes des indices p pour les groupes DT et TSL



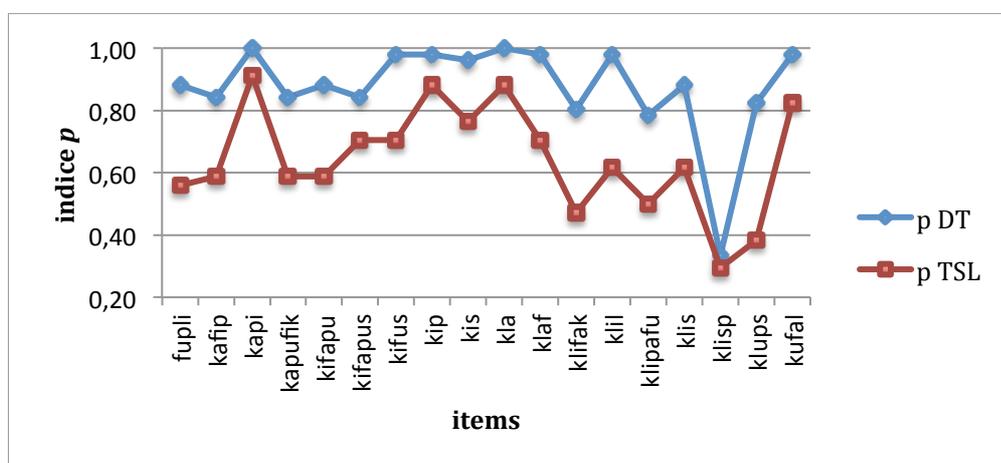
Si nous mettons en regard le nombre d'items en fonction de leur indice de difficulté pour chaque groupe (cf. tableau 11), nous remarquons cette fois-ci que 1,4% des items sont difficiles ou très difficiles pour le groupe DT (1 seul item : [klisp]), alors que ce calcul aboutit à 29,6% pour le groupe TSL. Seuls 4,2% des items sont de difficulté moyenne pour le groupe DT (32,4% pour le groupe TSL). Enfin, nous remarquons que 94,4% des items sont faciles ou très faciles pour le groupe DT (38% pour le groupe TSL).

Tableau 11 : Nombre d'items en fonction de leur indice de difficulté pour les groupes DT et TSL

	très facile		facile		moyen		difficile		très difficile	
	$p \geq 0,95$		$0,70 \leq p < 0,95$		$0,50 \leq p < 0,70$		$0,50 < p \leq 0,10$		$p < 0,10$	
Nb DT	24	33,8%	43	60,6%	3	4,2%	1	1,4%	0	0%
Nb TSL	0	0%	27	38%	23	32,4%	21	29,6%	0	0%
Nb ALL	6	8,5%	45	63,4%	18	25,3%	2	2,8%	0	0%

La comparaison entre les deux indices nous permet également de repérer les items dont l'indice de difficulté varie peu entre les deux groupes. Cela signifie que sur la base de cet indice, ces items discriminent peu. Sur le graphique 3, nous voyons par exemple que l'indice de difficulté de l'item [kapi] est élevé pour les deux groupes (item facile, réussi par la plupart des sujets), et que l'écart entre les indices pour chaque groupe est assez faible ($p=0,91$ pour le groupe TSL et $p=1$ pour le groupe DT, soit un écart de 0,09). De la même manière, l'indice de l'item [klisp] est bas pour les deux groupes et l'écart entre les deux indices est très faible (0,29 pour le groupe TSL et 0,33 pour le groupe DT, soit un écart de 0,04).

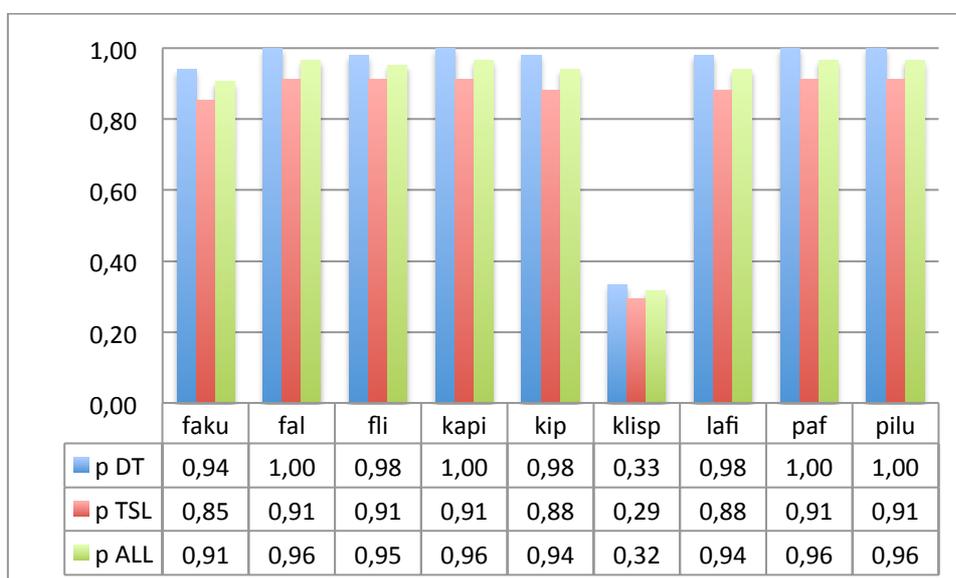
Graphique 3 : Exemple de comparaison de l'indice p DT/TSL sur 18 items



Il est donc possible d'isoler 9 items dont l'indice de difficulté varie peu ($\leq 0,1$) entre les groupes DT et TSL. Ce sont les items : [faku], [fal], [fli], [kapi], [kip], [klisp], [lafî], [paf] et [pilu] (cf. graphique 4).

Ces items seront à prendre en compte lors de la réduction de l'épreuve. Il s'agira de considérer attentivement leurs résultats aux autres types de calcul (indice de discrimination, écart-type, corrélation item-total) afin de déterminer lesquels d'entre eux seront sujets à élimination.

Graphique 4 : Items dont l'indice p varie peu selon les groupes (DT/TSL)



Nous observons que 8 items parmi ces 9 ont un certain nombre de caractéristiques communes :

- leur indice global de difficulté (p ALL) est assez élevé,
- ce sont des items contrôle (LD ou LI) : leur degré de complexité est faible.

L'item [klisp] est à part :

- son indice global de difficulté est bas,
- son degré de complexité est élevé.

Dans le cadre d'une épreuve utilisée en clinique, la présence des items contrôle est moins justifiable que dans un but de recherche. En effet, leur manque de complexité impacte

leur capacité discriminatoire, et risque donc d'avoir des conséquences sur celle de l'épreuve en totalité. Ces items sont de bons candidats à l'élimination.

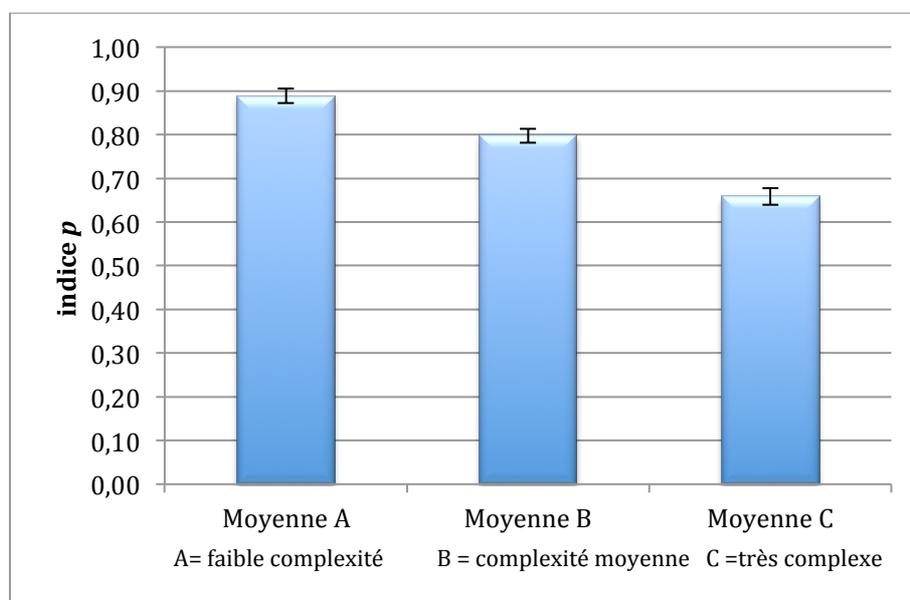
s

A ce stade, les items candidats à l'élimination sont :								
[faku]	[fal]	[fli]	[kapi]	[kip]	[klisp]	[laf]	[paf]	[pilu]

4.1.3 Indice de difficulté en fonction de la complexité

Nous avons ensuite mis en regard l'indice de difficulté p des items avec leur degré de complexité syllabique (cf. tableau 1 section 3.1.3). Les résultats sont présentés en annexe 3. Le graphique 5 présente les moyennes des indices p en fonction de la complexité. Nous observons que la difficulté se fait plus grande à mesure que la complexité s'accroît. Nous avons effectué un test de Kruskal-Wallis montrant que la différence est significative (p -value < 0,05).

Graphique 5 : Moyenne des indices p en fonction de la complexité



4.1.4 Indice de difficulté en fonction du type d'items : LI vs LD

Nous avons vu que l'épreuve comprenait deux catégories d'items : Langue-Indépendant et Langue-Dépendant (cf. section 3.1.2). Les items LD comportant plus de points de complexité que les items LI, ils devraient être moins bien réussis que ces derniers. Or, on observe dans le tableau 12 que la différence entre les moyennes est très légère. Un test de Mann-Whitney a été effectué, montrant que cette différence n'est pas significative ($p = 0,61$).

Tableau 12 : Moyenne des indices p selon le type d'items

	Moyenne des indices p
items LI	0,78
items LD	0,77

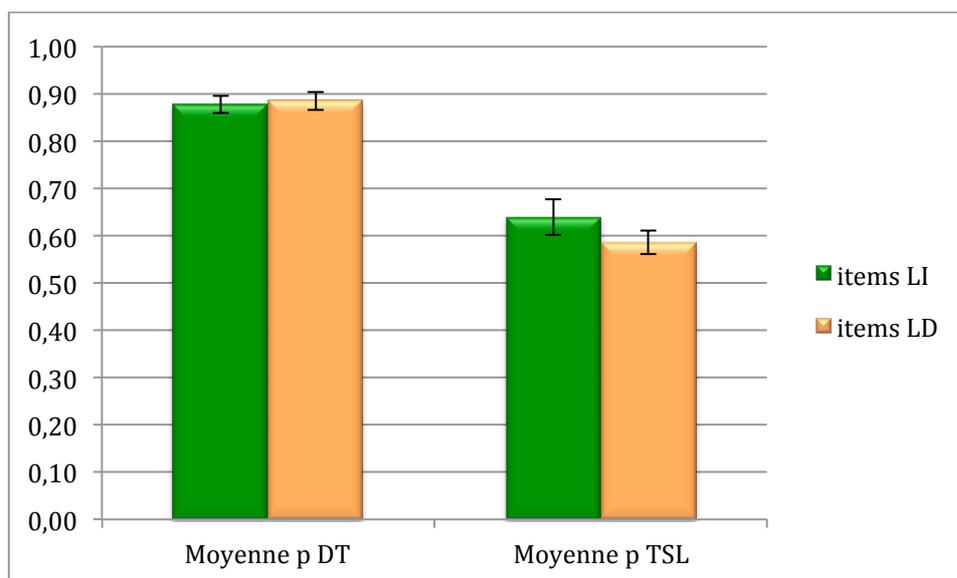
De la même manière, lorsque nous calculons la moyenne des indices de difficulté pour chacun des groupes DT et TSL (cf. tableau 13) en fonction du type d'items, nous nous apercevons qu'il y a très peu de différence entre les résultats aux items LI et LD dans chacun des groupes. Un test de Mann-Whitney a été effectué sur les moyennes pour chaque groupe, montrant effectivement que la différence n'est pas significative (p-value 'DT' = 0,58, p-value 'TSL' = 0,27).

Tableau 13 : Moyenne des indices p selon le type d'items par groupe DT/TSL

	Moyenne p DT	Moyenne p TSL
items LI	0,88	0,64
items LD	0,89	0,59

Le graphique 6 permet de se rendre compte que les items sont plus difficiles pour le groupe TSL, mais que dans chaque groupe, la différence entre les items LI et les items LD n'est pas importante.

Graphique 6: Comparaison des moyennes des indices p par groupe en fonction du type d'items



4.1.5 Items candidats à l'élimination par l'indice de difficulté

Jusqu'à présent, les calculs nous ont permis d'isoler 9 items ayant un indice de difficulté qui varie peu entre les groupes DT et TSL (cf. tableau 10). Cinq d'entre eux ([fal], [fli], [kapi], [paf] et [pilu]) font partie de la liste des 6 items « très faciles ».

Or, nous devons éliminer 21 items pour arriver à une épreuve comportant 50 items. Le tableau 14 présente les 22 items présentant les indices de difficulté les plus élevés, ainsi que l'item [klisp] qui présente un indice de difficulté plus bas, mais dont l'indice varie peu entre les groupes. Notons qu'il a fallu isoler 22 items pour en trouver 21, certains items présentant le même indice p .

Tableau 14 : Les 22 items les plus faciles selon l'indice p + « klisp »

Item	groupe	Indice difficulté p
fal	LD, contrôle	0,96
kapi	LI, contrôle	0,96
paf	LI, contrôle	0,96
pilu	LI, contrôle	0,96
fli	LI, contrôle	0,95
kla	LI, contrôle	0,95
kip	LI, contrôle	0,94
lafi	LI, contrôle	0,94
kufal	LD	0,92
plu	LI, contrôle	0,91
fuk	LI, contrôle	0,91
faku	LI, contrôle	0,91
spiku	LD	0,89
skap	LD	0,88
spu	LD, contrôle	0,88
kis	LD, contrôle	0,88
fluk	LD	0,88
fapul	LD	0,87
fiska	LD	0,87
kifus	LD	0,87
klaf	LD	0,87
fluka	LI	0,87
klisp	LD	0,32

 : items avec faible variation de p entre les groupes

À partir du tableau 14, un certain nombre de remarques peuvent être faites :

- Parmi les items les plus « faciles », nous retrouvons la totalité des items contrôle. Ces items ne présentant aucune complexité phonologique, ils sont globalement mieux réussis.
- On note la présence de 11 items LI et 11 items LD. A l'instar des résultats obtenus sur l'ensemble des items, le degré plus élevé de complexité des items LD ne se retrouve pas dans ce classement.
- Sur ces 22 items, aucun ne correspond à un item de structure très complexe (cf. tableau 1 section 3.1.3). 16 sont des items avec une structure de faible complexité, et 6 présentent une complexité moyenne.

Nous avons donc isolé 23 items candidats à l'élimination à partir du calcul de l'indice de difficulté. Cependant, selon la théorie classique du score vrai (cf. section 2.4), dans une analyse d'items, l'indice p à lui seul n'est pas suffisant pour pouvoir déterminer quels sont les items à éliminer. Nous allons donc procéder à un calcul de l'indice D de discrimination.

A ce stade, les items candidats à l'élimination sont :								
[faku]	[fal]	[fli]	[kapi]	[kip]	[klisp]	[laf]	[paf]	[pilu]
[kla]	[kufal]	[plu]	[fuk]	[spiku]	[skap]	[spu]	[kis]	[fluk]
[fapul]	[fiska]	[kifus]	[klaf]	[fluka]				

4.2 Calcul de l'indice D de discrimination

Pour rappel, l'indice de discrimination d'un item est la différence entre la proportion de réussite à cet item chez les sujets appartenant à un groupe « fort » (résultats élevés à l'ensemble de l'épreuve) et cette même proportion chez les sujets appartenant à un groupe « faible ». Nous avons donc déterminé les deux groupes d'après Kelley (1939, in Laveault et Grégoire 1997) comme indiqué en section 2.4.2. L'indice de discrimination pour l'ensemble des items est présenté en annexe 4. Nous présenterons ensuite les items candidats à l'élimination par l'indice D .

4.2.1 Détermination des groupes « fort » et « faible »

Le tableau 15 présente les groupes permettant le calcul de l'indice D . 23 enfants (27%) pour chaque groupe ont été choisis à partir de leur score total à l'ensemble des items (score/71).

Tableau 15 : Groupes « fort » et « faible »

Enfant	"Fort" (27%+)		Enfant	"Faible" (27%-)	
	Total réponses correctes	Groupe		Total réponses correctes	Groupe
FTM	66	Bi-DT	CLK	2	Mo-TSL
GED	66	Bi-DT	LOP	2	Mo-TSL
MLI	66	Mo-DT	LIB	3	Mo-TSL
SAC	66	Bi-DT	GAB	12	Bi-TSL
ANH	67	Bi-DT	AMO	16	Mo-TSL
ISL	67	Bi-DT	AUC	28	Mo-TSL
LDP	67	Bi-DT	CMU	30	Mo-TSL
MAD	67	Mo-DT	MAM	34	Mo-TSL
MLM	67	Mo-DT	ELV	35	Mo-TSL
ALB	68	Bi-DT	MAL	38	Mo-TSL
JOP	68	Bi-DT	AMT	40	Bi-TSL
KYP	68	Mo-DT	CEM	40	Mo-TSL
LEB	68	Mo-DT	DDS	41	Bi-TSL
LIT	68	Mo-DT	ELB	42	Bi-DT
VAH	68	Mo-DT	NBE	45	Mo-TSL
ESA	69	Bi-DT	GAV	46	Mo-TSL
KOS	69	Bi-DT	LEM	47	Bi-DT
LEG	69	Bi-DT	TVA	48	Bi-TSL
AMA	70	Bi-DT	VAM	48	Mo-TSL
CHI	70	Bi-DT	WYR	49	Mo-DT
MIE	70	Bi-DT	MAB	50	Bi-TSL
AMP	71	Bi-DT	TAS	50	Bi-TSL
ANB	71	Mo-DT	RAM	51	Bi-DT

Parmi les enfants du groupe « fort » :

- 14/23 (61 %) appartiennent au groupe des Bi-DT (36% de l'ensemble des Bi-TD),
- 8/23 (35%) appartiennent au groupe des Mo-DT (67% de l'ensemble des Mo-DT),
- 1/23 (4%) appartiennent au groupe des Bi-TSL (5% de l'ensemble des Bi-TSL),
- 0/23 (0%) appartiennent au groupe des Mo-TSL (0% de l'ensemble des Mo-TSL).

Parmi les enfants du groupe « faible » :

- 13/23 (57 %) appartiennent au groupe des Mo-TSL (87% des Mo-TSL),
- 6/23 (26%) appartiennent au groupe des Bi-TSL (32% de l'ensemble des Bi-TSL),
- 3/23 (13%) appartiennent au groupe des Bi-DT, (8% de l'ensemble des Bi-DT),
- 1/23 (4%) appartiennent au groupe des Mo-DT(8% de l'ensemble des Mo-DT).

4.2.2 Items candidats à l'élimination par l'indice D

À partir des indications de Ebel (1972) présentées en section 2.4.2, nous avons catégorisé les items selon leur degré de discrimination dans le tableau 16.

Tableau 16 : Classification des indices de discrimination (Ebel, 1972)

Type d'item	Catégorie	Indice D
Item qui discrimine très bien	TB	0,4 et plus
Item qui discrimine bien	B	0,3 à 0,39
Item qui discrimine peu	P	0,2 à 0,29
Item limite, à améliorer	L	0,1 à 0,19
Item sans utilité réelle pour l'examen	N	Moins de 0,1

Le tableau 17 présente un classement des items selon d'indice D. 13 items appartiennent aux trois dernières catégories (P, L et N) qui correspondent aux indices de discrimination les plus faibles. Ils sont candidats à l'élimination. Nous observons également que l'épreuve comporte une grande majorité d'items (82%) qui discriminent bien ou très bien.

Tableau 17 : Classement des items selon d'indice D

Item	Indice D	Item	Indice D	Item	Indice D	Item	Indice D
kapi	N	fapul	B	fapus	TB	kusp	TB
pilu	N	fikupla	B	fikapul	TB	kuspa	TB
fal	L	fiska	B	fikuspa	TB	pafuski	TB
fli	L	flis	B	filpa	TB	paklu	TB
kip	L	fluk	B	fips	TB	pifakup	TB
kla	L	fluka	B	flaplu	TB	pifukas	TB
lafi	L	fuk	B	flipuka	TB	piklafu	TB
paf	L	kifapus	B	flukif	TB	piks	TB
faku	P	kifus	B	fupli	TB	pilfu	TB
kis	P	klaf	B	kafip	TB	plaklu	TB
klisp	P	klis	B	kapufik	TB	plal	TB
kufal	P	pakifal	B	kifapu	TB	pliks	TB
plu	P	plifu	B	klifak	TB	plusk	TB
		pukif	B	klil	TB	pufaki	TB
		skap	B	klipafu	TB	pusk	TB
		spiku	B	klups	TB	skafu	TB
		spu	B	kufalpi	TB	skapufi	TB
				kufalpi	TB	skla	TB
				kupalfi	TB	spaf	TB
				kupifla	TB	spakifu	TB
						spli	TB

4.2.3 Comparaison indice D et indice p

Nous allons maintenant mettre en regard l'indice D de ces 13 items avec leur indice p . Le tableau 18 présente ces résultats.

Tableau 18 : Indice p des items les moins discriminants

Item	Discrimination	Indice p
kapi	N	0,96
pilu	N	0,96
fal	L	0,96
fli	L	0,95
kip	L	0,94
kla	L	0,95
lafi	L	0,94
paf	L	0,96
faku	P	0,91
kis	P	0,88
klisp	P	0,32
kufal	P	0,92
plu	P	0,91

Parmi ces 13 items, nous retrouvons les 9 items dont l'écart de valeur de p pour les groupes DT et TSL est réduit. D'autre part, 12 items parmi les 13 font partie de la liste des 22 items ayant un indice de difficulté p parmi les plus élevés. L'item [klisp] se comporte différemment. Il est peu discriminant, mais son indice de difficulté est bas. En fonction des autres calculs, ces items pourraient être éliminés lors de la réduction.

Les 10 items restants, qui présentent un indice de difficulté élevé, discriminent bien. Aucun d'entre eux ne discrimine très bien. Ils restent candidats à l'élimination.

4.2.4 Lien entre l'indice p et l'indice D

Nous avons calculé le coefficient de corrélation entre l'indice de difficulté p et l'indice de discrimination D :

$$r = -0,68$$

Les deux indices sont fortement corrélés négativement : lorsque p augmente, D diminue. Autrement dit, plus l'item est facile, moins il a de probabilités d'être discriminant. Le degré de difficulté d'un item est donc ici bien lié à sa capacité à discriminer.

A ce stade, les items candidats à l'élimination sont :								
[faku]	[fal]	[fli]	[kapi]	[kip]	[klisp]	[laf]	[paf]	[pilu]
[kla]	[kufal]	[plu]	[fuk]	[spiku]	[skap]	[spu]	[kis]	[fluk]
[fapul]	[fiska]	[kifus]	[klaf]	[fluka]				

Nous observons que le calcul de l'indice de discrimination a confirmé la présence de certains items dans la liste des items à éliminer (cellules gris foncé), mais n'a permis d'en ajouter aucun.

4.3 Calcul de l'écart-type

4.3.1 Items candidats à l'élimination par l'écart-type (ET)

Nous avons vu en section 2.4.3 que l'examen de l'écart-type était indissociable de celui de la moyenne (qui correspond elle-même à l'indice p pour des items dichotomiques). Le résultat des calculs de moyenne et écart-type pour chaque item sont présentés en annexe 5. Les items présentant des moyennes extrêmes et des écarts-types faibles ($<0,30$) sont présentés dans le tableau 19.

Tableau 19 : Items avec moyenne extrême et écart-type faible

item	Moyenne	Ecart-type
faku	0,91	0,29
fal	0,96	0,19
fli	0,95	0,21
fuk	0,91	0,29
kapi	0,96	0,19
kip	0,94	0,24
kla	0,95	0,21
kufal	0,92	0,28
lafi	0,94	0,24
paf	0,96	0,19
pilu	0,96	0,19
plu	0,91	0,29

Nous retrouvons ici 11 items contrôle, faisant partie des 13 items les moins discriminants et les plus faciles. Seul [kufal] n'est pas un item contrôle. Les valeurs de trois indices de ces 12 items (p , D et ET) fait d'eux d'excellents candidats à l'élimination. Calculons à présent l'écart-type des items candidats à l'élimination par les indices p et D qui ne font pas partie de cette liste.

4.3.2 Écart-type des items avec indice p trop élevé ou trop bas et indice D faible

Deux items présentent simultanément des indices p et D qui pourraient permettre leur élimination. Leur écart-type est calculé dans le tableau 20.

Tableau 20 : Écart-type des items avec p et D insuffisants

Item	Difficulté p	Discrimination D	Ecart-type
kis	0,88	0,26	0,32
klisp	0,32	0,22	0,47

: item avec faible variation de p entre les groupes

Nous observons que l'écart-type de l'item [kis] est légèrement supérieur à 0,30, mais que sa variation de p entre les groupes est suffisante. De son côté, l'item [klisp] possède un bon écart-type, mais une faible variation de p entre les groupes. Ces deux items restent donc candidats à l'élimination, mais pour des raisons différentes.

4.3.3 Écart-type des items avec indice p trop élevé et indice D correct

Le tableau 21 permet d'observer que la valeur des écarts-types de tous ces items se situe certes au-delà de 0,30, mais ne dépasse pas 0,40. Ils restent donc candidats à l'élimination.

Tableau 21 : Écart-type des items avec p trop élevé et D correct

Item	Difficulté p	Ecart-type
fapul	0,87	0,34
fiska	0,87	0,34
fluk	0,88	0,32
fluka	0,87	0,34
kifus	0,87	0,34
klaf	0,87	0,34
skap	0,88	0,32
spiku	0,89	0,31
spu	0,88	0,32

A ce stade, les items candidats à l'élimination sont :								
[faku]	[fal]	[fli]	[kapi]	[kip]	[klisp]	[lafī]	[paf]	[pilu]
[kla]	[kufal]	[plu]	[fuk]	[spiku]	[skap]	[spu]	[kis]	[fluk]
[fapul]	[fiska]	[kifus]	[klaf]	[fluka]				

4.4 Calcul de la corrélation r item/total

La corrélation item/total mesure la force de la liaison entre la performance des sujets à un item et celle que l'on observe en prenant en compte le score à l'ensemble des items. Les résultats sont présentés en annexe 6. Les coefficients de corrélation r de la plupart des items sont corrects. Seuls deux items ne présentent pas un coefficient de corrélation suffisant ($<0,30$) :

- [pilu] ($r = 0,17$)
- [klisp] ($r = 0,21$)

L'item [pilu] est un item contrôle. Nous avons déjà remarqué que cet item avait un indice de difficulté trop élevé et un indice de discrimination trop faible. L'item [klisp] est quant à lui trop difficile et son indice de discrimination est trop faible, même s'il présente un bon écart-type. Ces deux items sont donc de bons candidats à l'élimination.

A ce stade, les items candidats à l'élimination sont :								
[faku]	[fal]	[fli]	[kapi]	[kip]	[klisp]	[lafī]	[paf]	[pilu]
[kla]	[kufal]	[plu]	[fuk]	[spiku]	[skap]	[spu]	[kis]	[fluk]
[fapul]	[fiska]	[kifus]	[klaf]	[fluka]				

4.5 Premier essai de réduction de l'épreuve et premières interrogations

À la lumière des calculs précédents, nous pouvons donc procéder à la réduction de l'épreuve LITMUS-NWR 71. Le tableau 22 présente l'intégralité des items qui devraient éliminés, accompagnés de leurs indices respectifs.

Tableau 22 : Items candidats à l'élimination

Item	Type	Indice p	Indice D	Ecart-type	r item/total
faku	LI contrôle	0,91	0,22	0,29	0,58
fal	LD contrôle	0,96	0,13	0,19	0,56
fapul	LD	0,87	0,30	0,34	0,64
fiska	LD	0,87	0,39	0,34	0,69
fli	LI contrôle	0,95	0,13	0,21	0,55
fluk	LD	0,88	0,30	0,32	0,59
fluka	LI	0,87	0,39	0,34	0,64
fuk	LI contrôle	0,91	0,35	0,29	0,67
kapi	LI contrôle	0,96	0,09	0,19	0,38
kifus	LD	0,87	0,39	0,34	0,62
kip	LI contrôle	0,94	0,17	0,24	0,55
kis	LD contrôle	0,88	0,26	0,32	0,48
kla	LI contrôle	0,95	0,17	0,21	0,71
klaf	LD	0,87	0,39	0,34	0,65
klisp	LD	0,32	0,22	0,47	0,17
kufal	LD	0,92	0,22	0,28	0,65
lafi	LI contrôle	0,94	0,17	0,24	0,60
paf	LI contrôle	0,96	0,13	0,19	0,58
pilu	LI contrôle	0,96	0,04	0,19	0,21
plu	LI contrôle	0,91	0,26	0,29	0,41
skap	LD	0,88	0,43	0,32	0,77
spiku	LD	0,89	0,39	0,31	0,81
spu	LD contrôle	0,88	0,35	0,32	0,58

23 items sont donc candidats à l'élimination :

- tous les items contrôle,
- 11 items LI et 9 items LD.

Nous avons vu que la réduction devrait aboutir à un total de 40 à 50 items. Les items candidats à l'élimination étant au nombre de 23, nous choisissons à ce stade de réduire l'épreuve à 50 items, d'une part pour ne pas prendre le risque de retirer des items pertinents

pour l'évaluation, et d'autre part pour suivre les recommandations préconisées dans la littérature (cf. section 2.3.2). Pour ce faire, il faudrait éliminer 21 items. Nous devrions donc en garder 2. Le choix de ces 2 items devrait se faire parmi les items ayant au moins un indice correct :

- [skap], qui discrimine très bien,
- [fiska], [fluka], [kifus], [klaf], et [spiku] qui discriminent bien.

En premier lieu, nous choisissons de garder l'item [skap], dont l'indice de discrimination est très bon. Il reste à choisir 1 item parmi les 5 restants. Si l'épreuve réduite était destinée à la recherche, nous choisirions l'item [klaf], qui reste le seul représentant de la structure CCVC. Cependant, l'épreuve devra être utilisée en clinique, pour tester des enfants qui pourraient avoir un trouble du langage. Nous avons vu qu'il était important de garder des structures de faible complexité dans l'épreuve, pour que la motivation des enfants soit à son maximum. L'item [spiku], avec un indice de difficulté de 0,89, est mieux réussi que les autres selon les statistiques. Cependant, dans le tableau 1, nous voyons que sa structure (sCVCV) est de complexité moyenne. L'item [kifus] est légèrement moins bien réussi que [spiku], mais sa structure est de faible complexité (CVCVs).

Ici, nous pensons que les impératifs cliniques doivent l'emporter sur la statistique. Il ne reste plus que 5 items de faible complexité dans l'épreuve réduite : [fapus], [kafip], [kifapu], [pufaki] et [pufik]. Ils comportent 2 ou 3 syllabes. Il n'y a plus d'item très facile d'une seule syllabe dans l'épreuve. Aussi choisissons-nous de garder un item monosyllabique, éventuellement à placer en tout premier dans l'épreuve, pour que les enfants ne se sentent pas en difficulté dès le départ. L'item [kip] était proposé le premier dans les passations de l'épreuve LITMUS-NWR 71. Lors des passations que nous avons effectuées, nous avons remarqué que les enfants trouvaient amusant de le répéter après le petit extra-terrestre « Zoubidou ». Il est important pour le déroulement de l'épreuve dans son entier que le début soit très facile et motivant. Pour ces raisons, nous choisissons de le garder pour l'épreuve réduite.

Un premier choix des items à éliminer peut donc être fait. Le tableau 23 nous récapitule les items composant le premier essai de réduction de l'épreuve, que nous nommerons LITMUS-NWR 50-β.

Tableau 23 : LITMUS-NWR 50-β : premier essai de réduction de l'épreuve à 50 items

Item	Structure	Type	Item	Structure	Type
fapus	CVCVS	LD	kusp	CVSC	LD
fikapul	CVCVCL	LD	kuspa	CVSCV	LD
fikupla	CVCVCCV	LI	pafuski	CVCVSCV	LD
fikuspa	CVCVSCV	LD	pakifal	CVCVCL	LD
filpa	CVLCV	LD	paklu	CVCCV	LI
fips	CVCS	LD	pifakup	CVCVCVC	LI
flaplu	CCVCCV	LI	pifukas	CVCVCVS	LD
flipuka	CCVCVCL	LI	piklafu	CVCCVCL	LI
flis	CCVS	LD	piks	CVCS	LD
flukif	CCVCVC	LI	piifu	CVLCV	LD
fupli	CVCCV	LI	plaku	CCVCCV	LI
kafip	CVCVC	LI	plal	CCVL	LD
kapufik	CVCVCVC	LI	plifu	CCVCV	LI
kifapu	CVCVCV	LI	pliks	CCVCS	LD
kifapus	CVCVCVS	LD	plusk	CCVSC	LD
kip	CVC	LI	pufaki	CVCVCV	LI
klifak	CCVCVC	LI	pukif	CVCVC	LI
klil	CCVL	LD	pusk	CVSC	LD
klipafu	CCVCVCL	LI	skafu	SCVCV	LD
klis	CCVS	LD	skap	SCVC	LD
klups	CCVCS	LD	skapufi	SCVCVCV	LD
kufalpi	CVCVLCV	LD	skla	SCCV	LD
kuflapi	CVCCVCL	LI	spaf	SCVC	LD
kupalfi	CVCVLCV	LD	spakifu	SCVCVCV	LD
kupifla	CVCVCCV	LI	spli	SCCV	LD

Si nous effectuons maintenant une comparaison des deux épreuves en fonction du type d'items et de la complexité, nous obtenons les résultats indiqués dans le tableau 24.

Tableau 24 : Comparaison des épreuves en terme de type et de complexité des items

	LITMUS-NWR 71	LITMUS-NWR 50-β	
Type	30 items LI (42,3%)	20 items LI (40%)	
	41 items LD (57,7%)	30 items LD (60%)	
Complexité	A=21 (29,6%)	A=6 (12%)	A = faible complexité
	B=24 (33,8%)	B=18 (36%)	B = complexité moyenne
	C=26 (36,6%)	C=26 (52%)	C = très complexe

Comme nous l'avions remarqué, il ne reste plus que 7 items avec une structure de faible complexité, alors qu'un seul item très complexe a été retiré ([klisp]). En conséquence, le pourcentage des items de complexité moyenne et surtout celui des items très complexes s'en trouve fortement augmenté.

Nous avons vu (cf. section 2.4.5) que dans l'approche proposée par la théorie classique du score vrai, les scores ne dépendaient que du nombre d'items réussis, ce qui constitue une des limites de la théorie. Si l'épreuve est trop difficile, peu d'items seront réussis par les sujets, qui seront considérés comme étant « faibles », de manière un peu artificielle. L'épreuve comportant 71 items avait un bon niveau de difficulté : le niveau de complexité des items était bien réparti. L'épreuve réduite LITMUS-NWR 50- β , en privilégiant les items les plus « performants », a également affecté cette répartition. Or, dans l'étape 3 de la construction d'un test (cf. section 2.3.1) selon Laveault et Grégoire (1997), il s'agit entre autres de trouver le bon niveau de difficulté du test. Notre analyse des items a jusqu'ici négligé ce facteur.

La réduction proposée dans l'épreuve LITMUS-NWR 50- β ne nous semble donc pas satisfaisante pour une utilisation clinique. Nous devons tenter de mieux répartir le degré de difficulté des items pour deux raisons principales, que nous résumerions de la manière suivante :

- LITMUS-NWR 71 est une épreuve de dépistage de trouble phonologique basé sur la complexité syllabique. Si elle ne possède pas assez d'items monosyllabiques de faible complexité, elle ne permettra pas de faire une différence entre les sujets chutant sur des items très peu complexes (donc avec a priori un trouble important) et les sujets ayant des difficultés sur les items plus complexes. L'épreuve sera certes discriminante – ce qui est principalement demandé à une épreuve de dépistage, mais perdra en possibilité d'analyse clinique plus fine.
- LITMUS-NWR 71 est une épreuve destinée aux enfants à partir de 4 ans. Pour garder intacte l'envie de l'enfant, amorcée par le fort capital sympathie du petit extra-terrestre Zoubidou qui présente les non-mots (cf. section 3.1.5), il faut parsemer l'épreuve de quelques items courts et faciles. L'enfant doit pouvoir être remotivé après des séquences plus difficiles, d'autant que les items plus longs se

ressemblent pour certains d'entre eux (par exemple : [kuɫɫapi], [kuɫɫɔpi], [kuɫɫɔfi], [kuɫɫɔfi]). Couper le rythme dans la passation permettra également de minimiser l'impression de « déjà-dit ».

4.6 *Second essai de réduction de l'épreuve*

Pour cet essai de réduction prenant en compte les facteurs susmentionnés, il faudra tout de même faire attention à ne pas ajouter trop d'items très faciles. En effet, cela pourrait fortement faire baisser la capacité discriminatoire de l'épreuve. Nous proposons de répartir les items comme suit :

- items de faible complexité (A) : 10 (au lieu de 6 dans LITMUS-NWR 50-β)
- items de complexité moyenne (B) : 18 (les mêmes que dans LITMUS-NWR 50-β)
- items très complexes (C) : 22 (au lieu de 26 dans LITMUS-NWR 50-β)

À partir de la sélection déjà effectuée en section 4.1.5 (cf. tableau 14) il s'agit donc de retirer 4 items à la série des items très complexes, et d'en ajouter 4 à celle des items de faible complexité. Nous avons vu par ailleurs que les items qui manquaient étaient les items monosyllabiques – raison pour laquelle nous avons choisi de garder [kip]. Tous les calculs d'indices ayant été déjà effectués, nous les présentons dans leur ensemble pour ces items dans le tableau 25.

Tableau 25 : Indices des items de faible complexité monosyllabiques et des items très complexes

Item	Structure	Groupe	Complexité	Indice p	Indice D	Ecart-type	r item/total	p (DT-TSL)
faku	CVCV	LI	A	0,91	0,22	0,29	0,48	0,09
fal	CVL	LD	A	0,96	0,13	0,19	0,64	0,09
fli	CCV	LI	A	0,95	0,13	0,21	0,55	0,07
fuk	CVC	LI	A	0,91	0,35	0,29	0,67	0,19
kis	CVs	LD	A	0,88	0,26	0,32	0,48	0,20
kla	CCV	LI	A	0,95	0,17	0,21	0,71	0,12
paf	CVC	LI	A	0,96	0,13	0,19	0,58	0,09
plu	CCV	LI	A	0,91	0,26	0,29	0,41	0,14
spu	sCV	LD	A	0,88	0,35	0,32	0,58	0,24
fikapul	CVCVCL	LD	C	0,85	0,48	0,36	0,70	0,23
fikupla	CVCVCCV	LI	C	0,72	0,39	0,45	0,43	0,26
fikuspa	CVCVsCV	LD	C	0,71	0,74	0,46	0,62	0,44
flaplu	CCVCCV	LI	C	0,53	0,61	0,50	0,44	0,40
flipuka	CCVCVCV	LI	C	0,62	0,74	0,49	0,49	0,35
flukif	CCVCVC	LI	C	0,61	0,70	0,49	0,50	0,43
kapufik	CVCVCVC	LI	C	0,74	0,57	0,44	0,56	0,25
kifapus	CVCVCVs	LD	C	0,79	0,35	0,41	0,44	0,13
klifak	CCVCVC	LI	C	0,67	0,57	0,47	0,52	0,33
klipafu	CCVCVCV	LI	C	0,67	0,61	0,47	0,49	0,28
klups	CCVCs	LD	C	0,65	0,78	0,48	0,59	0,44
kufalpi	CVCVLCV	LD	C	0,60	0,83	0,49	0,58	0,37
kufłapi	CVCCVCV	LI	C	0,74	0,52	0,44	0,56	0,30
kupalfi	CVCVLCV	LD	C	0,61	0,83	0,49	0,64	0,48
kupifla	CVCVCCV	LI	C	0,49	0,61	0,50	0,43	0,29
pafuski	CVCVsCV	LD	C	0,69	0,83	0,46	0,68	0,47
pakifal	CVCVCL	LD	C	0,82	0,39	0,38	0,56	0,24
pifakup	CVCVCVC	LI	C	0,68	0,78	0,47	0,63	0,40
pifukas	CVCVCVs	LD	C	0,59	0,48	0,50	0,46	0,30
piklafu	CVCCVCV	LI	C	0,61	0,74	0,49	0,58	0,48
plaklu	CCVCCV	LI	C	0,62	0,48	0,49	0,39	0,40
pliks	CCVCs	LD	C	0,75	0,52	0,43	0,54	0,42
plusk	CCVsC	LD	C	0,64	0,57	0,48	0,48	0,28
skapufi	sCVCVCV	LD	C	0,67	0,78	0,47	0,62	0,48
spakifu	sCVCVCV	LD	C	0,71	0,87	0,46	0,69	0,54

x = mesure moins performante parmi les items très complexes

4.6.1 Choix des items très complexes à retirer

Parmi les 25 items de structure très complexe, le tableau 25 montre, surlignés en jaune, les indices les moins performants de certains d'entre eux. Cela permet d'isoler 8 items, que nous comparons avec ceux du même type de complexité. Nous choisissons de ne garder

que les items les plus difficiles et les plus discriminants parmi cette liste. Cela permettra de compenser le fait que nous allons ajouter par ailleurs des items plus faciles, et de n'entamer que faiblement la capacité discriminatoire de l'épreuve dans sa globalité. Le tableau 26 indique les indices les moins performants pour ces items.

Tableau 26 : Indices moins performants des items très complexes à éliminer

	p	ET	p (DT-TSL)	D
fikapul	X	X	X	
fikupla			X	X
kapufik			X	
kifapus			X	X
klipafu			X	
kupifla			X	
pakifal	X	X	X	X
plusk			X	

Nous choisissons d'éliminer les items qui possèdent le plus grand nombre d'indices les moins performants. L'item [pakifal] en possède 4. L'item [fikapul] en possède 3. Les items [fikupla] et [kifapus] en possèdent 2. Ce sont donc les 4 items qui seront éliminés.

4.6.2 Choix des items de faible complexité à ajouter

Le choix des items de faible complexité à ajouter à l'épreuve est moins aisé. Doit-on remettre les items les plus faciles (et/ou discriminants, etc.), ou choisir les plus difficiles ? Rappelons que notre choix de réinsérer des items faciles était guidé à la fois par l'idée de garder la motivation des enfants, et par celle de permettre à l'épreuve de détecter les enfants avec un trouble plus important. C'est à la lumière de cette dernière remarque que nous avons décidé de réinsérer les items faciles qui sont les plus échoués par le groupe des enfants avec TSL.

Le tableau 27 indique, surlignés en jaune, les indices de difficulté les moins élevés pour le groupe avec TSL sur ces items :

Tableau 27 : Indice p pour les items monosyllabiques (sauf kip)

Item	Structure	Groupe	Complexité	Indice p TSL
faku	CVCV	LI	A	0,85
fal	CVL	LD	A	0,91
fli	CCV	LI	A	0,91
fuk	CVC	LI	A	0,79
kis	CVs	LD	A	0,76
kla	CCV	LI	A	0,88
paf	CVC	LI	A	0,91
plu	CCV	LI	A	0,82
spu	sCV	LD	A	0,74

x = valeur de l'indice moins élevée

Les 4 items les moins bien réussis parmi cette liste par les enfants avec TSL sont : [fuk], [kis], [plu] et [spu]. Ils seront réintégrés dans l'épreuve.

4.7 Épreuve réduite LITMUS-NWR 50

Le tableau 28 présente donc les items présents dans l'épreuve réduite finale, que nous appellerons LITMUS-NWR 50. L'ensemble des résultats des calculs d'indices pour ces items est présenté en annexe 7.

Tableau 28 : LITMUS-NWR 50 : réduction de l'épreuve à 50 items

Item	Structure	Type	Item	Structure	Type
fapus	CVCVS	LD	kuspa	CVSCV	LD
fikuspa	CVCVSCV	LD	pafuski	CVCVSCV	LD
filpa	CVLCV	LD	paklu	CVCCV	LI
fips	CVCS	LD	pifakup	CVCVCVC	LI
flaplu	CCVCCV	LI	pifukas	CVCVCVS	LD
flipuka	CCVCVCV	LI	piklafu	CVCCVCV	LI
flis	CCVS	LD	piks	CVCS	LD
flukif	CCVCVC	LI	pilfu	CVLCV	LD
fuk	CVC	LI	plaklu	CCVCCV	LI
fupli	CVCCV	LI	plal	CCVL	LD
kafip	CVCVC	LI	plifu	CCVCV	LI
kapufik	CVCVCVC	LI	pliks	CCVCS	LD
kifapu	CVCVCV	LI	plu	CCV	LI
kip	CVC	LI	plusk	CCVSC	LD
kis	CVs	LD	pufaki	CVCVCV	LI
klifak	CCVCVC	LI	pukif	CVCVC	LI
klil	CCVL	LD	pusk	CVSC	LD
klipafu	CCVCVCV	LI	skafu	SCVCV	LD
klis	CCVS	LD	skap	SCVC	LD
klups	CCVCS	LD	skapufi	SCVCVCV	LD
kufalpi	CVCVLCV	LD	skla	SCCV	LD
kufłapi	CVCCVCV	LI	spaf	SCVC	LD
kupalfi	CVCVLCV	LD	spakifu	SCVCVCV	LD
kupifla	CVCVCCV	LI	spli	SCCV	LD
kusp	CVSC	LD	spu	sCV	LD

Nous pouvons maintenant comparer les trois épreuves, en fonction du type d'items, de leur degré de complexité, et des moyennes de leurs indices. Le tableau 29 nous donne les résultats de cette comparaison.

Tableau 29 : Comparaison des 3 épreuves

	LITMUS-NWR 71	LITMUS-NWR 50-β	LITMUS-NWR 50	
Type	30 items LI (42,3%)	20 items LI (40%)	21 items LI (42%)	
	41 items LD (57,7%)	30 items LD (60%)	29 items LD (58%)	
Complexité	A=21 (29,6%)	A=7 (14%)	A=11 (22%)	A= faible complexité
	B=24 (33,8%)	B=18 (36%)	B=18 (36%)	B= complexité moyenne
	C=26 (36,6%)	C=25 (50%)	C=21 (42%)	C= très complexe
Moyennes	$p = 0,77$	$p = 0,73$	$p = 0,74$	p = indice de difficulté
	$D = 0,47$	$D = 0,53$	$D = 0,55$	D = indice de discrimination
	$ET = 0,39$	$ET = 0,43$	$ET = 0,43$	ET = écart-type
	$r = 0,55$	$r = 0,55$	$r = 0,55$	r = corrélation item/total

Nous observons que la répartition entre les items de différente complexité est bien meilleure dans l'épreuve finale LITMUS-NWR 50. La moyenne des indices p de difficulté indique que les items de l'épreuve LITMUS-NWR 50 pris dans leur ensemble sont plus difficiles que ceux de LITMUS-NWR 71 et légèrement plus difficiles que ceux de LITMUS-NWR 50-β. Ils sont également plus discriminants dans l'épreuve LITMUS-NWR 50. L'écart-type moyen est le même pour LITMUS-NWR 50 et LITMUS-NWR 50-β, plus élevé que dans LITMUS-NWR 71. La corrélation item/total est la même pour les trois épreuves.

4.8 Estimation de la cohérence interne : calcul du coefficient alpha de Cronbach

Le coefficient alpha de Cronbach mesure la fidélité interne de l'épreuve. Nous allons le calculer pour l'épreuve de départ LITMUS-NWR 71 puis pour la première épreuve réduite LITMUS-NWR 50-β et enfin pour l'épreuve réduite finale LITMUS-NWR 50.

➤ Alpha de Cronbach pour LITMUS-NWR 71

Nous l'avons d'abord calculé à partir de l'épreuve originelle comportant 71 items :

$$\boxed{\text{Épreuve LITMUS-NWR 71 : } \alpha = 0,97}$$

Cette mesure indique que l'épreuve, dans sa version longue, a une excellente fidélité interne : les items dans leur ensemble mesurent bien le même construit. Le calcul de l'alpha sur l'épreuve réduite grâce à notre analyse d'items devra garder sensiblement le même niveau de cohérence. Comme nous l'avons vu, réduire une épreuve consiste à faire un compromis entre les contraintes de durée de passation et la fidélité, étant donné que plus l'épreuve est longue, plus sa fidélité risque d'être bonne.

➤ **Alpha de Cronbach pour LITMUS-NWR 50-β**

Au calcul de l'indice de fidélité interne de la première réduction de l'épreuve, nous obtenons le résultat suivant :

Épreuve LITMUS-NWR 50-β : $\alpha = 0,96$

Le résultat est légèrement inférieur, mais reste excellent.

➤ **Alpha de Cronbach pour LITMUS-NWR 50**

Au calcul de l'indice de fidélité interne de la réduction finale de l'épreuve, nous obtenons le résultat suivant :

Épreuve LITMUS-NWR 50: $\alpha = 0,96$
--

Le changement réalisé dans les choix des items pour l'épreuve réduite finale n'a pas fait baisser l'indice de fidélité interne, qui reste excellent.

L'analyse des items aboutissant à la réduction de l'épreuve LITMUS-NWR 71 ayant été effectuée, nous pouvons à présent discuter les résultats. Chaque analyse d'items étant liée à une épreuve, les résultats ne peuvent être comparés à ceux de la littérature dans ce domaine. Il s'agira donc de mettre en regard nos résultats avec la théorie qui a sous-tendu l'élaboration de l'épreuve.

5 Discussion

5.1 *Analyse des items de LITMUS-NWR : une analyse du construit*

Le but du présent travail était d'effectuer une analyse des items de l'épreuve LITMUS-NWR 71, afin d'en proposer une réduction, utilisable en clinique. Comme nous l'avons vu (cf. section 2.3.3), cette étape de l'élaboration d'une épreuve est également importante dans le cadre des hypothèses relatives au domaine qui ont servi à créer les items de départ : elle permet de mettre à l'épreuve la théorie, en lui amenant des arguments empiriques (Dickes et al., 1994).

L'analyse des résultats des indices de difficulté des items de l'épreuve permet d'observer que le niveau de difficulté des items impacte leur réussite par les enfants porteurs d'un trouble spécifique du langage. Les indices sont globalement plus bas (items plus difficiles) pour les enfants avec TSL que pour ceux au DT. La finalité clinique de l'épreuve est d'évaluer des sujets en difficulté dont les performances sont inférieures à celles attendues pour l'âge et le niveau scolaire. Les items ont été élaborés pour que l'épreuve plafonne pour les sujets au DT. Lorsque l'on met en regard les indices de difficulté pour chaque groupe (DT et TSL), on s'aperçoit qu'une grande majorité d'items est facile pour les enfants au DT. On peut donc considérer que ces résultats sont en accord avec le but de l'épreuve. Les items présentant peu de différence entre les groupes DT et TSL ont été éliminés lors de la réduction. D'après nos résultats, la création des items de l'épreuve LITMUS-NWR permet bien de proposer un construit valide pour différencier les enfants au DT des enfants avec TSL.

5.1.1 Complexité phonologique et TSL

L'analyse des items permet également d'observer que les items avec un degré de complexité syllabique moindre présentent un indice de difficulté plus élevé que ceux des items plus complexes sur l'ensemble des groupes. Cela signifie que pour notre groupe d'étude, plus l'item est phonologiquement complexe, moins il est réussi par les l'ensemble des enfants.

Pour aller plus loin dans l'analyse, nous avons comparé les moyennes des indices de difficulté des items pour les groupes d'enfants au DT et avec TSL. Nous nous sommes aperçu que les indices de difficulté des items pour le groupe des enfants avec TSL diminuent quand la complexité phonologique du non-mot augmente. Les enfants au développement typique présentaient quant à eux des performances qui ne sont pas en rapport avec la complexité phonologique de l'item. Cela signifie que plus le non-mot est complexe, moins il est réussi par les enfants avec TSL du groupe, alors que le facteur de complexité ne semble pas entrer en jeu dans la performance des enfants au développement typique du groupe. Nos résultats peuvent être mis en parallèle avec ceux des études de Gallon et al. (2007) et Ferré et al. (2012) qui ont révélé que les performances des enfants avec TSL diminuaient lorsque la complexité phonologique du non-mot augmentait.

D'autre part, nous avons vu que l'épreuve LITMUS-NWR, destinée aux enfants bilingues, était composée d'items « Langue-Indépendant » et d'items « Langue-Dépendant ». Les items LD présentant plus de points de complexité phonologique devraient être moins bien réussis par les enfants que les items LI : leur indice de difficulté devrait en moyenne être moins élevé. Or, nos résultats montrent que la différence n'est pas significative pour l'ensemble des enfants : les indices de difficulté sont très proches. Nous retrouvons des résultats similaires lorsque nous regardons les moyennes des indices de difficulté par groupe (DT et TSL). La différence entre les items LI et LD n'est donc pas significative sur notre échantillon.

5.1.2 Discrimination de l'épreuve

Le calcul de l'indice de discrimination a d'abord permis la répartition des enfants de notre population dans deux groupes : « fort » et « faible ». La majorité des enfants du groupe « fort » appartient au groupe des enfants au DT (Bi-DT=61%, Mo-DT=35%). 67% de l'ensemble des Mo-DT et 37% de l'ensemble des Bi-DT appartiennent donc au groupe ayant obtenu les meilleurs résultats à l'ensemble de l'épreuve. Dans le groupe « faible », une grande majorité des enfants (87%) appartient au groupe des Mo-TSL. Ces résultats indiquent que les meilleurs scores à l'épreuve sont obtenus par une majorité d'enfants au DT, et les scores les plus faibles sont obtenus par une majorité d'enfants avec TSL. Nous préférons en revanche ne pas tirer de conclusions hâtives sur les différences de performance entre les enfants bilingues

et monolingues, d'autant que l'étude étant encore en cours, le nombre d'enfants dans chacun des groupes va augmenter et ainsi permettre une analyse des données plus fine.

Grâce au calcul de l'indice D, nous avons pu remarquer que la majorité des items de l'épreuve LITMUS-NWR 71 discriminaient bien ou très bien. Les résultats de la corrélation entre l'indice de discrimination et l'indice de difficulté ont permis d'établir un lien entre la difficulté des items et leur capacité discriminatoire. Nous avons remarqué que plus le non-mot était difficile, plus il était discriminant. Or, nous avons vu également que le degré de difficulté du non-mot était lié à sa complexité phonologique. Dans le cadre de cette étude, nous observons donc que plus le non-mot est phonologiquement complexe, plus il a de chances de bien discriminer entre le groupe d'enfants avec TSL et le groupe d'enfant au DT.

5.2 Réduction de LITMUS-NWR : signification statistique et pertinence clinique

L'analyse des items est un outil statistique à la disposition des concepteurs de tests destinés à être utilisés en recherche comme en clinique. L'outil LITMUS-NWR 71 a été élaboré à des fins de recherche. Afin que l'épreuve soit plus facilement utilisable en clinique, nous avons fait appel à des données statistiques pour la réduire. Les raisons pour lesquelles cette réduction était nécessaire sont les suivantes :

- obtenir une épreuve plus courte, car il s'agit ici d'un test de dépistage dont la longueur ne devrait pas excéder cinq minutes : pour cela, nous avons retiré 21 items, pour arriver à une épreuve contenant 50 items,
- augmenter la capacité discriminatoire de l'épreuve, en retirant les items les moins pertinents.

Nous avons pu observer, lors de notre première tentative de réduction, que le résultat obtenu ne prenait pas en compte un facteur qui se révèle important pour une utilisation du test en clinique : la distribution du degré de difficulté des items. En effet, l'épreuve est destinée à de jeunes enfants, suspectés d'avoir des difficultés de langage. Avoir trop d'items difficiles risquait de compromettre la motivation qu'ils devraient garder tout le long de l'épreuve. Cette motivation, ainsi que la fatigabilité des enfants, doivent être prises en compte lors de la phase d'élaboration d'une épreuve. Comme le souligne Tourrette (2006 : 34), « l'évaluation des

enfants doit s'adapter à leur niveau supposé de développement de façon à maximiser leur coopération ». Nous avons donc réintroduit des items monosyllabiques plus faciles et retiré certains d'entre eux plus difficiles. Ces changements n'ont pas affecté les capacités discriminatoires de l'épreuve.

Ces observations nous permettent de souligner ici l'importance du maintien d'une perspective de pertinence clinique lors d'une analyse d'items basée sur la signification statistique. Même si l'attitude clinique relève plus de l'interprétation que de la norme, il nous a semblé important de nous rendre compte que les statistiques ne sont qu'un outil d'aide à la décision : mesurer est essentiel, mais n'est pas suffisant. Comprendre dans quel but et comment un outil de mesure tel que l'épreuve LITMUS-NWR a été conçu, pour quelle population les items ont été créés et quelles sont les conditions d'application clinique de cette épreuve doit permettre d'ajuster les résultats obtenus par les outils statistiques.

5.3 *Limites et perspectives*

5.3.1 Population

L'analyse des items, pour être pertinente, doit se baser sur un échantillonnage de population représentatif et suffisant. Notre population de 85 enfants représente une cohorte souvent supérieure à celles étudiées dans de nombreuses publications sur le sujet. Cependant, elle n'est pas suffisante pour proposer une analyse d'items basée sur la nouvelle approche statistique proposée par les MRI, et notamment par le modèle de Rasch (cf. section 2.5), qui demande une population d'au moins 100 sujets. Cette nouvelle approche aurait permis une analyse plus solide des items, avec des résultats ne dépendant pas de l'échantillon de population sur lequel les items ont été testés.

De plus, la répartition de notre population de référence n'est pas égale entre les enfants au DT (51 enfants) et ceux avec TSL (34 enfants). Les difficultés de recrutement des enfants bilingues suivis en orthophonie (avec une contrainte pour notre étude sur les langues parlées) n'ont pas permis d'avoir un nombre d'enfants équivalent dans les deux groupes. À l'intérieur même des groupes, la répartition n'est pas homogène. Même si les outils statistiques dont

nous nous sommes servis se basent sur des calculs minimisant ces biais, il aurait été intéressant d'effectuer ces calculs avec une plus large population d'enfants avec TSL. En outre, les enfants monolingues avec TSL ont été recrutés pour la plupart dans un Centre de Référence Langage. Ces centres voient arriver une population d'enfants porteurs a priori de troubles plus graves que ceux que l'on peut rencontrer dans les cabinets d'orthophonie. Notre analyse des items aurait gagné en pertinence si elle avait pu comporter plus de résultats d'enfants avec un trouble moins important.

L'âge des enfants de notre échantillon de population peut également être sujet à discussion. Comme nous l'avons précisé, les contraintes d'âge imposées par l'action de recherche ont été fixées pour pouvoir comparer les acquisitions en phonologie et en syntaxe. Les résultats de notre analyse d'items se basent donc sur une population d'enfants en partie trop âgés en ce qui concerne l'acquisition de la phonologie. C'est pourquoi nous n'avons pas pu présenter de résultats en fonction de l'âge des enfants testés. Il serait donc intéressant de faire passer l'épreuve à des enfants aux alentours de l'âge de 4 ans pour une plus grande pertinence des résultats.

5.3.2 Construction des items : cadre de recherche vs utilisation clinique

Les items de l'épreuve LITMUS-NWR 71 ont été créés d'abord dans une perspective de recherche. Or, comme le notent Tourrette et al. (2012), les objectifs d'une évaluation sont différents selon que l'on se trouve dans un cadre de pratique clinique ou dans un cadre de recherche. La recherche va d'abord permettre de comprendre des fonctionnements, ce qui va dans un second temps aider à mettre au point des instruments valides. Dans le cadre de l'élaboration de l'épreuve LITMUS-NWR 71, les chercheurs ont créé des items permettant, par exemple, de déterminer si les séquences labiale...dorsale sont plus problématiques dans leur réalisation que les séquences dorsale...labiale (de Almeida et al., 2014). Ces études permettront de mieux cerner l'acquisition de la phonologie chez l'enfant, et de mieux comprendre la pathologie.

Cependant, cette perspective de recherche a conduit à créer des items qui doivent pouvoir être comparés, donc qui comportent les mêmes phonèmes placés à des endroits

différents (ex : [kifapu] vs [pufaki]). De plus, comme il s'agit d'une épreuve testant des enfants bilingues, il a fallu isoler des phonèmes utilisés dans le plus grand nombre de langues possibles. Or, dans une perspective clinique, on pourrait objecter à l'épreuve qu'elle comporte beaucoup d'items qui se ressemblent du point de vue de l'enfant, mais aussi de celui de l'orthophoniste. Cette « ressemblance » peut apparaître pour les cliniciens comme un biais possible lors de la passation, l'enfant pouvant avoir un sentiment de « déjà-dit ». Pourtant, lors des passations que nous avons effectuées, nous n'avons pas observé de remarques traduisant ce sentiment de la part des enfants. Il sera donc nécessaire de bien expliquer les enjeux d'une telle épreuve aux orthophonistes qui souhaiteraient l'utiliser, ainsi que la nécessité d'utiliser des non-mots phonétiquement proches pour permettre l'évaluation bilingue dans le plus grand nombre de langues possibles.

5.3.3 Conditions de passation

Les non-mots de l'épreuve LITMUS-NWR 71 ont été enregistrés, afin que les conditions de passation soient les plus contrôlées possibles pour obtenir une épreuve standardisée. Cependant, même si l'utilisation de l'ordinateur pour faire passer des tests est de plus en plus répandue dans les cabinets d'orthophonie, elle n'est pas généralisée. Certains cliniciens, peu habitués aux outils informatiques, sont réticents quant à leur utilisation. De plus, les possibilités offertes par l'ordinateur comme le calcul du temps de réponse ne sont pas ici exploitées, car elles sont plus pertinentes dans le cadre d'une épreuve d'accès au lexique par exemple que dans celui d'une répétition de non-mots évaluant la phonologie.

D'autre part, l'évaluation par ordinateur des enfants avec un trouble du langage peut poser un certain nombre de difficultés, liées au trouble lui-même : il sera difficile pour l'orthophoniste de revenir sur des non-mots non répétés par l'enfant pour des raisons dues à la fatigue générée par la surcharge cognitive exigée par l'exercice, le manque d'estime de soi, ou par la possible intimidation de l'enfant générée par la situation même de bilan. Même si l'évaluateur s'éloigne ainsi de l'objectivité que requiert une évaluation normée, et risque par là-même d'invalider l'outil dont il se sert, il s'agit également en situation clinique d'arriver à placer l'enfant dans des conditions optimales pour être évalué (Tourrette, 2006). Ces conditions requièrent bien souvent de pouvoir « adapter » l'épreuve à l'enfant, ce qui est difficile quand l'épreuve est pilotée par un ordinateur. Tourrette (2006) indique également

que les conditions standardisées de passation prévues par le protocole du test sont rarement présentes en clinique et parle « d'aménagements réfléchis » (2006 : 35). Pour ce que nous avons pu observer et ce dont nous avons pu faire nous-même l'expérience, l'orthophoniste, en situation clinique, fait sans cesse des compromis entre les contraintes de passation imposées par les manuels de tests et les contraintes imposées par les enfants eux-mêmes. Une épreuve enregistrée tient certainement compte de la première contrainte, mais ne prend pas en compte la seconde.

Il serait intéressant de faire passer l'épreuve LITMUS-NWR 71 sans le support de l'ordinateur, l'orthophoniste faisant répéter les non-mots à l'enfant. Dans ce cas, il sera important de garder la dimension ludique apportée par le petit extra-terrestre (l'orthophoniste pourrait, par exemple, dire à l'enfant qu'un extra-terrestre lui a appris des mots dans sa langue). Comme le note Tourrette (2006 : 35) à propos des tests psychologiques, « les situations proposées doivent avoir du sens car l'interprétation que l'on fera des résultats obtenus par l'enfant est largement fonction de la façon dont il s'est engagé dans la tâche ». Cette réflexion nous semble également valable pour les tests de langage.

Si l'on prend soin de vérifier la stabilité des paramètres d'items lorsqu'on les transpose d'un support informatisé à un support papier-crayon (Martin, 2003) et si les résultats permettent de valider ses capacités discriminatoires, l'épreuve pourrait être proposée en version papier-crayon. Néanmoins, il faut aussi tenir compte des autres épreuves qui seront proposées dans le cadre de la batterie de dépistage. Si la batterie a pour vocation d'être totalement informatisée, cette suggestion n'est pas valable. Si, en revanche, l'épreuve LITMUS-NWR 71 peut être proposée seule, une version papier-crayon pourrait être une option.

5.3.4 Transcription des non-mots répétés

Afin qu'elle puisse être correctement évaluée, la production des enfants doit être transcrite aussi fidèlement que possible. Dans le cadre de notre étude, nous avons vu que les passations de l'épreuve LITMUS-NWR 71 étaient enregistrées. Les productions étaient ensuite transcrites par un premier expérimentateur, puis la transcription vérifiée par un

deuxième expérimentateur. Nous avons pu nous apercevoir que, même dans ces conditions, les transcriptions pouvaient fortement varier d'une personne à l'autre. Cependant, le système de notation dichotomique (0/1) de l'épreuve permet de minimiser ce biais. En effet, même si les deux expérimentateurs n'étaient pas d'accord, la discussion portait le plus souvent sur l'exacte différence entre le non-mot répété et le non-mot cible, et non sur le fait que le non-mot était bien répété ou non.

Dans le cadre d'une utilisation clinique, cette relecture est évidemment impossible. L'enregistrement lui-même peut être difficile à mettre en place en clinique, et nécessite une réécoute qui prend du temps. Il serait intéressant dans des perspectives d'études futures d'évaluer des transcriptions effectuées sans l'aide de l'écoute de l'enregistrement. Cela permettrait de savoir si l'épreuve pourrait être proposée en transcription directe sans perdre ses capacités discriminatoires.

5.3.5 Temps de passation

Nous n'avons pas pu tester le temps de passation moyen de l'épreuve LITMUS-NWR 50 sur des enfants présentant un TSL. Néanmoins, nos essais sur cinq enfants au développement typique de l'âge de la population visée situent la durée de passation de l'épreuve aux alentours de 5 minutes, en comptant l'explication des consignes et la mise en condition de l'enfant (ajustement du casque audio, etc.). L'épreuve en elle-même dure environ 2 minutes. Il serait important de généraliser ces premiers essais dans des études futures, afin de pouvoir indiquer un temps moyen de passation.

5.3.6 Ordre des non-mots dans l'épreuve LITMUS-NWR 50

L'ordre des items dans l'épreuve LITMUS-NWR 71 a été établi de façon « pseudo-aléatoire ». Dans un premier temps, les items ont été distribués aléatoirement. Puis, cet ordre s'est vu modifié afin d'éviter que des items trop proches phonologiquement se suivent dans l'épreuve. L'analyse des items, qui a abouti à l'élaboration de l'épreuve écourtée LITMUS-NWR 50, a permis d'ajouter à ce facteur l'indice de difficulté p des items. Nous pourrions utiliser cette nouvelle donnée afin de proposer un ordre qui en tiendrait également compte. Les paramètres entrant en ligne de compte pour l'établissement d'un ordre dans les items de l'épreuve LITMUS-NWR 50 seraient donc les suivants :

- a) Le degré de difficulté des non-mots devrait être réparti dans l'épreuve en prenant soin de proposer un item facile après une série plus difficile.

Dans les tests qui sont des échelles de développement, il est souvent d'usage de commencer par des items faciles puis de compliquer la tâche au fil de l'examen par des items de plus en plus difficiles (Tourrette, 2006). Cette procédure ne nous semble pas pertinente dans le cadre de l'épreuve LITMUS-NWR 50. Pour un enfant avec TSL dont les troubles sont d'ordre phonologique, il serait difficile et peu motivant de rester sur un constat d'échec à la fin de l'épreuve. De plus, mettre tous les items difficiles en fin d'épreuve reviendrait à prendre le risque que l'enfant « baisse les bras » après une série d'échecs et refuse de terminer l'épreuve.

- b) Comme dans LITMUS-NWR 71, les non-mots devraient être répartis afin d'éviter que des items phonologiquement trop proches ne se suivent.

En effet, nous avons vu que les items étaient tous composés des mêmes consonnes [p], [k], [f], [l] et [s]. Nous avons vu également que ce choix était dicté par des contraintes dues à l'application de l'épreuve dans un contexte de bilinguisme. Les non-mots possibles sont donc limités, et nous avons déjà discuté du sentiment de « déjà-dit » généré par ces impératifs de création des items, ainsi que des difficultés qu'il peut engendrer (cf. section 5.3.2). Afin d'éviter au maximum l'effet de « ressemblance », il est possible d'établir un ordre qui prenne en compte la répartition des consonnes et des voyelles au sein des items. Par exemple, faire suivre les items [flukif] et [kifapu] permet de ne pas faire entrer en jeu le facteur « ressemblance ». D'autre part, les items monosyllabiques (faciles ou difficiles), même s'ils sont composés des mêmes consonnes que les autres, permettent de couper le rythme et « sonnent » différemment. Ainsi, si l'on insère par exemple [kis] entre [flukif] et [fupli], on devrait minimiser l'effet.

- c) Le premier non-mot devrait être un item facile monosyllabique.

Le non-mot qui ouvre l'épreuve déclenche également la motivation de l'enfant. Comme nous l'avons indiqué en section 4.5, nous choisissons l'item [kip], qui se trouvait également en première position dans LITMUS-NWR 71. Son indice de difficulté p est de 0,94, ce qui le place dans les items les plus faciles, donc les plus

réussis par l'ensemble des enfants. Son indice p pour le groupe des enfants avec TSL est de 0,88, ce qui en fait un des items les plus faciles également pour ce groupe.

En suivant les contraintes imposées par ces paramètres, nous sommes donc à même de proposer dans le tableau 30 un ordre possible pour la passation de l'épreuve LITMUS-NWR 50.

Tableau 30 : Ordre possible des items de LITMUS-NWR 50

Ordre	Item	Structure	Type	Complexité	p	Ordre	Item	Structure	Type	Complexité	p
1	kip	CVC	LI	A	0,94	26	kufalpi	CVCVLCV	LD	C	0,60
2	fapus	CVCVs	LD	A	0,76	27	plifu	CCVCV	LI	B	0,86
3	kusp	CVsC	LD	B	0,64	28	kafip	CVCVC	LI	A	0,74
4	paklu	CVCCV	LI	B	0,81	29	pifakup	CVCVCVC	LI	C	0,68
5	flipuka	CCVCVCV	LI	C	0,62	30	fupli	CVCCV	LI	B	0,75
6	plusk	CCVsC	LD	C	0,64	31	skla	sCCV	LD	B	0,80
7	pukif	CVCVC	LI	A	0,75	32	klipafu	CCVCVCV	LI	C	0,67
8	skafu	sCVCV	LD	B	0,82	33	klil	CCVL	LD	B	0,84
9	klis	CCVs	LD	B	0,78	34	pufaki	CVCVCV	LI	A	0,85
10	kupalpi	CVCVLCV	LD	C	0,61	35	filpa	CVLCV	LD	B	0,79
11	plu	CCV	LI	A	0,91	36	plaklu	CCVCCV	LI	C	0,62
12	spakifu	sCVCVCV	LD	C	0,71	37	pafuski	CVCVsCV	LD	C	0,69
13	plal	CCVL	LD	B	0,85	38	skap	sCVC	LD	B	0,88
14	kapufik	CVCVCVC	LI	C	0,74	39	klups	CCVCs	LD	C	0,65
15	fips	CVCs	LD	B	0,60	40	flipuka	CCVCVCV	LI	C	0,62
16	piklafu	CVCCVCV	LI	C	0,61	41	fuk	CVC	LI	A	0,91
17	spu	sCV	LD	A	0,88	42	pifukas	CVCVCVs	LD	C	0,59
18	klifak	CCVCVC	LI	C	0,67	43	kufalpi	CVCCVCV	LI	C	0,74
19	flis	CCVs	LD	B	0,81	44	spaf	sCVC	LD	B	0,78
20	kupifla	CVCVCCV	LI	C	0,49	45	kuspa	CVsCV	LD	B	0,76
21	flaplu	CCVCCV	LI	C	0,53	46	pusk	CVsC	LD	B	0,75
22	kis	CVs	LD	A	0,88	47	spli	sCCV	LD	B	0,82
23	skapufi	sCVCVCV	LD	C	0,67	48	fikuspa	CVCVsCV	LD	C	0,71
24	piks	CVCs	LD	B	0,76	49	plik	CCVCs	LD	C	0,75
25	flukif	CCVCVC	LI	C	0,61	50	kifapu	CVCVCV	LI	A	0,76

6 Conclusion

L'épreuve de répétition de non-mots LITMUS-NWR 71 a été conçue dans le but d'étudier la nature du déficit phonologique des enfants bilingues avec TSL, ainsi que de pallier le manque de tests proposés pour l'évaluation du langage de l'enfant dans un contexte de bilinguisme. Elle constitue ainsi un outil de dépistage des troubles phonologiques, en se basant sur l'hypothèse de la sensibilité à la complexité syllabique. Elle comporte 71 items, dont une partie a été créée pour être utilisable dans une majorité de langues du monde (LI), l'autre partie comprenant des contraintes phonotactiques spécifiques, entre autres, au français (LD). Notre étude consistait à réduire cette épreuve afin qu'elle soit plus facilement utilisable en clinique. Cette réduction a abouti à une épreuve comportant 50 items, LITMUS-NWR 50, dont la durée de passation ne devrait pas excéder 5 minutes.

Nous avons effectué une analyse des items en suivant les préceptes de la théorie classique du score vrai. Notre échantillon de population n'était pas suffisant pour faire une analyse selon les MRI.

Le calcul des différents indices nous a permis de classer les items en fonction de leur degré de difficulté, de discrimination, de leur écart-type et de leur corrélation avec l'ensemble de l'épreuve. À partir de ces calculs, nous avons pu isoler un certain nombre d'items à éliminer pour proposer une première tentative de réduction, que nous avons mise à l'épreuve de critères cliniques. Ces critères nous ont permis de réajuster nos résultats pour enfin présenter une proposition de réduction définitive. La fidélité de l'épreuve a été contrôlée par le calcul du coefficient alpha de Cronbach. Enfin, nous avons proposé un ordre de passation des items au sein de l'épreuve réduite.

Nos résultats ont également abouti à la validation du construit de l'épreuve pour différencier les enfants au DT des enfants avec TSL. La comparaison de l'indice de difficulté avec le degré de complexité des items, établi au préalable selon l'hypothèse de la complexité phonologique, a également permis de montrer la sensibilité des enfants avec TSL à ce facteur, que l'on ne retrouve pas chez les enfants au DT.

Notre étude montre également de manière plus large l'importance pour l'orthophoniste de la connaissance des outils d'évaluation qu'il utilise dans sa pratique quotidienne des tests, ainsi que l'importance de la prise en compte du potentiel thérapeutique de l'outil par les

concepteurs de ces mêmes tests. Connaître la manière dont une épreuve est construite, les cadres théoriques qui la sous-tendent, le choix des items, est nécessaire à une utilisation pertinente et adaptée de l'outil d'évaluation. Prendre le temps de comprendre la manière dont l'outil peut être utilisé, c'est prendre en compte le fait qu'évaluer, c'est déjà engager un « processus de changement » (Tourrette, 2006 : 4). Cette perspective de changement se doit de reposer sur l'assurance que les concepteurs de l'épreuve et leurs utilisateurs ont reconnu sa portée pour l'avenir de l'enfant.

7 BIBLIOGRAPHIE

Abi-Aad, K. et Atallah, C. (2012) « Phonologie, Plurilinguisme et Trouble Spécifique du Langage Oral au Liban : Etude Pilote sur la Pertinence d'un Test de Répétition de Non-Mots ». Mémoire en vue de l'obtention du Master recherche en orthophonie, Université Saint-Joseph de Beyrouth.

Angoujard, J-P. (1997) *Théorie de la syllabe*. Paris : CNRS Editions.

Armon-Lotem, S. et Chiat, S. (2012) « How do Sequential Bilingual Children Perform on Nonword Repetition Tasks ». *Proceedings of the 36th annual Boston University Conference on Language Development*. Somerville, MA : Cascadilla Press : ed. Alia K. Biller et al: 63-73.

Bernaud, J-L. (1998) *Les méthodes d'évaluation de la personnalité* (2ème édition). Paris : Editions Dunod.

Bertrand, R. et Blais J-G (2004) *Modèles de mesure. L'apport de la théorie des réponses aux items*. Ste-Foy (Québec) Canada : Presse de l'Université du Québec.

Bowen, C. (2007) *Les difficultés phonologiques de l'enfant*. Montréal (Québec) : Chenelière Education.

Brousseau, A-M. et Nikiema, E. (2001) *Phonologie et Morphologie du français*. Montréal : Editions Fides.

Chartier, P. et Loarer, E. (2008) *Evaluer l'intelligence logique*. Paris : Dunod.

Chen, W-H, Lenderking, W., Jin, Y., Wyrwich, K.W., Gelhorn, H., Revicki, D.A. (2014) « Is Rasch Model Analysis Applicable in Small Sample Size Pilot Studies for Assessing Item Characteristics? An Example Using PROMIS Pain Behavior Item Bank Data ». *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation* 23, (2): 485–93.

Chevrie-Muller, C. et Plaza, M. (2001) *Nouvelles épreuves pour l'examen du langage, N-EEL*. Paris : ECPA.

Coady, J. A. et Evans, J. L. (2008) « Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI) ». *International Journal of*

Language & Communication Disorders 43: 1-40. 106.

Coady, J. A., Evans, J. L., Kluender, K. R. (2010) « Role of phonotactic frequency in nonword repetition by children with specific language impairments ». *International Journal of Language & Communication Disorders* 45: 494-509.

COST Action IS0804 (2011) Questionnaire for Parents of Bilingual Children (Pabi Questionnaire). Consulté le 25 février 2014. <http://www.bi-sli.org>.

Cronbach, L.J. (1951) « Coefficient alpha and the internal structure of tests ». *Psychometrika* 16(3): 297-334.

de Almeida, L., dos Santos C., Ferré, S. (2014). When Place changes Place: is /kifapu/ the same as /pufaki/ ?. Communication orale présentée aux 12^{èmes} rencontres du Réseau français de phonologie, 30 juin 2014, Lille, France.

Dickes, P., Tournois, J., Flieller, A., Kop, J-L. (1994) *La psychométrie*. Vendôme : Editions puf

Dollaghan, C. et Campbell, T. (1998) « Nonword Repetition and Child Language Impairment ». *Journal of Speech, Language, and Hearing Research* 41: 1136-1146

dos Santos, C. (2007) « Développement phonologique en français : une étude de cas ». Thèse de doctorat, département des sciences du langage, université Lumière Lyon 2.

Ebel, R. L. (1972) *Essentials of educational measurement*. Oxford, England : Prentice-Hall

Ferré, S. (2006). La complexité phonologique à 3 ans. [PDF] consulté le 18 mars 2014.

http://ia37.tice.ac-orleans-tours.fr/eva/sites/ia37__mission-maitrise-langue/IMG/File/conferences/La-complexite-phonologique-3ans-synthese.pdf

Ferré, S., Tuller, L., Piller, A-G., Barthez, M-A., (2011) « Strategies of avoidance in (a)typical development of French », in Dominguez, L. et Guijarres-Fuentes, P. (eds.) *Selected proceedings of the Romance Turn III*. Cambridge: Cambridge Scholar Publisher.

Ferré, S., Tuller, L., Sizaret, E., Barthez, M-A. (2012) « Acquiring and avoiding phonological complexity in SLI vs. typical development of French : The case of consonant clusters » in Hoole, Bombien, Pouplier, Mooshammer et Kühnert (eds.) *Consonant Clusters and Structural Complexity*. Germany: DE GRUYTER, pp: 285-308.

Frisch, S. A., Large, N. R., Pisoni, D. B. (2000) « Perception of Wordlikeness : Effects of

Segment Probability and Length on the Processing of Nonwords ». *Journal of Memory and Language* 42 : 481-4

Gallon, N., Harris, J. et van der Lely, H. (2007) « Non-word repetition : an investigation of phonological complexity in children with Grammatical SLI ». *Clinical Linguistics & Phonetics* 21(6): 445-455. 107

Gathercole, S. E., Willis, C. S., Baddeley, A.D. et Emslie, H. (1994) « The Children's Test of Nonword Repetition: a test of phonological working memory ». *Memory* 2 : 103-127.

Glaser, R. (1963) « Instructional Technology & The Measurement Of Learning Outcomes ». *American Psychologist* 18: 519-521

Grégoire, J. (2005) « Aspects psychométriques de la construction d'épreuves d'évaluation du langage », in Pierart, B. *Le langage de l'enfant, comment l'évaluer*. De Boeck supérieur, Bruxelles, pp 35-48

Grimm, A., Ferré, S., dos Santos, C., Shiat, S. (2014). Can nonwords be language-independent ? Cross-linguistic evidence from monolingual and bilingual acquisition of French, Lebanese and German. Communication orale présentée au IASCL 2014, 14-18 juillet 2014, Amsterdam, Pays-Bas.

Gutierrez-Clellen, V. F. et Simon-Cereijido, G. (2010). « Using nonword repetition tasks for the identification of language impairment in spanish-english-speaking children : does the language of assessment matter? ». *Learning Disabilities Research & Practice* 25: 48-58.

Harris, J. et Gussmann, E. (1998). « Final Codas : why the west was wrong », in Cyran, E. (ed.), *Structure and interpretation in phonology : studies in phonology*, Lublin : Folia, pp. 139-162.

Huteau, M. et Lautrey, J. (2006) « Les tests d'intelligence ». *Repères* : 3-4

Iluz-Cohen, P. (2010). Tasks testing executive control: Bilingual and Monolingual children with TLD and SLI. Communication orale présentée au COST Action IS0804 Cyprus meeting, 28 octobre 2010, Larcana, Chypre.

Kaye, J. et Lowenstamm, J. (1984) « De la syllababilité », in Dell, F., Hirst, D., Vergnaud J-R. (eds.) *Forme sonore du langage*. Paris : Hermann, pp. 123-161.

Khomsis, A., Khomsis, J., Pasquet, F., Parbeau-Gueno, A. (2007) *Bilans informatisés de langage oral (BILO)*, ECPA.

Laveault, D. et Grégoire, J. (1997) *Introduction aux théories des tests en sciences humaines*. Bruxelles : De Boeck Université.

Lelièvre, L. (2013) « Troubles Spécifiques du Langage et bilinguisme : intérêts d'un test de répétition de non-mots ». Mémoire en vue de l'obtention du certificat de capacité d'orthophoniste, Université François Rabelais de Tours, Ecole d'orthophonie.

Levelt, C.C., Schiller, N.O., Levelt, W.J. (1999) « The Acquisition of syllable types ». *Language acquisition* 8(3), 237-264.

Maddieson, I. (2006) « Correlating phonological complexity : data and validation ». *Linguistic Topology* 10 : 108-125

Marshall, C., Ebbels, S., Harris, J., van der Lely, H.K.J. (2002) « Investigating the impact of prosodic complexity on the speech of children with Specific Language Impairment ». *UCL Working Papers in Linguistics* 14: 43-66.

Martin, O. (1997) *La mesure de l'esprit : origines et développement de la psychométrie 1900-1950*. Paris : L'Harmattan

Martin, R. (2003) « Le testing adaptatif par ordinateur dans la mesure en éducation : potentialités et limites ». *Psychologie et Psychométrie*, 24 (2/3): 89-116

McLeod, S., Van Doorn, J., & Reed, V. A. (2001) « Normal acquisition of consonant clusters ». *American Journal of Speech-Language Pathology*, 10: 99–110.

Meynadier, Y. (2001) « La syllabe phonétique et phonologique : une introduction ». *Travaux interdisciplinaires du laboratoire Parole et Langage. Volume 20* : pp 91-148.

Nunnally, J. C. et Bernstein, I. H. (1994) *Psychometric Theory, Third Edition*. New-York : McGraw-Hill.

OCDE – Nations Unies/DEAS (2013) *Les migrations internationales en chiffres*. Consulté le 22 août 2014, à l'adresse <http://www.oecd.org/fr/els/mig/les-migrations-internationales-en-chiffres.pdf>

Paradis, J., Crago, M., Genesee, F. et Rice, M. (2003) « French-english bilingual children

with SLI: How do they compare with their monolingual peers ? ». *Journal of speech, language and hearing research* 46: 113-127.

Paradis, J. (2005) « Grammatical Morphology in Children Learning English as a Second Language : Implications of Similarities With Specific Language Impairment». *Language, Speech, and Hearing Services in Schools* 36 : 172-187.

Paradis, J. (2010) « The interface between bilingual development and specific language impairment ». *Applied Psycholinguistics* 31 : 227-252.

Paradis, J., Emmerzael, K. et Sorenson Duncan, T. (2010) « Assessment of English language learners : Using parent report on first language development ». *Journal of communication disorders* 43: 474-497.

Parisse, C. et Maillart, C. (2005) « Interférences entre phonologie et syntaxe en pathologie développementale du langage ». *Hommage à Elizabeth Bates, Le Langage et l'Homme*, N° Spécial, vol.40/2 : 127-147

Pichot, P. (1997) *Les tests mentaux* (15e édition refondue). Que sais-je n°626, Paris, puf.

Raven, J.C. (1964). *Standard progressives matrices. Séries A, B, C, D et E.* (révision 1956). PM38-T. Paris : EAP Editions Scientifiques et Psychotechniques.

Rondal J.A., Esperet E., Gombert J.E., Thibaut J.-P. et Comblain A. (2003) « Développement du langage oral » in Rondal J.A. et Seron X. (eds) *Troubles du langage. Bases théoriques, diagnostic et rééducation.* Mardaga, pp 114-120.

Rothweiler, M., Chilla, S., Clahsen, H. (2012) « Subject-verb agreement in specific language impairment : a study of monolingual and bilingual German-speaking children » *Bilingualism : Language and Cognition*, 15(1) : 39-57

Rvachew, S., Marquis, A., Brosseau-Lapré, F., Paul, M., Royle, P., Gonnerman L-M, (2013) « Speech articulation performance of francophone children in the early school years : Norming of the Test de Dépistage Francophone de Phonologie ». *Clinical Linguistics & Phonetics*, 27(12): 950-968

Saumon, C. (2012) « Dépistage du TSL chez l'enfant bilingue : intérêt du questionnaire parental dans la population arabophone ». Mémoire en vue de l'obtention du certificat de capacité d'orthophoniste, Université François Rabelais, Ecole d'orthophonie, Tours.

- Schulz, P. et Tracy, R. (2011). *Linguistische Sprachstandserhebung – Deutsch als Zweitsprache (LiSe-DaZ)*. Göttingen: Hogrefe Verlag.
- Selltiz, C., Wrightsman, L. S., Balch, G. I., Cook, S. W., Hofstetter, R., Bickman, L. (1977) *Les méthodes de recherche en sciences sociales*. Montréal : Les Editions HRW, Paris diffusion Doin.
- Stokes, S.F. et Klee, T. (2009) « The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children ». *Journal of Speech, Language and Hearing Research* 52: 872-882.
- Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N. et Chilingaryan, G. (2011) « Sensitivity and Specificity of French Language and Processing Measures for the Identification of Primary Language Impairment at Age 5 ». *Journal of Speech, Language, and Hearing Research* 54: 580-597.
- Thordardottir, E. et Brandeker, M. (2012) « The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores ». *Journal of Communication Disorders* 46(1) : 1-16.
- Thorndike, E.L. (1926) *The measurement of intelligence*. New-York : Teacher's College, Columbia University.
- Tourrette, C. (2006) *Evaluer les enfants avec déficiences ou troubles du développement*. Paris : Dunod.
- Tourrette, C., Guedeney, A. (2012) *L'évaluation clinique du jeune enfant*. Paris : Dunod
- van der Lely, H.K.J. et Harris, J. (1999) *The Test of Phonological Structure*. London UK : University College London.
- Verhoeven, L. et Vermeer, A. (2002). *Taaltoets Alle Kinderen (TAK)*. Tilburg: Zwijsen.
- Vrignaud, P., Castro, D., Mogenet, J-L. (2003) « Recommandations internationales pour l'utilisation des tests ». *Pratiques Psychologiques* Numéro spécial hors-série juin 2003, Adaptation française.
- Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., Jones, M. (2000) « Nonword Repetition Performance in School-Age Children With and Without

Language Impairment ». *Journal of Speech, Language, and Hearing Research* 43: 865-878.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—4th Edition (WISC-IV®)*. San Antonio, TX: Harcourt Assessment.

Yavas, M. (2004) « Bilingualism : speech issues » in Kent, R.D (ed.) *The MIT encyclopedia of communication disorders*. Cambridge : The MIT press, pp: 119-121.

ANNEXE 1 : liste des sujets de l'échantillon de population

Sujet	Age L2 session (french)	Groupe	Sujet	Age L2 session (french)	Groupe
CLR	5;2	Bi-DT	KYP	6;5	Mo-DT
SAH	5;2	Bi-DT	MAH	6;6	Bi-DT
GAB	5;3	Bi-TSL	HAC	6;6	Bi-DT
TVA	5;5	Bi-TSL	AMB	6;7	Bi-DT
FAA	5;5	Bi-DT	AMT	6;8	Bi-TSL
OMD	5;6	Bi-TSL	ISA	6;8	Bi-TSL
TAS	5;6	Bi-TSL	LIC	6;9	Bi-DT
RAM	5;6	Bi-DT	ESA	6;9	Bi-DT
ELB	5;6	Bi-DT	IME	6;10	Bi-TSL
VAH	5;6	Mo-DT	MOB	6;10	Mo-TSL
ADB	5;7	Bi-DT	PAM	7;0	Bi-TSL
CHI	5;7	Bi-DT	GED	7;0	Bi-DT
TSS	5;7	Bi-DT	AND	7;1	Bi-DT
LMP	5;7	Bi-DT	GAV	7;1	Mo-TSL
MAS	5;7	Mo-DT	CEM	7;2	Mo-TSL
OCS	5;7	Mo-DT	LIB	7;2	Mo-TSL
THP	5;8	Mo-DT	LOP	7;4	Mo-TSL
LEB	5;9	Mo-DT	CLM	7;5	Bi-TSL
SAC	5;10	Bi-DT	MAM	7;5	Mo-TSL
FTM	5;10	Bi-DT	JOC	7;7	Bi-TSL
RAB	5;10	Bi-DT	LDP	7;7	Bi-DT
MAD	5;10	Mo-DT	AMP	7;9	Bi-DT
MIE	5;11	Bi-DT	AUC	7;9	Mo-TSL
SAM	6;0	Bi-TSL	ELV	7;9	Mo-TSL
KHT	6;0	Bi-DT	AMA	7;10	Bi-DT
LEM	6;0	Bi-DT	ALB	7;11	Bi-DT
ANB	6;0	Mo-DT	FAS	7;11	Bi-DT
IBM	6;1	Bi-DT	KOS	7;11	Bi-DT
ISL	6;1	Bi-DT	JOP	8;2	Bi-TSL
WYR	6;1	Mo-DT	LEP	8;2	Bi-TSL
AYT	6;2	Bi-TSL	LEG	8;2	Bi-DT
ANH	6;2	Bi-DT	CHB	8;2	Bi-DT
BEA	6;2	Bi-DT	NBE	8;2	Mo-TSL
MLM	6;2	Mo-DT	LMA	8;3	Bi-TSL
MAB	6;3	Bi-TSL	CHD	8;3	Mo-TSL
CLK	6;3	Mo-TSL	VAM	8;3	Mo-TSL
CMU	6;3	Mo-TSL	JOV	8;4	Bi-TSL
CHN	6;4	Bi-DT	DDS	8;5	Bi-TSL
CHR	6;4	Bi-DT	AMO	8;5	Mo-TSL
DSC	6;4	Bi-DT	MAL	8;7	Mo-TSL
ALA	6;4	Bi-DT			
LIT	6;4	Mo-DT			
MLI	6;4	Mo-DT			
ZOA	6;5	Bi-DT			

ANNEXE 2 : Indice de difficulté des items

Item	Indice difficulté p	Item	Indice difficulté p
faku	0,91	kufalpi	0,60
fal	0,96	kuflapi	0,74
fapul	0,87	kupalfi	0,61
fapus	0,76	kupifla	0,49
fikapul	0,85	kusp	0,64
fikupla	0,72	kuspa	0,76
fikuspa	0,71	lafi	0,94
filpa	0,79	paf	0,96
fips	0,60	pafuski	0,69
fiska	0,87	pakifal	0,82
flaplu	0,53	paklu	0,81
fli	0,95	pifakup	0,68
flipuka	0,62	pifukas	0,59
flis	0,81	piklafu	0,61
fluk	0,88	piks	0,76
fluka	0,87	pilfu	0,65
flukif	0,61	pilu	0,96
fuk	0,91	plaklu	0,62
fupli	0,75	plal	0,85
kafip	0,74	plifu	0,86
kapi	0,96	pliks	0,75
kapufik	0,74	plu	0,91
kifapu	0,76	plusk	0,64
kifapus	0,79	pufaki	0,85
kifus	0,87	pukif	0,75
kip	0,94	pusk	0,75
kis	0,88	skafu	0,82
kla	0,95	skap	0,88
klaf	0,87	skapufi	0,67
klifak	0,67	skla	0,80
klil	0,84	spaf	0,78
klipafu	0,67	spakifu	0,71
klis	0,78	spiku	0,89
klisp	0,32	spli	0,82
klups	0,65	spu	0,88
kufal	0,92		

ANNEXE 3 : indice de difficulté p et complexité syllabique des items

item	structure	type	complexité	difficulté p	item	structure	type	complexité	difficulté p
faku	CVCV	LI	A	0,91	plal	CCVL	LD	B	0,85
fal	CVL	LD	A	0,96	plifu	CCVCV	LI	B	0,86
fapul	CVCVL	LD	A	0,87	pusk	CVsC	LD	B	0,75
fapus	CVCVs	LD	A	0,76	skafu	sCVCV	LD	B	0,82
fli	CCV	LI	A	0,95	skap	sCVC	LD	B	0,88
fuk	CVC	LI	A	0,91	skla	sCCV	LD	B	0,80
kafip	CVCVC	LI	A	0,74	spaf	sCVC	LD	B	0,78
kapi	CVCV	LI	A	0,96	spiku	sCVCV	LD	B	0,89
kifapu	CVCVCV	LI	A	0,76	spli	sCCV	LD	B	0,82
kifus	CVCVs	LD	A	0,87	fikapul	CVCVCVL	LD	C	0,85
kip	CVC	LI	A	0,94	fikupla	CVCVCCV	LI	C	0,72
kis	CVs	LD	A	0,88	fikuspa	CVCVsCV	LD	C	0,71
kla	CCV	LI	A	0,95	flaplu	CCVCCV	LI	C	0,53
kufal	CVCVL	LD	A	0,92	flipuka	CCVCVCV	LI	C	0,62
lafi	LVCV	LI	A	0,94	flukif	CCVCVC	LI	C	0,61
paf	CVC	LI	A	0,96	kapufik	CVCVCVC	LI	C	0,74
pilu	CVCV	LI	A	0,96	kifapus	CVCVCVs	LD	C	0,79
plu	CCV	LI	A	0,91	klifak	CCVCVC	LI	C	0,67
pufaki	CVCVCV	LI	A	0,85	klipafu	CCVCVCV	LI	C	0,67
pukif	CVCVC	LI	A	0,75	klisp	CCVsC	LD	C	0,32
spu	sCV	LD	A	0,88	klups	CCVCs	LD	C	0,65
filpa	CVLVCV	LD	B	0,79	kufalpi	CVCVLVCV	LD	C	0,60
fips	CVCs	LD	B	0,60	kufalpi	CVCCVCV	LI	C	0,74
fiska	CVsCV	LD	B	0,87	kupalfi	CVCVLVCV	LD	C	0,61
flis	CCVs	LD	B	0,81	kupifla	CVCVCCV	LI	C	0,49
fluk	CCVC	LD	B	0,88	pafuski	CVCVsCV	LD	C	0,69
fluka	CCVCV	LI	B	0,87	pakifal	CVCVCVL	LD	C	0,82
fupli	CVCCV	LI	B	0,75	pifakup	CVCVCVC	LI	C	0,68
klaf	CCVC	LD	B	0,87	pifukas	CVCVCVs	LD	C	0,59
klil	CCVL	LD	B	0,84	piklafu	CVCCVCV	LI	C	0,61
klis	CCVs	LD	B	0,78	plaklu	CCVCCV	LI	C	0,62
kusp	CVsC	LD	B	0,64	pliks	CCVCs	LD	C	0,75
kuspa	CVsCV	LD	B	0,76	plusk	CCVsC	LD	C	0,64
paklu	CVCCV	LI	B	0,81	skapufi	sCVCVCV	LD	C	0,67
piks	CVCs	LD	B	0,76	spakifu	sCVCVCV	LD	C	0,71
pilfu	CVLVCV	LD	B	0,65					

A = item de faible complexité

B = item de complexité moyenne

C = item très complexe

ANNEXE 4 : Indices de discrimination des items

Item	Indice discrimination D	Item	Indice discrimination D
faku	0,22	kufalpi	0,83
fal	0,13	kufłapi	0,52
fapul	0,30	kupalfi	0,83
fapus	0,48	kupifla	0,61
fikapul	0,48	kusp	0,61
fikupla	0,39	kuspa	0,43
fikuspa	0,74	łafi	0,17
filpa	0,57	paf	0,13
fips	0,83	pafuski	0,83
fiska	0,39	pakifal	0,39
flaplu	0,61	paklu	0,48
fli	0,13	pifakup	0,78
flipuka	0,74	pifukas	0,48
flis	0,39	pikłafu	0,74
fluk	0,30	piks	0,48
fluka	0,39	pilfu	0,74
flukif	0,70	pilu	0,04
fuk	0,35	plaklu	0,48
fupli	0,43	plal	0,43
kafip	0,48	plifu	0,30
kapi	0,09	pliks	0,52
kapufik	0,57	plu	0,26
kifapu	0,57	plusk	0,57
kifapus	0,35	pufaki	0,57
kifus	0,39	pukif	0,39
kip	0,17	pusk	0,57
kis	0,26	skafu	0,57
kla	0,17	skap	0,43
klaf	0,39	skapufi	0,78
klifak	0,57	skla	0,52
klil	0,52	spaf	0,57
klipafu	0,61	spakifu	0,87
klis	0,39	spiku	0,39
klisp	0,22	spli	0,48
klups	0,78	spu	0,35
kufal	0,22		

ANNEXE 5 : Moyenne et écart-type des items

Item	Moyenne	Ecart-type	Item	Moyenne	Ecart-Type
faku	0,91	0,29	kufalpi	0,60	0,49
fal	0,96	0,19	kufalpi	0,74	0,44
fapul	0,87	0,34	kupalpi	0,61	0,49
fapus	0,76	0,43	kupifla	0,49	0,50
fikapul	0,85	0,36	kusp	0,64	0,48
fikupla	0,72	0,45	kuspa	0,76	0,43
fikuspa	0,71	0,46	lafi	0,94	0,24
filpa	0,79	0,41	paf	0,96	0,19
fips	0,60	0,49	pafuski	0,69	0,46
fiska	0,87	0,34	pakifal	0,82	0,38
flaplu	0,53	0,50	paklu	0,81	0,39
fli	0,95	0,21	pifakup	0,68	0,47
flipuka	0,62	0,49	pifukas	0,59	0,50
flis	0,81	0,39	piklafu	0,61	0,49
fluk	0,88	0,32	piks	0,76	0,43
fluka	0,87	0,34	pilfu	0,65	0,48
flukif	0,61	0,49	pilu	0,96	0,19
fuk	0,91	0,29	plaklu	0,62	0,49
fupli	0,75	0,43	plal	0,85	0,36
kafip	0,74	0,44	plifu	0,86	0,35
kapi	0,96	0,19	pliks	0,75	0,43
kapufik	0,74	0,44	plu	0,91	0,29
kifapu	0,76	0,43	plusk	0,64	0,48
kifapus	0,79	0,41	pufaki	0,85	0,36
kifus	0,87	0,34	pukif	0,75	0,43
kip	0,94	0,24	pusk	0,75	0,43
kis	0,88	0,32	skafu	0,82	0,38
kla	0,95	0,21	skap	0,88	0,32
klaf	0,87	0,34	skapufi	0,67	0,47
klifak	0,67	0,47	skla	0,80	0,40
klil	0,84	0,37	spaf	0,78	0,42
klipafu	0,67	0,47	spakifu	0,71	0,46
klis	0,78	0,42	spiku	0,89	0,31
klisp	0,32	0,47	spli	0,82	0,38
klups	0,65	0,48	spu	0,88	0,32
kufal	0,92	0,28			

ANNEXE 6 : corrélation r item/total des items

Item	r item/total	Item	r item/total
faku	0,48	kufalpi	0,58
fal	0,64	kufalpi	0,56
fapul	0,53	kupalfi	0,64
fapus	0,50	kupifla	0,43
fikapul	0,70	kusp	0,45
fikupla	0,43	kuspa	0,49
fikuspa	0,62	lafi	0,60
filpa	0,61	paf	0,58
fips	0,58	pafuski	0,68
fiska	0,69	pakifal	0,56
flaplu	0,44	paklu	0,51
fli	0,55	pifakup	0,63
flipuka	0,49	pifukas	0,46
flis	0,41	piklafu	0,58
fluk	0,59	piks	0,59
fluka	0,64	pilfu	0,49
flukif	0,50	pilu	0,21
fuk	0,67	plaklu	0,39
fupli	0,46	plal	0,70
kafip	0,45	plifu	0,49
kapi	0,38	pliks	0,54
kapufik	0,56	plu	0,41
kifapu	0,56	plusk	0,48
kifapus	0,44	pufaki	0,78
kifus	0,62	pukif	0,44
kip	0,55	pusk	0,56
kis	0,48	skafu	0,71
kla	0,71	skap	0,77
klaf	0,65	skapufi	0,62
klifak	0,52	skla	0,58
klil	0,73	spaf	0,61
klipafu	0,49	spakifu	0,69
klis	0,40	spiku	0,81
klisp	0,17	spli	0,63
klups	0,59	spu	0,58
kufal	0,65		

ANNEXE 7 : épreuve réduite finale LITMUS-NWR 50

Item	Structure	Type	Complexité	Indice p	Indice D	Ecart-type	r item/total	p : DT-TSL
fapus	CVCVs	LD	A	0,76	0,48	0,43	0,50	0,29
fikuspa	CVCVsCV	LD	C	0,71	0,74	0,46	0,62	0,44
filpa	CVLCV	LD	B	0,79	0,57	0,41	0,61	0,13
fips	CVCs	LD	B	0,60	0,83	0,49	0,58	0,46
flaplu	CCVCCV	LI	C	0,53	0,61	0,50	0,44	0,40
flipuka	CCVCVCV	LI	C	0,62	0,74	0,49	0,49	0,35
flis	CCVs	LD	B	0,81	0,39	0,39	0,41	0,27
flukif	CCVCVC	LI	C	0,61	0,70	0,49	0,50	0,43
fuk	CVC	LI	A	0,91	0,35	0,29	0,67	0,19
fupli	CVCCV	LI	B	0,75	0,43	0,43	0,46	0,32
kafip	CVCVC	LI	A	0,74	0,48	0,44	0,45	0,25
kapufik	CVCVCVC	LI	C	0,74	0,57	0,44	0,56	0,25
kifapu	CVCVCV	LI	A	0,76	0,57	0,43	0,56	0,29
kip	CVC	LI	A	0,94	0,17	0,24	0,55	0,10
kis	CVs	LD	A	0,88	0,26	0,32	0,48	0,20
klifak	CCVCVC	LI	C	0,67	0,57	0,47	0,52	0,33
klil	CCVL	LD	B	0,84	0,52	0,37	0,73	0,36
klipafu	CCVCVCV	LI	C	0,67	0,61	0,47	0,49	0,28
klis	CCVs	LD	B	0,78	0,39	0,42	0,40	0,26
klups	CCVCs	LD	C	0,65	0,78	0,48	0,59	0,44
kufalpi	CVCVLCV	LD	C	0,60	0,83	0,49	0,58	0,37
kuflapi	CVCCVCV	LI	C	0,74	0,52	0,44	0,56	0,30
kupalfi	CVCVLCV	LD	C	0,61	0,83	0,49	0,64	0,48
kupifla	CVCVCCV	LI	C	0,49	0,61	0,50	0,43	0,29
kusp	CVsC	LD	B	0,64	0,61	0,48	0,45	0,13
kuspa	CVsCV	LD	B	0,76	0,43	0,43	0,49	0,34
pafuski	CVCVsCV	LD	C	0,69	0,83	0,46	0,68	0,47
paklu	CVCCV	LI	B	0,81	0,48	0,39	0,51	0,22
pifakup	CVCVCVC	LI	C	0,68	0,78	0,47	0,63	0,40
pifukas	CVCVCVs	LD	C	0,59	0,48	0,50	0,46	0,30
piklafu	CVCCVCV	LI	C	0,61	0,74	0,49	0,58	0,48
piks	CVCs	LD	B	0,76	0,48	0,43	0,59	0,49
piifu	CVLCV	LD	B	0,65	0,74	0,48	0,49	0,39
plaklu	CCVCCV	LI	C	0,62	0,48	0,49	0,39	0,40
plal	CCVL	LD	B	0,85	0,43	0,36	0,70	0,33
plifu	CCVCV	LI	B	0,86	0,30	0,35	0,49	0,11
pliks	CCVCs	LD	C	0,75	0,52	0,43	0,54	0,42
plu	CCV	LI	A	0,91	0,26	0,29	0,41	0,14
plusk	CCVsC	LD	C	0,64	0,57	0,48	0,48	0,28
pufaki	CVCVCV	LI	A	0,85	0,57	0,36	0,78	0,28
pukif	CVCVC	LI	A	0,75	0,39	0,43	0,44	0,17
pusk	CVsC	LD	B	0,75	0,57	0,43	0,56	0,42
skafu	sCVCV	LD	B	0,82	0,57	0,38	0,71	0,39
skap	sCVC	LD	B	0,88	0,43	0,32	0,77	0,29
skapufi	sCVCVCV	LD	C	0,67	0,78	0,47	0,62	0,48
skla	sCCV	LD	B	0,80	0,52	0,40	0,58	0,30
spaf	sCVC	LD	B	0,78	0,57	0,42	0,61	0,21
spakifu	sCVCVCV	LD	C	0,71	0,87	0,46	0,69	0,54
spli	sCCV	LD	B	0,82	0,48	0,38	0,63	0,19
spu	sCV	LD	A	0,88	0,35	0,32	0,58	0,24

Réduction d'une épreuve de répétition de non-mots à destination des enfants bilingues

Le manque d'outil spécifique d'évaluation du langage dans un contexte d'acquisition bilingue a fait naître un nouveau champ de recherche linguistique : l'étude des enfants bilingues avec trouble spécifique du langage (TSL) et la recherche des marqueurs diagnostiques de ces troubles. Il a été montré qu'une faible performance dans une tâche de répétition de non-mot pouvait être un marqueur de TSL. Dans ce cadre, l'épreuve expérimentale LITMUS-NWR a été créée pour évaluer la phonologie. Elle comporte actuellement 71 items. Une partie des items a été élaborée en prenant en compte des contraintes phonologiques répertoriées dans de nombreuses langues du monde ; l'autre partie teste certains points de complexité phonologique du français. Le caractère discriminant de cette épreuve ayant été vérifié, il s'agit d'en proposer ici une version écourtée, plus facilement utilisable en clinique.

Nous avons donc procédé à une analyse d'items, à partir des données provenant des résultats à ce test d'une population de 85 enfants âgés de 5;2 à 8;10 ans, monolingues et bilingues, au développement typique et avec TSL. Après calcul des différents indices statistiques (difficulté, discrimination, corrélation item-total, écart-type), les résultats de l'analyse d'items ont permis de procéder à l'élimination des items les moins discriminants. Le degré d'homogénéité de l'épreuve ainsi réduite a ensuite été estimé au moyen du coefficient alpha de Cronbach.

Cette étape de réduction amène des arguments empiriques à la théorie qui a permis l'élaboration de l'épreuve : elle met en évidence le rôle de la complexité syllabique dans les productions d'enfants avec TSL. Elle permet ainsi à la fois d'avancer dans l'étude de la nature de leur déficit phonologique et de proposer aux orthophonistes un outil adapté aux besoins des populations bilingues.

Mots-clés : Test, Répétition de non-mots, Bilinguisme, Complexité phonologique, Analyse d'items, Trouble Spécifique du Langage.