



Université Paris 8 Vincennes – Saint-Denis
Ecole Doctorale 224 Cognition Langage Interaction

Thèse présentée en vue de l'obtention du grade de Docteur en Psychologie

Evaluer la lecture électronique : une approche multidimensionnelle

Présentée par
Jean-Luc PERRIN

Dirigée par
Thierry BACCINO,
Laboratoire Cognitions Humaine et Artificielle, EA 4004

Responsable en entreprise :
Damien PAILLÉ,
Essilor International Recherche & Développement Optique, Sciences de la Vision

Membres du jury :

M. Yann COELLO	Université Lille 3	Rapporteur
Mme. Olga MEGALAKAKI	Université de Picardie	Rapporteuse
M. Charles TIJUS	Université Paris 8	Examineur
M. Jean-Luc VELAY	Université Aix-Marseille	Examineur
M. Thierry BACCINO	Université Paris 8	Directeur
M. Damien PAILLÉ	Essilor International	Responsable

Soutenue le 1^{er} septembre 2015

Remerciements

Je tiens tout d'abord à remercier les deux personnes qui ont su habilement guider cette thèse et m'offrir d'excellentes conditions pour parvenir à réaliser ce travail. Merci à Thierry Baccino de m'avoir donné l'occasion de découvrir ce domaine de recherche, en me prodiguant de nombreux conseils mais aussi en me permettant d'ajouter ma propre pierre à l'édifice. Merci à Damien Paillé pour son soutien continu, sa disponibilité ainsi que sa bienveillance : ces qualités m'ont offert un cadre idéal pour pouvoir avancer sereinement.

Merci à Olga Megalakaki et à Yann Coello d'avoir accepté de réviser ces travaux, et à Jean-Luc Velay et Charles Tijus de faire parti du jury.

Je remercie l'entreprise Essilor International et l'Association Nationale de la Recherche et de la Technologie d'avoir mis à disposition les moyens matériels pour la réalisation de ce projet.

Mes remerciements vont également aux collaborateurs des équipes R&D Essilor, pour les nombreuses aides apportées sur le plan technique et pour leur convivialité et leur curiosité.

Merci aux collègues du laboratoire Cognitions Humaine et Artificielle et de la plateforme LUTIN pour leur accueil, ainsi qu'aux membres de l'école doctorale Cognition Langage et Interaction.

Je voudrais remercier celles et ceux qui ont donné de leur temps pour participer aux expérimentations.

Pour leur soutien, merci à Asma et merci aux amis de Nancy (au sens large) et de Paris. Enfin, un grand merci à toute ma famille : mes parents, mon frère, mes tantes et oncles, mes cousines et cousins et ma grand-mère.

Une pensée pour mes grands-parents Alice et Jean-Pierre, ainsi que pour mon oncle André.

Digital reading assessment: a multidimensional approach

The aim of this thesis is to offer methods of evaluating performance during digital reading, in order to be able to compare different display devices. Tests were developed allowing repeated measurements for classic measures of digital reading: visibility (psychometric characterization), readability (letter identification, lexical decision, reading speed on automatically generated sentences) and comprehension (memorization and inference measurement while reading automatically generated texts). Beyond these classic dimensions, we also conducted two studies in order to examine the link between digital reading and the posture adopted by the device users. The first revealed that a tablet-use-related posture (lateral head tilt) does not affect reading performance. The second postural study demonstrated the effect of text difficulty on eye-screen distance during the use of an e-reader. Subjects approached the screen as their attentional level increased. The constructed tests and a postural measure were aggregated into a battery of tests permitting the evaluation of reading performance associated with a device. The resulting indicator is called the “Digital Reading Quotient”. This indicator has been computed for a computer screen, a projector and a tablet. The experimental results show that these tests can be used in order to compare different devices, especially in terms of visibility and readability.

Keywords: Digital reading ; Assessment ; Psychophysics ; Performance ; Visibility ; Readability ; Comprehension

The work which is described in this thesis have been achieved in the laboratory Cognitions Humaine et Artificielle (CHArt EA 4004, Université Paris 8, 2 rue de la liberté, 93526 Saint-Denis, France) and in the R&D Optics Departement of Essilor International (13 rue Moreau, 75012 Paris, France).

Evaluer la lecture électronique : une approche multidimensionnelle

Le but de cette thèse est de proposer des méthodes pour pouvoir évaluer les performances durant la lecture sur support électronique, afin de pouvoir comparer plusieurs dispositifs d'affichage. Des tests permettant d'effectuer des mesures de performances de manière répétée ont été construits pour chacune des dimensions classiques de la lecture électronique : visibilité (test de caractérisation psychométrique), lisibilité (identification de lettres, décision lexicale, vitesse de lecture de phrases générées automatiquement) et compréhension (mesure de la mémorisation et des inférences pendant la lecture de textes générés automatiquement). Au-delà de ces dimensions classiques, nous avons réalisé deux études pour examiner le lien entre la lecture sur support électronique et la posture des utilisateurs. La première montre qu'une posture pouvant être engendrée par l'utilisation de tablette (inclinaison latérale de tête) n'affecte pas la performance de lecture. La seconde étude posturale montre l'effet de la difficulté d'un texte sur la distance à l'écran pendant l'utilisation d'une liseuse électronique ; les sujets se rapprochent de l'écran lorsque leur niveau d'attention augmente. L'ensemble des tests développés et une mesure de posture ont été regroupés dans une batterie de tests visant à quantifier la performance de lecture associée à un support donné ; l'indicateur résultant étant appelé « Quotient de Lecture Electronique ». Cet indicateur a été calculé pour un écran d'ordinateur, un vidéoprojecteur et une tablette. Les résultats expérimentaux montrent que les tests permettent de comparer les supports, notamment en termes de visibilité et de lisibilité.

Mots-clés : Lecture électronique ; Evaluation ; Psychophysique ; Performance ; Visibilité ; Lisibilité ; Compréhension

Les travaux décrits dans cette thèse ont été réalisés au sein du laboratoire Cognitions Humaine et Artificielle (CHArt EA 4004, Université Paris 8, 2 rue de la liberté, 93526 Saint-Denis) et du département R&D Optique de Essilor International (13 rue Moreau, 75012 Paris).

Table des matières

Chapitre 1 : Introduction	10
Chapitre 2 : Etat de l'art	13
2.1 Les dimensions classiques de la lecture électronique	13
2.1.1 La visibilité.....	13
2.1.1.1 Les dispositifs d'affichage électronique et la lumière.....	14
2.1.1.2 Le système visuel humain	19
2.1.1.3 L'œil et l'écran	22
2.1.2 La lisibilité.....	24
2.1.2.1 Les constituants du texte et leur affichage électronique.....	25
2.1.2.1.1 Les lettres.....	25
2.1.2.1.2 Les mots.....	27
2.1.2.1.3 Les phrases	28
2.1.2.2 Bases neurophysiologiques et psycholinguistiques de la lecture	29
2.1.2.2.1 Le traitement cérébral des lettres et mots	30
2.1.2.2.2 Le parcours des yeux et la lecture des phrases	32
2.1.2.3 La lisibilité des écrans	34
2.1.2.4 La mesure de la lisibilité des dispositifs d'affichage.....	36
2.1.3 La compréhension	39
2.1.3.1 Le texte : sémantique et affichage électronique	39
2.1.3.1.1 Aspects sémantiques des textes	39
2.1.3.1.2 Affichage électronique des textes : l'interface	42
2.1.3.2 La compréhension en lecture	44
2.1.3.3 Comment les interfaces agissent sur la compréhension	47
2.1.3.4 Evaluation de la compréhension.....	49
2.1.3.4.1 Les mesures du produit de la lecture	49
2.1.3.4.2 Les mesures concomitantes à la lecture.....	51
2.2 La dimension posturale de la lecture électronique	54
2.2.1 La posture	55
2.2.2 La posture pendant l'utilisation d'appareils électroniques.....	58
2.2.2.1 Caractéristiques des dispositifs d'affichage électronique.....	58
2.2.2.2 Approche de l'ergonomie physique.....	59
2.2.2.3 Posture et dimensions visuelle et cognitive de la lecture électronique	60
2.2.2.3.1 Posture et vision	60
2.2.2.3.2 Posture et aspects cognitifs.....	61
Chapitre 3 : Problématique.....	64
Chapitre 4 : L'évaluation de la visibilité	67
4.1 Les méthodologies d'évaluation de la visibilité.....	67
4.2 Technique de caractérisation psychométrique	69
Chapitre 5 : L'évaluation de la lisibilité	74
5.1 Test d'identification de lettres et test de décision lexicale.....	74
5.1.1 Sélection des stimuli.....	74
5.1.1.1 Stimuli pour le test de d'identification de lettres.....	74
5.1.1.2 Stimuli pour le test de décision lexicale	78
5.1.2 Procédure.....	80
5.1.3 Mesure du temps de réponse de base	81
5.1.4 Traitement des données	81

5.2	La mesure de la vitesse de lecture maximale	82
5.2.1	Introduction	83
5.2.2	Principes de fonctionnement du générateur de phrases.....	86
5.2.2.1	Gestion des valeurs de vérité	86
5.2.2.2	Choix des objets	89
5.2.2.3	Génération des phrases et lecture	90
5.2.3	Validation des phrases générées dans l'évaluation de la vitesse de lecture	91
5.2.3.1	Protocole.....	91
5.2.3.1.1	Participants	91
5.2.3.1.2	Appareillage	92
5.2.3.1.3	Tests de lecture	92
5.2.3.1.4	Procédure et mesures	92
5.2.3.1.5	Analyse des données.....	94
5.2.3.2	Résultats	95
5.2.4	Discussion / conclusion	99
Chapitre 6 :	Chapitre 6 : L'évaluation de la compréhension.....	103
6.1	Objectifs du test de compréhension	103
6.2	Un générateur automatique de textes pour l'évaluation de la compréhension	104
6.2.1	Concepts manipulés par le générateur.....	104
6.2.2	Attribution de caractéristiques au concept	106
6.2.3	Construction de traits généraux liés au concept	107
6.2.4	Construction des textes.....	109
6.3	Attributions des valeurs aux textes et aux questionnaires.....	111
6.3.1	Méthodes possibles de calcul des valeurs finales.....	111
6.3.2	Principe de calcul des valeurs finales par réseaux bayésiens.....	114
6.3.3	Recueil d'observations pour le corpus d'apprentissage	115
6.3.4	Construction des réseaux bayésiens	118
6.3.5	Apprentissage des réseaux bayésiens	119
6.3.6	Vérification de la fiabilité.....	120
6.3.7	Sélections des textes.....	121
6.4	Procédure.....	122
6.5	Traitement des données	123
6.6	Tâche « Remember-Know »	123
6.7	Mesure de la charge attentionnelle.....	125
Chapitre 7 :	Aspects posturaux de la lecture électronique	126
7.1	L'effet de la posture sur la lecture électronique	126
7.1.1	Introduction	127
7.1.2	Expérience 1	129
7.1.2.1	Méthode.....	129
7.1.2.1.1	Participants	129
7.1.2.1.2	Appareillage	129
7.1.2.1.3	Stimuli	131
7.1.2.1.4	Plan expérimental et procédure	131
7.1.2.1.5	Analyse des données.....	132
7.1.2.2	Résultats	133
7.1.2.3	Discussion.....	134
7.1.3	Expérience 2.....	135
7.1.3.1	Méthode.....	136

7.1.3.1.1	Participants, appareillage et stimulus	136
7.1.3.1.2	Plan expérimental et procédure	136
7.1.3.1.3	Analyse des données.....	136
7.1.3.2	Résultats	138
7.1.3.3	Discussion.....	141
7.1.4	Discussion générale et conclusion.....	142
7.2	L'effet de la lecture électronique sur la posture	145
7.2.1	Introduction	145
7.2.2	Méthode.....	147
7.2.2.1	Participants	147
7.2.2.2	Stimuli	148
7.2.2.3	Appareillage	149
7.2.2.4	Mesures.....	150
7.2.2.4.1	Indicateurs de la charge attentionnelle	150
7.2.2.4.2	Indicateur postural	151
7.2.2.5	Plan expérimental et procédure	151
7.2.3	Analyse des données et résultats	152
7.2.3.1	Données manquantes	152
7.2.3.2	Effet de la difficulté.....	153
7.2.3.3	Liens avec la charge attentionnelle.....	154
7.2.4	Discussion	156
Chapitre 8 :	Le quotient de lecture électronique	159
8.1	Les tests de la batterie du Quotient de Lecture Electronique.....	159
8.2	Matériel et sujets	160
8.2.1	Dispositifs d'affichage électroniques	160
8.2.2	Salle d'expérimentation.....	161
8.2.3	Taille des stimuli	161
8.2.4	Interfaces de réponse	162
8.2.5	Matériel de mesure	162
8.2.6	Participants	163
8.3	Recueil des données	164
8.3.1	Procédure générale	164
8.3.2	Procédures associées aux tests	164
8.3.2.1	Test de visibilité	164
8.3.2.2	Tests de lisibilité.....	165
8.3.2.3	Test de compréhension	167
8.4	Tests d'hypothèses	169
8.4.1	Visibilité	169
8.4.1.1	Test de caractérisation subjective	169
8.4.1.2	Mesure des profils de luminance des écrans	170
8.4.2	Lisibilité	171
8.4.2.1	Temps de réponse de base	171
8.4.2.2	Test d'identification de lettres	172
8.4.2.3	Test de décision lexicale.....	176
8.4.2.4	Durée réelle d'affichage des phrases	180
8.4.2.5	Vitesse de lecture.....	181
8.4.3	Compréhension et posture	181
8.4.3.1	Test de compréhension	181
8.4.3.1.1	Durées de lecture	181
8.4.3.1.2	Erreurs de mémorisation.....	183

8.4.3.1.3	Erreurs d'inférences.....	184
8.4.3.1.4	Durées de réponse.....	186
8.4.3.2	Posture	187
8.4.3.2.1	Distance moyenne à l'écran.....	187
8.4.3.2.2	Ecart-type de la distance à l'écran.....	189
8.4.3.3	Tâche R-K	191
8.5	Modèle de calcul du QLE.....	192
8.5.1	Calcul des indicateurs pour les tests.....	192
8.6	Résultats et implications.....	196
8.6.1	Visibilité	196
8.6.2	Lisibilité	197
8.6.2.1	Temps de réponse de base	197
8.6.2.2	Identification de lettres	197
8.6.2.3	Décision lexicale	198
8.6.2.4	Vitesse de lecture.....	198
8.6.3	Compréhension et posture.....	199
8.6.3.1	Compréhension.....	199
8.6.3.1.1	Facteurs expérimentaux propres au test	199
8.6.3.1.2	Effets des supports.....	199
8.6.3.2	Posture	200
8.6.3.2.1	Facteurs expérimentaux propres au test	200
8.6.3.2.2	Effets des supports.....	200
8.6.3.3	Tâche R-K	201
8.6.4	Quotient de lecture	201
Chapitre 9 :	Discussion générale - Conclusion.....	202
9.1	Tester la lecture électronique : dimensions classiques et posture	202
9.2	Une approche multidimensionnelle : la batterie de tests du Quotient de	
Lecture Electronique		205
9.3	Limites et ouvertures	206
Liste des références bibliographiques		210
Résumé long.....		223

Chapitre 1 : Introduction

L'avènement de l'ordinateur a marqué l'Histoire, offrant à l'humanité des capacités de calcul jusqu'alors inégalées. Peu à peu, l'ordinateur a également fourni la possibilité de stocker d'immenses quantités d'informations. En manipulant des volumes importants de données symboliques, les ordinateurs ont permis d'automatiser des tâches qui pouvaient se révéler fastidieuses pour l'être humain.

Néanmoins, l'ordinateur et ses capacités de calcul et de stockage ne seraient d'aucune utilité s'ils ne disposaient pas de moyens permettant de traduire les informations électroniques en un langage compréhensible par l'opérateur. Si les premiers ordinateurs acheminaient l'information en l'imprimant sur des rouleaux de papier (par exemple, sur l'ENIAC en 1946), le moyen privilégié de transmission est rapidement devenu l'écran (comme sur le SSEM en 1948).

Les technologies n'ont cessé d'évoluer, et offrent aujourd'hui une multitude d'écrans et d'autres dispositifs d'affichage électronique afin que nous puissions prendre connaissance des informations stockées et générées par les ordinateurs. Leurs formats sont nombreux, ils varient en tailles et en fonctionnalités : du désormais classique écran de bureau, à la tablette électronique, en passant par les vidéoprojecteurs ou les liseuses électroniques, ces dispositifs sont présents sous diverses formes dans notre environnement.

L'arrivée massive des ordinateurs dans un contexte professionnel a rapidement conduit les psychologues et les ergonomes à questionner l'efficacité de la lecture sur support numérique en comparaison au papier (pour une revue de la littérature sur cette question, se référer à (Dillon, 1992)). Ces nouveaux dispositifs d'affichage électroniques avaient des caractéristiques propres, qui pouvaient perturber la lecture, ce qui se traduisait éventuellement par des baisses de productivité dans l'entreprise. A ces premiers questionnements, s'ajoute le fait que, depuis, les objets numériques ne se cantonnent plus au bureau, mais sont également devenu un compagnon dans nos activités quotidiennes ainsi que dans nos loisirs¹. Nous lisons donc quotidiennement des informations fournies par des dispositifs d'affichage électronique.

¹ Des études sur les usages des technologies telles que les rapports IPSOS Technology Trackers montrent la progression des smartphones, tablettes et l'implantation des ordinateurs dans le contexte extra-professionnel.

Ce sont ces deux constats – la multiplicité des dispositifs et leur omniprésence – qui motivent le projet de cette thèse. Nous poserons ici la question suivante : comment évaluer la lecture sur support électronique ? Mais avant de pouvoir répondre à cette question, il est nécessaire de définir quelques concepts. Cette question repose sur la notion de « lecture électronique », qui est liée à deux aspects : d'un côté le processus de lecture et de l'autre les dispositifs d'affichage électroniques.

L'activité de lecture peut se définir comme étant le processus répété d'extraction de l'information visuelle représentée par des caractères, aboutissant à la compréhension par le lecteur des connaissances et éventuellement des enjeux soulevés par le texte. La lecture électronique correspond à la réalisation de l'activité de lecture lorsque l'information est affichée au moyen d'un dispositif électronique. Nous considérerons ici la notion de dispositif d'affichage électronique au sens large : il s'agit de tout appareil alimenté électriquement qui émet ou module la lumière de manière à traduire visuellement l'état interne d'un système électronique, notamment sous forme textuelle.

La lecture électronique est un processus complexe, impliquant des fonctions de bas niveau telles que la détection de stimulations lumineuses, et des fonctions de plus haut niveau telles que la construction de représentations mentales du contenu d'un texte. Cette multi-dimensionnalité de l'objet de cette thèse nous a poussé à prendre comme cadre de recherche le découpage en trois dimensions de la lecture électronique (Baccino, 2004). Un dispositif de lecture électronique peut être évalué en termes de visibilité, de lisibilité et de compréhension. Ces dimensions seront détaillées dans la section *2.1 Les dimensions classiques de la lecture électronique*. Néanmoins, partant de l'observation des formes multiples prises par les supports électroniques, cette thèse cherchera à y ajouter une dimension posturale. En effet, au-delà de prendre des formes qui vont stimuler différemment notre vision, de fournir des contenus variables qui vont être traités par notre cognition, les supports vont nous faire adopter des postures variées selon leurs caractéristiques physiques.

Le but des travaux présentés dans cette thèse est de proposer une méthodologie d'évaluation de la performance de lecture, qui tient compte de ces quatre dimensions. Ce projet nécessite de définir les éléments pertinents à observer pour évaluer la qualité de lecture sur un support donné, ainsi que de construire des tests permettant de réaliser ces mesures. Les données qui seront recueillies à partir de ces tests devront être synthétisées pour fournir des indicateurs sur la performance de lecture associée à un support. Nos travaux portent surtout

sur la performance, les aspects liés à la lecture sur le moyen ou long-terme, tels que la fatigue visuelle ou l'effet de l'usage des technologies sur le comportement et la personnalité, n'y seront pas abordés.

Nous présenterons dans un premier temps un état de l'art sur les dimensions classiques de la lecture électronique et sur la dimension posturale associée. Un chapitre portant sur la problématique permettra ensuite d'identifier les questions à traiter pour pouvoir parvenir à réaliser une évaluation de la lecture électronique. Les trois chapitres suivants présenteront les tests que nous avons construits pour évaluer la visibilité, la lisibilité et la compréhension associées à un support donné. Le septième chapitre décrira deux expérimentations que nous avons réalisées pour mieux cerner le lien entre la posture et la lecture électronique. Nous présenterons ensuite la réalisation concourante de tous les tests construits, et le modèle d'analyse permettant de dégager un indicateur de la performance : le « quotient de lecture électronique ». La thèse se terminera sur une discussion et une conclusion générales.

Chapitre 2 : Etat de l'art

Ce chapitre vise à présenter les dimensions de la lecture électronique. Pour chacune de ces dimensions, nous exposerons les processus psychologiques mis en œuvre par le lecteur, les aspects techniques liés aux dispositifs d'affichage et l'interaction entre le lecteur et le support. Nous présenterons également les mesures qui peuvent être réalisées pour rendre compte de la performance de lecture. Dans une première section seront abordées les trois dimensions classiques de la lecture électronique, puis dans la seconde nous traiterons des aspects posturaux liés à l'utilisation des supports.

2.1 Les dimensions classiques de la lecture électronique

Dans cette première section nous abordons les trois dimensions classiques de la lecture électronique (Baccino, 2004) : à savoir la visibilité, la lisibilité et la compréhension.

2.1.1 La visibilité

Les écrans, et plus généralement les dispositifs d'affichage électronique sont au cœur des interactions homme/machine. La plupart des systèmes informatiques rendent compte de leur état interne à travers un ou plusieurs dispositifs d'affichage électronique. Ces dispositifs transmettent une information visuelle qui peut varier dans le temps. Nous les considérerons donc comme les moyens de transcrire visuellement des informations produites et stockées sous une forme et/ou à une échelle qui ne sont pas accessible à l'utilisateur : les données. Une des manières les plus courantes de traduire ces données est de les afficher sous forme de symboles et de caractères. Si ce n'est pas la fonction unique des dispositifs d'affichage –ils peuvent également afficher des images, des vidéos...–, les textes sont omniprésents sur les écrans, et se pose la question de la lecture électronique, ainsi que de la qualité de l'affichage.

Un dispositif d'affichage émet un signal lumineux qui doit être perçu par l'utilisateur, afin que celui-ci puisse disposer de l'information contenue dans le support électronique. Les qualités du dispositif vont permettre à l'utilisateur de détecter ces signaux avec une efficacité plus ou moins grande. Cette transmission de signaux dépend du dispositif, mais également des caractéristiques propres au système visuel de l'utilisateur. La visibilité consiste en cette capacité du dispositif à transmettre une information visuelle avec efficacité. Dans cette partie nous allons présenter les différents types d'affichages ainsi que leurs caractéristiques, le traitement de l'information lumineuse par le système visuel humain, et enfin l'interaction entre l'écran et ce système visuel.

2.1.1.1 Les dispositifs d'affichage électronique et la lumière

Les dispositifs d'affichage remplissent leur fonction (à savoir la traduction des données en signal visuel perceptible) en émettant ou en modulant la lumière de façon à générer un espace où la valeur de luminance (absolue ou relative) et la couleur de chaque point est contrôlée. Les dispositifs ne construisent pas un espace continu, mais un espace discrétisé, on parlera alors de « pixel » pour définir la surface correspondant à chaque point de l'affichage. Pour les dispositifs d'affichage en couleur, les pixels sont généralement composés de trois sous-pixels : le rouge, le vert et le bleu, dont l'activation simultanée à différents niveaux va permettre d'attribuer la couleur requise au pixel.

Différentes technologies sont mises en œuvre pour y parvenir (Thomas, 2009). Les écrans émissifs émettent la lumière qui va former l'affichage demandé. Dans le cas des écrans cathodiques (couramment abrégés CRT, de l'anglais *Cathode Ray Tube*), des phosphores convertissent l'énergie électrique en lumière. Dans le cas des écrans à plasma, c'est un mélange de gaz qui est stimulé électriquement pour émettre de la lumière. Pour les écrans émissifs basés sur la technologie LED (diodes électroluminescentes ou *Light-Emitting Diodes*), des diodes sont activées individuellement de manière à former l'image. Les dénominations de ces écrans varient avec le type de LED utilisé (par exemple, OLED pour *Organic Light-Emitting Diode*).

D'autres écrans, dits non émissifs, viennent moduler une lumière (de l'environnement, ou produite par le dispositif électrique) afin d'attribuer à chaque pixel la valeur de luminance et de couleur demandée. Les écrans qui produisent une lumière et qui la modulent pour obtenir l'affichage souhaité sont dits transmissifs (ou rétro-éclairés). Le rétro-éclairage de l'écran est réalisé par une source lumineuse : des tubes lumineux, ou alors des LED placées sur les côtés (*edge LED*) ou sous forme de matrice (*full LED*). Certains écrans ne produisent pas de lumière, et utilisent la lumière incidente de l'environnement (la lumière du soleil ou de l'éclairage se reflète sur un miroir) ; on parle alors d'écran « réfléchif », c'est par exemple le cas des calculatrices où l'on ne peut lire ce qui est affiché à l'écran que lorsque l'ambiance lumineuse est suffisante. Une fois la lumière produite ou réfléchie, elle doit traverser un milieu qui va moduler la lumière à chaque pixel pour obtenir l'affichage désiré. La technologie la plus utilisée est celle des cristaux liquides (pour écrans LCD : *Liquid Cristal Displays*). Des plaques successives de matériaux polarisent la lumière, des cristaux liquides sont orientés de manière à laisser passer plus ou moins cette lumière polarisée. Plusieurs

technologies de LCD sont employées et peuvent avoir un impact sur la fluidité de l'affichage, ainsi que sur les angles de vision acceptables. Une alternative au LCD, qui doit consommer de l'électricité pour maintenir l'état des cristaux liquides, est proposée par l'encre électronique : des microcapsules pouvant prendre au moins deux états (noir ou blanc), font varier leurs états respectifs de manière à obtenir l'affichage désiré. La technologie de l'encre électronique est utilisée dans les liseuses (ou en anglais, *E-Readers*).

Les écrans diffèrent également par le traitement optique de leur face extérieure, qui peut être ainsi rendue brillante ou mate.

Les dispositifs d'affichage numérique ne sont pas restreints aux écrans, on peut noter l'utilisation des vidéoprojecteurs. Le principe de fonctionnement est similaire à celui des écrans transmissifs, particulièrement dans le cas des projecteurs LCD. Une autre technologie employée est le DLP (*Digital Light Processing*, traitement digital de la lumière), pour laquelle des faisceaux de lumière sont envoyés sur des matrices de miroirs qui renvoient ou non la lumière selon l'affichage nécessaire.

L'éventail des technologies est large, et répond à des besoins spécifiques. Ces dernières permettent d'obtenir un nombre plus ou moins important de pixels. Cette quantité est représentée par la « définition » de l'écran, c'est-à-dire le nombre de colonnes de pixels multiplié par le nombre de lignes de pixels. Certaines paires de ces valeurs numériques sont désignées par des acronymes (640 colonnes * 480 lignes correspond à VGA, 800 * 600 à SVGA, 1920 * 1080 à Full HD...). Le produit des deux valeurs donne le nombre total de pixels à l'écran. Le quotient des deux valeurs donne le ratio de l'écran (les plus courants étant le 4:3, le 16:9 et 16:10). Puisque les écrans ont des tailles différentes (souvent exprimées par la longueur de leur diagonale, en pouces), la définition de l'écran n'est pas suffisante pour juger de la qualité d'un écran. En effet, pour deux écrans ayant la même définition, mais des tailles différentes, la taille d'un pixel ne sera pas la même. On emploie donc également la « résolution » de l'écran ; la résolution s'obtient en divisant le nombre de pixels par la longueur physique associée (on peut par exemple diviser le nombre de colonnes de pixels de l'écran par sa largeur en cm). L'unité de longueur physique la plus employée dans cette configuration est le pouce, on parle alors de « pixels par pouces », abrégé « ppi » (de l'anglais *Pixel Per Inch*). La résolution ainsi que la définition de l'écran nous renseignent sur la qualité spatiale d'affichage de l'écran, sur sa capacité à afficher des détails plus ou moins fins. Outre le nombre de pixels, on pourra aussi s'intéresser à leur surface relative à la taille de l'écran,

ainsi qu'à l'espace qui sépare deux pixels (« espace interpixel »). En effet, les pixels ne sont pas strictement accolés les uns aux autres : on peut s'en rendre compte en regardant de très près l'image affichée par un vidéoprojecteur. Néanmoins, cet espace étant majoritairement trop petit pour être discerné dans des conditions d'utilisation habituelles, nous ne l'étudierons pas dans le contexte de cette thèse.

Les différentes technologies ont aussi un impact sur les valeurs de luminance des pixels. Pour chaque écran, on peut mesurer un contraste maximum. Le contraste peut être calculé de plusieurs manières, en utilisant les valeurs de luminance minimale et maximale pouvant être affichées par l'écran (typiquement, la luminance de l'écran lorsqu'il affiche un blanc uni et lorsqu'il affiche un noir uni). Les formules de calcul du contraste sont les suivantes, la lettre L représentant la luminance, L_{max} la luminance du point le plus lumineux, L_{min} la luminance du point le moins lumineux, L_{fond} la luminance de l'arrière-plan et L_{objet} la luminance de l'objet présent devant l'arrière-plan (par exemple, un mot noir sur un fond blanc) :

- Le ratio de luminance :

$$\frac{L_{max}}{L_{min}}$$

Il s'agit de la mesure la plus directe du contraste. Lorsque l'écran affiche une couleur unie (qu'il n'y a pas de contrastes), le ratio s'élève à 1. Lorsque l'on affiche deux zones ayant des luminances différentes, le ratio augmente linéairement avec la luminance de la partie la plus lumineuse.

- Le contraste de Weber :

$$\frac{(L_{fond} - L_{objet})}{L_{fond}}$$

Le contraste de Weber est utilisé dans le cas où l'on peut définir une partie de l'affichage comme étant le fond, et une autre comme étant un objet distinct. Quand le fond et l'objet sont confondus (pas de contrastes), la valeur de contraste est de 0. La valeur est positive lorsque la luminance de fond est plus élevée que celle de l'objet (par exemple, un texte noir sur fond blanc, on parle de « polarité positive »), elle est négative quand la luminance de fond est inférieure à celle de l'objet (texte blanc sur fond noir, « polarité

négative »). Plus l'écart entre les deux valeurs de luminance est important, plus la valeur du contraste est proche de 1, ou de -1, selon la polarité.

- Le contraste de Michelson (Michelson, 1927) :

$$\frac{(L_{max} - L_{min})}{(L_{max} + L_{min})}$$

Cette mesure du contraste est utilisée lorsque l'on ne peut pas déterminer clairement dans l'affichage quelle partie correspond au fond et quelle partie correspond à l'objet (par exemple, une texture sinusoïdale). Lorsque l'affichage ne contient pas de forme contrastée, la valeur de contraste est de 0. Plus le contraste est fort, plus la valeur de contraste de Michelson est proche de 1.

Il existe d'autres manières de calculer le contraste d'une image, notamment lorsque l'image est complexe (Peli, 1990). Néanmoins, comme l'objet de cette thèse porte essentiellement sur l'affichage de textes, seules les trois formules détaillées précédemment seront utilisées.

La valeur de contraste n'est pas suffisante pour renseigner la qualité du dispositif d'affichage, dans la mesure où cette valeur est fonction de la luminosité minimale et de la luminosité maximale ; on peut augmenter le contraste de l'écran en diminuant la luminosité de l'écran lorsqu'il affiche du noir (luminosité résiduelle). Mais le contraste peut aussi être accru en augmentant par exemple le rétro-éclairage de l'écran, et en augmentant ainsi la luminosité associée à un affichage blanc. Deux écrans peuvent donc avoir un contraste mesuré équivalent, mais avec des valeurs de luminance différentes (e.g. un affichage blanc peu lumineux sur un fond très noir versus un affichage blanc très lumineux sur un fond grisâtre). Il est donc important de connaître à la fois la valeur du contraste, et celle de la luminance minimale (ou maximale) pour pouvoir juger de la qualité d'un dispositif d'affichage.

Par extension, il peut être pertinent de connaître la relation entre chaque niveau d'entrée et la luminance de sortie : cette évaluation est appelée « caractérisation » de l'affichage (Ban & Yamamoto, 2013; Gille, Arend, & Larimer, 2004). La fonction associant la valeur d'entrée (en volts, ou en code numérique) à la valeur de luminance de sortie est appelée « fonction de transfert ». Le niveau d'entrée est généralement décrit par un triplet RGB (*Red Green Blue*), qui correspond aux valeurs d'intensité des couleurs rouge, verte et

bleue. Le niveau (0, 0, 0) correspond à la valeur minimale d'intensité de l'écran (fond noir) et le niveau (255, 255, 255) à la valeur maximale d'intensité (fond blanc). On réalise généralement la caractérisation d'un écran en affichant chaque niveau RGB à l'écran et en mesurant la valeur de luminance correspondante à l'aide d'un luminancemètre (ou d'un colorimètre si l'on veut connaître les valeurs associées au rendu des couleurs de l'écran). La caractérisation des écrans est critique du fait de la non-linéarité de la sortie lumineuse des écrans, en particulier pour les écrans cathodiques. La valeur de luminance émise par le CRT n'est pas proportionnelle au voltage reçu en entrée, mais suit plutôt une fonction puissance de la forme $f_a : x \rightarrow x^a$, avec x la valeur de l'entrée exprimée en taux du voltage d'entrée maximal (compris entre 0 et 1), a une constante généralement comprise entre 2,2 et 2,5, et $f_a(x)$ la valeur de sortie exprimée en taux de luminance (0 correspondant à la luminance de l'écran lorsqu'il est noir, L_{min} , et 1 correspondant à la luminance maximale de l'écran, L_{max}). Cette relation peut être appliquée aux variables exprimées dans leurs unités de mesure en introduisant 3 constantes : les valeurs de luminance minimale et maximale en cd/m^2 (respectivement L_{min} et L_{max}) et le voltage maximal d'entrée V_{max} exprimé en mV, la relation s'écrit alors $f_a : x \rightarrow (x / V_{max})^a * (L_{max} - L_{min}) + L_{min}$. Dans les faits, le même écart de voltage produira un changement faible de luminance dans les couleurs foncées alors qu'il provoquera un grand changement de luminance dans les couleurs claires. Il est donc nécessaire d'appliquer une correction sur la valeur d'entrée pour compenser la non-linéarité et obtenir une fonction linéaire de transfert. Cette correction est appelée « correction gamma » pour les écrans cathodiques : on applique au voltage d'entrée une fonction $f_\gamma : x \rightarrow x^\gamma$ avec γ l'inverse du paramètre a (donc entre 1/2,5 et 1/2,2). Pour les types d'écrans actuels, la fonction de transfert ne suit pas nécessairement une fonction puissance, et peut nécessiter des corrections plus spécifiques (Ban & Yamamoto, 2013).

Enfin, les différents types d'écran ont un impact sur la qualité des couleurs et la qualité temporelle de l'affichage (réactivité, rémanence, traînée...). Le sujet de nos travaux portant sur la lecture, et ces dimensions étant plutôt associées à d'autres activités telles que le visionnage de vidéo, l'observation d'images ou le jeu vidéo, nous nous sommes concentrés en majeure partie sur les qualités de résolution spatiale et de luminance des dispositifs d'affichage et n'avons pas traité ces autres problématiques.

2.1.1.2 Le système visuel humain

Si les écrans ont été conçus pour émettre une image qui sera vue par l'utilisateur, il est nécessaire que la lumière émise soit traitée par le système visuel humain pour que l'information puisse être exploitée. L'anatomie du système visuel humain a été largement étudiée par le passé (Hubel, 1994) ; le système visuel chez un sujet « normal » est constitué de deux yeux, des systèmes musculaires associés, de deux nerfs optiques, des voies optiques et du cortex visuel.

Au niveau de l'œil, la lumière traverse la cornée, l'humeur aqueuse, la pupille, le cristallin puis l'humeur vitrée pour se projeter sur la rétine. Chaque composant de l'œil remplit une fonction optique ou d'homéostasie. La cornée et le cristallin assurent la focalisation de la lumière : c'est-à-dire qu'ils permettent aux rayons lumineux parallèles de se concentrer, pour un œil emmétrope (i.e. qui n'a pas de défaut visuel), sur un même point appelé fovéa qui se trouve sur la rétine. La forme du cristallin peut être modifiée par les muscles ciliaires afin d'adapter sa puissance de focalisation pour pouvoir assurer la netteté de l'image de l'objet observé, on parle d'accommodation (cette capacité d'accommodation diminue à partir d'environ 40-45 ans, c'est la presbytie). La taille de la pupille varie en fonction de la contraction ou de la dilatation de l'iris afin de moduler la quantité de lumière qui entre dans l'œil. La pupille se contracte lorsque l'ambiance lumineuse augmente, et se dilate lorsque cette dernière diminue. Le contrôle de la quantité de lumière est la fonction principale de la variation du diamètre pupillaire, mais il varie également avec d'autres facteurs : il diminue avec l'augmentation de l'accommodation, mais il est également lié à des aspects cognitifs tels que la charge attentionnelle (Beatty, 1982) ou les processus émotionnels (Hess, 1965). Une fois que la lumière a traversé l'humeur vitrée, elle atteint la rétine. Elle traverse les premières couches de la rétine, composées des cellules ganglionnaires, amacrines, bipolaires et horizontales, pour aboutir aux photorécepteurs : les cônes et les bâtonnets dont le rôle est de répondre à une stimulation lumineuse par une décharge électrique. Les bâtonnets sont sensibles aux faibles intensités lumineuses ainsi qu'au mouvement, alors que les cônes assurent la détection des couleurs dans des conditions de forte luminosité. Les influx électriques sont transmis aux différentes cellules des premières couches de la rétine. Les axones de ces cellules se rassemblent pour former le nerf optique. Les deux nerfs optiques gauche et droit se croisent au niveau du chiasma optique, et les axones se divisent de telle sorte que les signaux provenant des hémichamps gauche et droit se projettent respectivement dans les corps géniculés latéraux droit et gauche (vision des couleurs, des détails) ainsi que

dans les colliculus supérieurs droit et gauche (vision du mouvement, pour le contrôle oculomoteur). Les corps géniculés (anciennement appelés « corps genouillés ») projettent alors l'information dans le cortex visuel primaire. Nous reviendrons sur le traitement cérébral de l'information visuelle dans la prochaine sous-section 2.1.2 *La lisibilité*.

Les capacités de discernement de l'œil sont limitées à deux niveaux : par l'acuité et par la sensibilité aux contrastes.

Les photorécepteurs tapissent la rétine comme le font les pixels sur un écran. Le traitement de la lumière consiste donc en un échantillonnage de la lumière reçue. Les détails ne peuvent donc être traités par l'œil que de manière limitée, l'acuité visuelle fait référence à cette limite. Elle correspond à la plus petite distance nécessaire entre deux objets pour qu'un sujet puisse distinguer ces objets, on emploie le terme de « minimum séparable ». Cette distance s'exprime relativement au sujet, on peut donc la calculer en une valeur d'angle visuel. La taille en angle visuel se calcule au moyen de la formule suivante :

$$\alpha = 2 * \arctan\left(\frac{T}{2 * d}\right)$$

La fonction arctan s'exprimant en degrés, avec α la taille en degrés d'angle visuel, T la taille de l'objet en cm et d la distance entre l'œil et le centre apparent de l'objet en cm. En multipliant la valeur par 60, on obtient une valeur en minutes d'angle visuel.

Une valeur en minutes d'angle visuel ainsi obtenue peut être inversée pour obtenir la notation d'acuité en 10èmes (couramment utilisée en France), ou être passée à la fonction logarithme pour obtenir la notation LogMAR (pour *Logarithm of the Minimum Angle of Resolution* : logarithme de l'angle de résolution minimal).

L'acuité n'est pas la même sur toute la rétine. Elle est beaucoup plus importante au niveau de la région centrale de la rétine : la fovéa. Cette région de la rétine compte un nombre plus élevé de cônes, de cellules ganglionnaires et projette dans une région plus importante du cortex strié relativement à la rétine périphérique (on parle du facteur d'agrandissement cortical). L'acuité visuelle augmente durant l'enfance et se stabilise à l'adolescence (puis diminue en raison de la presbytie à partir de 40-45 ans). L'acuité varie également entre les sujets. Cette variation est essentiellement due aux défauts de l'œil (amétropies) : défauts de forme qui engendrent une défocalisation de l'image (myopie et hypermétropie) ou un étalement (astigmatisme), défauts de capacités accommodatives (presbytie), ou alors à des

pathologies oculaires ou cérébrales (cataracte, dégénérescence maculaire liée à l'âge, glaucome, cécité corticale...).

La sensibilité aux contrastes forme également une limite à la perception visuelle. Pour être distingué, un objet doit non seulement être de taille suffisante, mais il doit aussi être associé à une valeur de luminance suffisamment différente de celle du fond qui l'entoure : le contraste doit être suffisamment élevé (ou alors l'objet et le fond doivent avoir des couleurs suffisamment différentes). Comme pour les variations de l'acuité, la sensibilité aux contrastes diminue avec les pathologies précédemment citées et avec l'excentricité par rapport à la fovéa. La sensibilité aux contrastes diminue également avec l'âge (Owsley, Sekuler, & Siemsen, 1983), phénomène que l'on peut expliquer par l'augmentation du bruit interne (distorsions dues au vieillissement de l'œil et du système nerveux) et par la diminution de la capacité à gérer ce bruit interne (Barten, 1999; Allard, Renaud, Molinatti, & Faubert, 2013). La sensibilité aux contrastes diminue également avec une hausse de la charge attentionnelle (Huang & Dobkins, 2005).

Le système visuel n'est pas uniformément sensible aux contrastes, la sensibilité variant avec le niveau lumineux du fond. Selon Gustav Fechner (1860), qui s'appuie sur les travaux de Ernst Weber, la sensation (quel que soit le sens utilisé) est linéairement liée au logarithme de l'intensité de la stimulation. La différence d'intensité nécessaire pour percevoir la différence entre deux stimulations est linéairement liée à l'intensité du stimulus.

$$SD = k * \log(I),$$

où SD est le seuil différentiel : la plus petite différence d'intensité qui peut être perçue, k une constante qui dépend du sens mis en œuvre et I l'intensité du stimulus comparé. En appliquant cette loi à la vision des contrastes, on imagine que le contraste nécessaire pour discerner un objet d'un arrière-plan est plus important pour un arrière-plan très lumineux que pour un arrière plan sombre.

Mais cette loi n'est vérifiée que pour des valeurs de luminance moyennes. La valeur k n'est pas constante (ou seulement pour une plage limitée de valeurs d'intensité, voir la *Figure 1*). La courbe de sensibilité à la lumière a notamment été mesurée par Gonzalez et Woods (Gonzalez & Woods, 2002). Le système visuel est moins sensible pour les faibles intensités lumineuses, ainsi que pour les fortes intensités lumineuses (qui vont jusqu'à l'éblouissement).

Ces différences de sensibilités vont donc avoir un impact sur la capacité de l'œil à discerner des contrastes pour certaines conditions lumineuses.

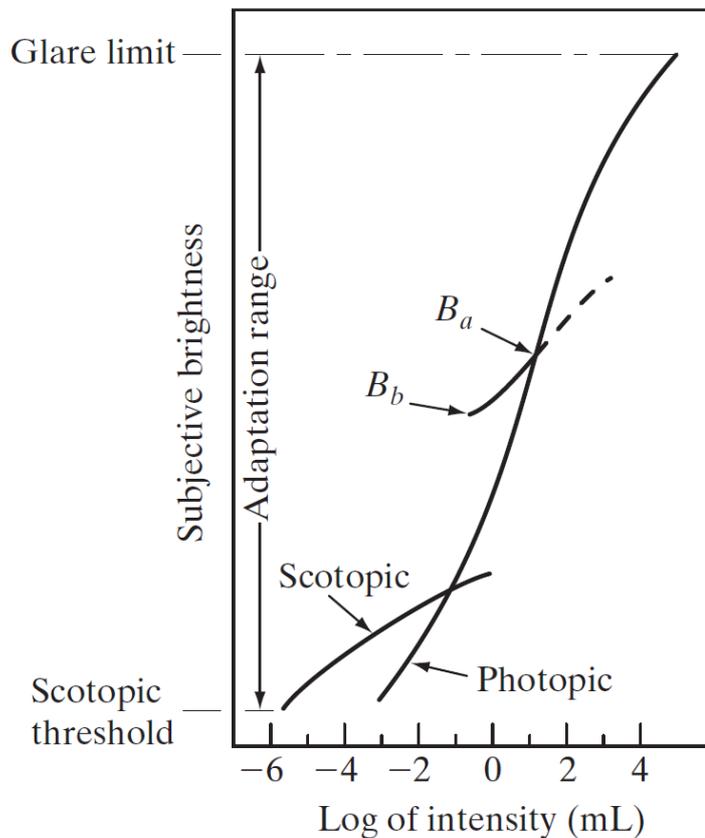


Figure 1. Sensation subjective de luminosité en fonction du logarithme de l'intensité lumineuse mesurée. Tirée de (Gonzalez & Woods, 2002)

2.1.1.3 L'œil et l'écran

Dans les parties précédentes, nous avons montré deux aspects qui limitent les performances du système visuel humain : l'acuité et la sensibilité aux contrastes.

Parallèlement, nous avons montré que les dispositifs d'affichage électroniques sont limités par leur résolution spatiale et par leur profil de luminance. L'affichage de texte va mettre en jeu ces deux limites : afficher une lettre consiste à attribuer des niveaux de luminance définis à certains pixels.

Le niveau de détail de la lettre (empâtement, pseudo-courbes, anti-crénelage) dépend de la définition de l'écran. Plus le nombre de pixels employés est grand, meilleur sera le niveau de détail (voir *Figure 2*).

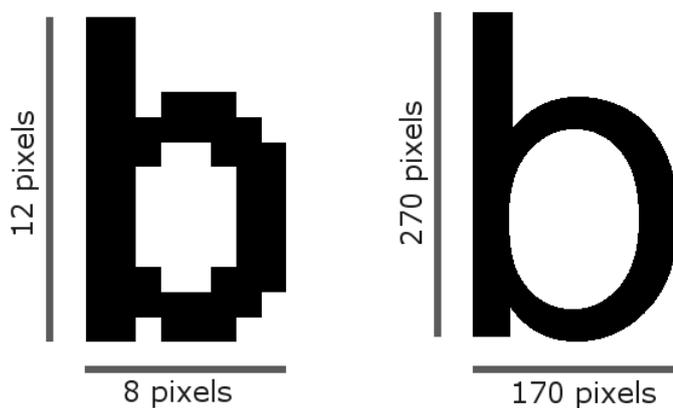


Figure 2. Deux lettres affichées avec un nombre différent de pixels.

L'œil quant à lui, ne peut saisir les détails d'une scène que dans les limites de son acuité. Il est donc intéressant de connaître non seulement la résolution de l'écran (nombre de pixel par unité de distance) et/ou la taille (absolue) des pixels, mais aussi la taille des pixels relative à l'observateur. En effet, si un écran est composé de pixels de grande taille, mais se situe à une distance éloignée des yeux (par exemple, un projecteur de cinéma), la taille relative des pixels pourra être équivalente voire inférieure à celle d'un écran disposant de pixels de petite taille (e.g. un écran de tablette, de smartphone...). Nous avons montré précédemment comment calculer la taille d'un objet en degrés d'angle visuel. Si l'on connaît la taille absolue du pixel, il suffit de connaître la distance qui sépare l'œil de l'écran pour déduire la taille angulaire du pixel.

Le contraste global d'affichage doit être suffisant pour que les objets affichés soient visibles, il est par exemple recommandé d'avoir un ratio de contraste d'au moins 3/1, voire 6/1 selon l'AFNOR (Association française de normalisation, 2002). L'affichage des lettres tire également parti des différents niveaux de luminance qui peuvent être produits par l'écran afin d'adoucir les contours. Le « lissage » (ou « anticrénelage », en anglais « *anti-aliasing* ») consiste à appliquer algorithmiquement des dégradés sur les contours de l'objet afin de diminuer progressivement le contraste entre les pixels isolés et le fond. Puisque la capacité à discerner différents niveaux de luminance est limitée dans le système visuel, il est nécessaire que les différents niveaux RGB utilisés engendrent des niveaux de luminance suffisamment distinguables. Cela peut ne pas être le cas notamment lorsqu'une correction gamma inadaptée est appliquée (voir *Figure 3*).



Figure 3. 1 : caractère sans lissage (voir *Figure 2*) - 2 : caractère avec lissage - 3 : zoom sur 2, on voit les points gris plus ou moins clairs sur les contours de la lettre - 4 : caractère 3 pour lequel la fonction de transfert est une fonction puissance, la plupart des pixels est très foncée - 5 : caractère 3 pour lequel une très forte correction gamma est appliquée, les pixels gris sont très clairs.

Les valeurs de luminance peuvent également être impactées par l'environnement. En effet, l'écran peut côtoyer diverses conditions lumineuses, engendrant parfois des reflets qui peuvent augmenter la luminance de certains pixels, et diminuer alors la visibilité (Huang & Menozzi, 2014).

Par extension, les aspects liés à la visibilité (contrastes, résolution spatiale) vont avoir un effet sur la lisibilité de l'affichage. Par exemple, la vitesse de lecture est affectée par le contraste d'affichage (Legge, Rubin, & Luebker, 1987; Johansson, Pansell, Ygge, & Seimyr, 2014). Il est donc important, lorsque l'on cherche à évaluer la qualité d'un dispositif d'affichage, de connaître ses caractéristiques en termes de résolution spatiale et de profil de luminance, et ce dans un contexte d'utilisation écologique (lumière, distance à l'écran,...).

2.1.2 La lisibilité

Dans la sous-section précédente, nous avons vu comment un dispositif d'affichage produit ou module la lumière, qui va traverser les milieux oculaires, pour être ensuite traitée par le système visuel. La capacité des écrans à faire varier la couleur et l'intensité de leurs pixels est utilisée pour afficher des caractères, qui vont former des mots et des phrases. L'entrée et l'affichage de caractère est une modalité prépondérante des interfaces homme-

machine, certaines interfaces utilisateurs étant uniquement textuelles (notamment le BIOS², des systèmes d'exploitations tels que MS-DOS³,...). Les interfaces se sont enrichies en devenant graphiques, mais le texte reste omniprésent dans les interfaces. Il existe quelques cas très rares d'interfaces sans texte, mais elles ne sont destinées qu'à une population spécifique comme les personnes illettrées (Medhi, Sagar, & Toyama, 2006), ou alors il s'agit de projets artistiques⁴.

Durant l'interaction, les utilisateurs identifient sans cesse les lettres et les mots qui sont affichés. Les lettres et mots affichés peuvent prendre une multitude de formes, et leurs qualités visuelles vont varier avec les caractéristiques des écrans. La lisibilité d'un dispositif d'affichage est fonction des qualités visuelles des lettres, des mots et phrases, qui vont permettre de lire efficacement les textes affichés. Un écran très lisible permet une lecture confortable. La lisibilité étant, par définition, uniquement liée aux caractéristiques visuelles, il convient de la distinguer des aspects de compréhension, qui peuvent aussi faire varier la qualité de la lecture (nous traiterons de ces aspects dans la sous-section 2.1.3 *La compréhension*). Dans cette sous-section, nous traiterons d'abord des constituants du texte : les lettres, les mots et les phrases, ainsi que de leur affichage. Ensuite, nous nous intéresserons au processus de lecture, d'un point de vue neurophysiologique, psycholinguistique et oculométrique. Enfin, nous traiterons de la lisibilité et de sa mesure.

2.1.2.1 Les constituants du texte et leur affichage électronique

2.1.2.1.1 Les lettres

Chaque système d'écriture contient un ensemble de symboles visuels. C'est la juxtaposition de ces symboles qui va permettre de créer des textes. Dans cette thèse, puisque nous étudions la lecture de textes écrits en français, nous nous intéresserons uniquement à l'un de ces systèmes de symbole : l'alphabet latin. Cet alphabet contient 26 lettres de base. Ces

² BIOS pour *Basic Input Output System* : système d'entrée-sortie basique, c'est le logiciel qui contrôle le démarrage de certains ordinateurs.

³ MS-DOS pour Microsoft Disk Operating System.

⁴ Le projet « Wordless Web », développé par Ji Lee et Cory Forsyth, est un plug-in qui efface tous les caractères affichés sur une page web, ne laissant visibles que les illustrations.

lettres sont complétées par des diacritiques et des lettres supplémentaires selon la langue orale qu'ils transcrivent. En français, les diacritiques sont notamment les accents, le tréma et la cédille ; l'alphabet compte une lettre supplémentaire : œ (l'E dans l'O). L'alphabet latin est une écriture bicamérale, c'est-à-dire que les 26 lettres peuvent être déclinées en majuscule ou en minuscule (on parle de la « casse » des caractères). Par analogie à la typographie, nous utiliserons le terme « caractère » pour définir la représentation visuelle d'une lettre, ou d'un signe de ponctuation. Par exemple, la lettre « e » peut prendre la forme des caractères « E » et « e ».

Chaque caractère est constitué d'un ensemble de traits et de courbes. Il existe différentes manières de décrire les primitives qui composent les lettres et/ou les caractères, ces manières étant liées à des hypothèses sur le traitement visuel et cérébral des lettres. Nous reviendrons sur ces hypothèses dans *2.1.2.2.1 Le traitement cérébral des lettres et mots*. Les caractéristiques des traits et courbes ne sont pas strictement figées, et les lettres peuvent prendre différentes formes : il paraît évident que deux caractères manuscrits « a », écrits par deux personnes différentes, ou par la même personne à deux moments séparés, ne sont pas strictement les mêmes. Mais les caractères affichés sur écran peuvent également subir des modifications, notamment en fonction du choix de la police utilisée. Les deux caractères « a » de la *Figure 4* forment un exemple très explicite : les courbes des deux caractères ne sont pas les mêmes, pourtant les deux formes représentent bien la même lettre.



Figure 4. Deux caractères représentant la lettre « a » avec des polices différentes : à gauche, la police utilisée est « Roboto Light » (Google), celle de droite est « CaviarDreams » (Lauren Thompson).

Les polices de caractères varient avec les courbes utilisées, les empattements (formes ajoutées aux terminaisons des caractères), la largeur et la hauteur des caractères, la largeur dédiée à l'affichage du caractère et de l'espace qui le sépare des autres (le terme employé en typographie est la « chasse », qui peut être fixe ou proportionnelle à la largeur du caractère).

Les caractères peuvent être affichés à différentes tailles, les tailles de polices peuvent s'exprimer dans des unités de longueur physiques (généralement le point « pt », qui équivaut à 1/72 pouce). L'utilisation de longueurs physiques est en référence à l'impression des

documents sur papier. Les tailles peuvent également s'exprimer comme le ratio d'une taille de police par défaut du support : l'unité utilisée est alors le cadratin (abréviation : « em »). Enfin, la taille des caractères peut également s'exprimer dans une unité relative à la taille des pixels de l'écran, l'unité est alors le px, mais 1px ne correspond pas forcément à un pixel à l'écran, c'est une unité définie par le World Wide Web Consortium (W3C) pour les feuilles de styles (CSS) qui correspond à un espace qui peut être affiché de manière nette et visible. En fonction de la définition d'un écran, ainsi que de sa distance d'utilisation prévue, la largeur 1px peut correspondre à un pixel de largeur, mais aussi à deux pixels...⁵ Les unités choisies, et les tailles appliquées à la police vont avoir un effet sur la netteté des caractères, notamment en raison de l'application de l'anticrénelage qui a été défini dans la sous-section précédente 2.1.1 *La visibilité*. Enfin, la couleur des caractères n'est pas forcément noire, mais peut être définie par un triplet RGB.

2.1.2.1.2 *Les mots*

Un ensemble de lettres ou une lettre isolée peuvent former un graphème, c'est-à-dire la plus petite entité de l'écriture. Un graphème correspond à la forme écrite d'un phonème, qui est la plus petite unité distinctive dans la parole. Ces phonèmes peuvent être symbolisés par un alphabet phonétique ; un phonème peut correspondre à un ou plusieurs graphèmes : par exemple, le phonème /o/ peut s'écrire « o », « au », « eau ». Réciproquement, un graphème peut correspondre à plusieurs phonèmes : le graphème « y » peut être prononcé /j/ comme dans « ayant » ou /i/ comme dans « cycle ».

La combinaison de plusieurs graphèmes va former un morphème : la plus petite unité de sens de la langue (dans certains cas, un seul graphème suffit, par exemple : « à »). Les mots sont des ensembles constitués d'un ou plusieurs morphèmes (par exemple : « parleurs » est constitué de « parl » lié à l'action de parler, « eur » lié à la personne qui effectue l'action et « s » qui représente le pluriel). Ces ensembles de lettres représentent alors des concepts. Chaque langue est constituée d'un corpus de mots plus ou moins important. Il est difficile de quantifier avec exactitude le nombre de mots d'une langue, dans la mesure où la définition de mot peut être sujette à interprétation (« pomme de terre » correspond à un seul ou à trois mots ?), et où la prise en compte de mots spécifiques à un domaine, la désuétude et à l'apport

⁵ Le site du W3C décrit les types d'unité utilisables pour le CSS dans <http://www.w3.org/Style/Examples/007/units.fr.html>.

de certains mots nouveaux (néologismes ou mots issus de langues étrangères) vont modifier ce comptage. Le recensement des mots, ainsi que la quantification de certaines propriétés de ces mots relèvent de la lexicométrie ainsi que de la psycholinguistique.

Lorsque l'on étudie un mot, on peut s'intéresser à ses aspects grammaticaux : un mot appartient à une catégorie grammaticale (nom, verbe, adjectif,...), il peut avoir un genre et un nombre, une conjugaison, un temps... Chaque mot a une longueur, il s'agit du nombre de caractères qui le composent.

D'autres variables sont construites au regard du rapport entre le mot et les autres mots de la langue. La fréquence lexicale d'un mot correspond au nombre d'occurrences d'un mot dans un corpus d'une certaine taille. La base « Lexique 3 » de New, Pallier, Ferrand et Matos (2001) recense 135 000 mots du français, et contient deux valeurs de fréquences lexicales. La première est basée sur le corpus « Frantext » de l'ATILF⁶ qui est composé de textes littéraires, philosophiques, techniques et scientifiques ; l'autre repose sur des données de sous-titrages de films et séries (dans l'idée de représenter les fréquences liées au langage oral et courant plutôt que littéraire). Il est également possible de calculer le nombre de voisins orthographiques d'un mot. Un mot a des voisins orthographiques si le remplacement d'une lettre dans ce mot permet d'obtenir un autre mot de la langue (« chat » devient « char » ou « chas » en changeant la dernière lettre). On peut compter de la même manière le nombre de voisins phonologiques d'un mot (en changeant l'un des phonèmes, par exemple : « aurore » [œʁœʁ] et « horreur » [œʁœʁ]).

Toutes les suites de lettres ne forment pas nécessairement des mots ; les chercheurs manipulent également des suites de lettres qui respectent les règles de prononciation et d'écriture du français, mais qui ne sont pas des mots appartenant à la langue : les « pseudo-mots » (par exemple : « parnire », « dicope », « rismaté »). Des suites de lettres qui ne respectent pas les règles de la langue sont aussi manipulées, il s'agit des « non-mots » (comme « ejdndks »).

2.1.2.1.3 *Les phrases*

⁶ Base textuelle FRANTEXT, ATILF – CNRS & Université de Lorraine <http://www.frantext.fr>

Pour former une langue, les mots se distribuent, s'accordent et s'agrémentent de ponctuation au sein de phrases, afin de former des liens entre les différents concepts qui les constituent. Des règles influencent la construction de la phrase, en s'appliquant en fonction des mots utilisés, de leurs rôles et de leurs catégories. L'emploi des mots et la construction des phrases vont traduire des idées, des rapports entre concepts, ... L'aspect sémantique est prédominant au regard de la phrase, mais nous n'en traiterons que dans la prochaine sous-section *2.1.3 La compréhension*. Pourtant, la lecture des phrases peut aussi être influencée par les écrans, et la phrase est un enjeu de la lisibilité, au même titre que la lettre et le mot.

Une même phrase peut être affichée de manières diverses. Ces modalités d'affichages sont liées à la mise en forme du texte. La mise en forme est un ensemble de paramètres tels que différentes valeurs d'espace. L'espace inter-mots est la longueur correspondant au vide entre deux mots successifs (la taille d'un caractère espace). Cet espace peut être une valeur fixe, ou varier selon un autre paramètre : l'alignement (le texte peut être centré, aligné à gauche ou à droite, ou alors justifié, auquel cas toutes les lignes ont la même largeur, l'espace inter-mot étant modulé de manière à y parvenir). La longueur de la ligne peut aussi être spécifiée, et va être très dépendante du découpage visuel du texte (présentation en bloc, ou en plusieurs colonnes). L'espace inter-lignes correspond à la longueur du vide situé entre deux lignes successives. Selon le type de document, d'autres paramètres peuvent être manipulés : par exemple, la présence d'un retrait de première ligne, et l'espace inter-paragraphes peuvent être contrôlés dans un document qui est composé de paragraphes.

L'une des différences majeures entre l'affichage électronique de textes par rapport au papier, est que la structure globale du texte peut prendre des formes nouvelles. Un texte peut être affiché page par page, comme dans un livre, mais il peut aussi prendre la forme d'une longue suite de paragraphes, qui ne contient pas de coupure ni d'unité d'organisation telle que la page. Le lecteur doit alors faire défiler le texte afin de pouvoir accéder à l'ensemble du contenu. La disposition de l'information peut être également non-linéaire, c'est le cas des hypertextes, où le lecteur peut passer d'un constituant du texte à un autre en ouvrant des liens. L'ordre de succession des informations est alors laissé en partie au choix du lecteur.

2.1.2.2 Bases neurophysiologiques et psycholinguistiques de la lecture

La lecture et l'écriture sont un procédé artificiel, qui, plus qu'avoir marqué l'Histoire, l'a fait démarrer. En effet, les historiens situent la fin de la préhistoire vers -3500, au moment de l'invention de l'écriture. La lecture est donc un processus relativement récent. Cette

invention a eu de grande retombées pour l'humanité. Plusieurs disciplines scientifiques en ont fait leur objet. La lecture peut être observée du point de vue de la linguistique, qui cherche à identifier les modèles qui peuvent décrire les langues. La psychologie et les neurosciences cherchent à étudier les processus mis en jeu chez le lecteur quand il lit. Nous présentons ici quelques éléments issus de ces disciplines.

2.1.2.2.1 Le traitement cérébral des lettres et mots

Nous avons montré dans *2.1.1.2 Le système visuel humain* comment l'œil traite l'information lumineuse de manière à la transférer aux corps géniculés puis au cortex visuel primaire (qui peut être noté V1, parfois appelé « cortex strié »). Le cortex visuel primaire est responsable de la réception des informations visuelles. A chaque zone de la rétine correspond une zone corticale. Le cortex visuel primaire va mutualiser les informations locales afin de dégager des motifs de plus grande taille : il va coder notamment les barres orientées, les bordures. Des neurones de V1 s'activent ou non selon le motif présent, sa position, son orientation, sa taille, son mouvement (ou son absence de mouvement) et le sens du mouvement. L'information en provenance des neurones de V1 va se projeter dans le cortex visuel secondaire (V2, ou cortex extra-strié). V2 va combiner les influx liés à la présence de barres orientées, et activer des réponses à des formes plus complexes lorsqu'elles sont présentes (coins et angles).

Les connexions de V2 vont ensuite se prolonger sur deux voies : la voie ventrale et la voie dorsale. La voie ventrale est plutôt associée à l'identification et la reconnaissance des objets, alors que la voie dorsale est liée à la localisation et au contrôle visuo-moteur (Schenk & McIntosh, 2010). Etant donné le sujet de cette thèse, nous nous intéresserons surtout à la voie ventrale. En se prolongeant dans la voie ventrale, les informations de V2 vont vers V4, puis vers la région occipito-temporale. C'est dans cette voie que les informations sur les objets isolés vont se combiner pour aboutir à une réaction à des objets spécifiques (visages, objets naturels ou artificiels). Toutefois, la voie dorsale ne doit pas être ignorée, dans le sens où la lecture de lettres active également des zones liées aux aspects moteurs de l'écriture (Longcamp, Anton, Roth, & Velay, 2003) ; et dans la mesure où des résultats comportementaux montrent que le traitement spatial de l'information et le traitement sémantique peuvent interférer entre eux, laissant penser que les deux voies ne sont pas strictement indépendantes (Coello, Bartolo, & Weisbecker, 2007).

Le cerveau traite donc l'information visuelle à partir d'éléments de petites tailles, pour ensuite intégrer des formes plus complexes. Ce schéma de traitement semble donc légitimer les travaux, déjà anciens, de recherche d'éléments de base constituant les lettres : les « *features* », pouvant être traduit par « traits » (Gibson, 1969). En effet, si le cerveau aboutit à la perception des objets en procédant à des découpages de l'image en une série de motifs spécifiques, il est possible d'émettre l'hypothèse que le traitement des lettres suit le même schéma. Les lettres de l'alphabet peuvent être décomposées en un ensemble de traits, qui sont des formes géométriques ou des caractéristiques : lignes horizontales, verticales, diagonales, courbes fermées, ouvertes, présence de redondances (symétrie...) ou encore discontinuité. Geyer & DeWald (1973) ont comparé plusieurs typologies de traits. Pour évaluer empiriquement la validité d'un ensemble de traits comme descripteur des lettres de l'alphabet, les auteurs utilisent des matrices de confusion (Bouma, 1971; Townsend, 1971). Les matrices de confusion sont obtenues en mesurant les niveaux de confusions de chaque lettre de l'alphabet en comparaison à toutes les autres (par exemple, « g » et « q » sont plus fréquemment confondues que « g » et « x »). Des études comportementales plus récentes et basées sur des marqueurs cérébraux tendent à confirmer le traitement des lettres par sélection des motifs constituant ces lettres (Grainger, Rey, & Dufau, 2008; Dehaene, Cohen, Sigman, & Vinckier, 2005). Le fait d'utiliser et de reconvertir des fonctions attribuées à une capacité très ancienne pour l'être humain (détecter des objets de l'environnement) afin de répondre à un besoin récent du point de vue évolutionnaire (lire des textes), est proposé par l'hypothèse du « recyclage neuronal » (Dehaene, 2007). Cette hypothèse, appliquée à la lecture, conçoit le traitement des lettres de la même manière que le traitement des objets (Dehaene et al., 2005) : une fois les motifs de base des lettres identifiés dans V1, ceux-ci sont combinés dans V2. V4 répondrait aux formes des caractères, en restant sensible aux différentes formes pouvant être prises par les lettres (notamment la casse). L'unité de traitement suivante associerait une réponse unique aux formes spécifiques des caractères (l'ensemble des caractères « e, E, e, e » pouvant représenter la lettre « e » engendrer la même réponse). Une fois le traitement des lettres réalisé, le cerveau procéderait à des traitements de plus en plus globaux : il répondrait aux bigrammes (suites de deux lettres), puis aux mots courts et aux suites de lettres fréquentes (trigrammes, morphèmes, etc.).

Une fois les morphèmes identifiés, les traitements ne se limitent plus uniquement à la forme visuelle du langage. Plusieurs modèles de la lecture supposent un traitement parallèle de l'information orthographique et de l'information phonologique (Grainger & Ferrand, 1994;

Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). L'activation simultanée des deux voies de traitement va aboutir ensuite à un codage sémantique des éléments lus. Les formes lues activent des entrées d'un lexique mental, qui associe les morphèmes et les mots à la dimension sémantique qu'ils représentent.

2.1.2.2.2 Le parcours des yeux et la lecture des phrases

Nous venons de voir comment le cerveau traite les lettres et les mots. Mais certains mots sont trop longs pour être contenus entièrement dans le champ de la fovéa, et ne sont donc pas traités en une seule fois. En outre, il est rare que l'activité de lecture se limite à l'identification d'un seul mot isolé ; la lecture porte généralement sur des phrases. Le lecteur déplace son regard afin de pouvoir intégrer les phrases dans leur globalité.

Le lecteur utilise les mouvements de ses yeux (son oculomotricité) pour parvenir à lire. Chaque œil est capable de mouvements rapides et précis grâce à 3 paires de muscles antagonistes. Les muscles droits inférieurs et supérieurs permettent aux yeux de s'élever et de s'abaisser. Les muscles droits médians et latéraux assurent les mouvements des yeux vers la gauche et la droite. Lorsque le mouvement d'un œil est dans un sens symétrique à celui de l'autre œil, on parle de mouvement de vergence : ce mouvement permet notamment de faire converger les axes de regard des deux yeux sur des cibles à différentes distances. Quand le mouvement rapproche les deux axes de regard, on parle donc de convergence ; quand les axes s'éloignent, on parle alors de divergence. Le mécanisme de vergence est lié au mécanisme d'accommodation présenté dans la sous-section *2.1.1 La visibilité* : quand les yeux convergent, ils accommodent et quand ils divergent, ils désaccommodent ; le système visuel maximise ainsi la netteté des objets regardés à différentes distances. Il est également à noter que le mouvement de convergence est associé à un myosis (diminution du diamètre pupillaire), qui permet de limiter les aberrations générées par l'accommodation. L'action simultanée de ces trois mécanismes (vergence, accommodation et variation de la taille de pupille) est connue sous le nom de « triade d'accommodation ». Enfin, les muscles obliques inférieurs et supérieurs sont responsables des mouvements de cyclotorsions. Assez peu étudiés dans les études portant sur la lecture, ceux-ci sont néanmoins détaillés dans la section *7.1 L'effet de la posture sur la lecture*.

Les mouvements des yeux associés à la lecture ne forment pas un déplacement continu, mais se détaillent en de nombreuses séquences formées de deux phases. Les yeux se déplacent à certains endroits de la phrase, et marquent des pauses : on parle de saccades et de

fixations. Les saccades sont des mouvements balistiques très rapides (moins de 50 ms). Les fixations sont des arrêts marqués, de l'ordre de 200-250 ms (mais la durée peut varier selon le mot lu, la tâche, le lecteur, le texte). C'est durant la fixation que le mot observé est traité par le système visuel ; durant la saccade, le système est insensible au changement. En réalité, la fixation ne constitue pas un moment d'immobilité stricte pour l'œil, car celui-ci est continuellement en train de réaliser des mouvements de très petite amplitude (micro-saccades et micro-fixations). D'autres mouvements oculaires peuvent être observés, mais ils sont liés à des activités distinctes de la lecture. Ces mouvements sont liés au suivi d'objets en mouvement, à la compensation des mouvements du corps ou à des pathologies : ce sont la poursuite, le réflexe vestibulaire, le réflexe opto-cinétique et le nystagmus.

L'oculométrie (étude des mouvements oculaires) est une technique largement employée par les chercheurs qui étudient la lecture, comme en atteste la revue de littérature de Keith Rayner (Rayner, 1998). Pour un texte de langue occidentale, la lecture se fait de gauche à droite et de haut en bas : les yeux réalisent plusieurs fixations et saccades vers la droite, et lorsque la fin de la ligne est atteinte, ils se dirigent vers la ligne suivante au moyen d'une large saccade vers la gauche. Néanmoins, des saccades dirigées vers une partie du texte déjà lue auparavant peuvent survenir (saccades vers la gauche ou vers le haut), on parle alors de « rétro-saccades » (ou « régressions saccadiques », « saccades de régression »).

Les fixations et saccades dans le texte s'expliquent par la taille réduite de la fovéa, mais elle n'est pas la seule partie du champ visuel utilisée dans la lecture. En effet la vision parafovéale (autour de la fovéa) est également employée. La zone de la vision qui sert effectivement à la lecture est appelée « l'empan perceptif » (McConkie & Rayner, 1975). Cet empan est asymétrique, il englobe plus de caractères à droite qu'à gauche du point de fixation chez un lecteur occidental (en moyenne, 14 ou 15 caractères à droite contre 3 ou 4 à gauche). La taille de l'empan perceptif peut varier selon les lecteurs, et est un facteur influençant l'efficacité de la lecture, il est notamment lié à la vitesse de lecture (Risse, 2014).

La durée totale de lecture d'un texte est fonction du nombre de fixations et de saccades, ainsi que de la durée moyenne de ces fixations et saccades. Si un lecteur lit un texte en réalisant peu de saccades, mais qui sont de grande taille, ou si ses fixations sont très brèves, sa vitesse de lecture sera plus rapide que celle d'un lecteur réalisant de petites saccades et des fixations longues.

Cette vitesse de lecture varie avec de nombreux facteurs. D'un côté, elle est impactée par des paramètres propres au texte. La fréquence lexicale des mots influence leur durée de traitement (Kliegl, Grabner, Rolfs, & Engbert, 2004), plus un mot est fréquent dans la langue, plus vite il sera lu. La difficulté globale du texte affecte également la vitesse de lecture (Just, Carpenter, Keller, Eddy, & Thulborn, 1996). D'un autre côté, des paramètres liés à l'interaction entre le lecteur et le texte modulent la vitesse. Chaque lecteur dispose ou non de connaissances a priori sur les informations évoquées dans un texte. Le fait d'avoir des connaissances congruentes avec le texte présenté va augmenter la vitesse de lecture (Kendeou & van den Broek, 2007). La vitesse est aussi en lien avec les stratégies de lecture mises en œuvre. Un lecteur n'effectue pas les mêmes mouvements des yeux selon la tâche qu'il réalise (lire un texte pour le comprendre, l'apprendre, chercher un mot dans une liste,...). Les durées de fixations peuvent être traitées par un algorithme afin de déduire quel type de stratégie de lecture le sujet met en œuvre (Cole et al., 2011) : les auteurs distinguent le comportement de lecture de celui de balayage. Les lecteurs alternent régulièrement leur stratégie au cours de la lecture, modulant ainsi leur vitesse. Enfin, des paramètres propres au lecteur vont aboutir à des vitesses spécifiques. La capacité de lecture n'est pas acquise à la naissance, mais nécessite un apprentissage long. Le niveau d'éducation est l'un des facteurs liés aux différences inter-individuelles (Friedman et al., 1999; Landerl & Wimmer, 2008). Les capacités cognitives (comme la mémoire du travail) sont également une source de différences de vitesse de lecture (Traxler et al., 2012). Mais la vision joue aussi un rôle, un déficit de la vision entraînant une chute de la vitesse de lecture (Legge, Rubin, Pelli, & Schleske, 1985; Friedman et al., 1999). Les pathologies se traduisant par une faible vitesse de lecture ne sont pas nécessairement visuelles, en effet la dyslexie développementale (trouble cognitif) aboutit à une vitesse de lecture plus faible que celle des sujets sains (Tressoldi, Stella, & Faggella, 2001).

C'est la combinaison des traitements cérébraux des lettres et des mots, qui, associés aux mouvements oculaires, permettent la lecture de phrases.

2.1.2.3 La lisibilité des écrans

Les dispositifs électroniques affichent des textes, mais de nombreux paramètres vont influencer la forme visuelle finale de ces textes. Ces paramètres, ainsi que les caractéristiques propres des dispositifs vont influencer la lisibilité : ils vont permettre au lecteur de traiter l'information visuelle avec plus ou moins d'efficacité à différents niveaux.

Concernant les lettres, ce sont les tailles et les polices employées qui vont affecter la lisibilité. Une petite taille de police diminue la lisibilité (Lee, Shieh, Jeng, & Shen, 2008) : ces auteurs ont montré que la durée pour identifier des lettres cibles parmi un groupe de lettres est inversement corrélée à la taille des caractères. La taille de caractère influence aussi la vitesse de lecture : la vitesse de lecture diminue à partir du moment où la taille des caractères est plus petite que 2,5 fois la taille liée à la plus petite acuité de caractère identifiable par le lecteur (Chung, Mansfield, & Legge, 1998). La police utilisée est un autre facteur pouvant influencer la vitesse de lecture : Mansfield, Legge & Bane (1996) ont comparé les vitesses de lecture pour une police sans empattements et à chasse fixe (Courier) et une police avec empattements et à chasse proportionnelle (Times New Roman). Chez les lecteurs ne souffrant pas de problèmes de vision, et pour des tailles de caractères permettant une lecture fluide, les auteurs ont constaté que la police Times entraînait une lecture plus rapide. Dans une autre étude (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006), la police Times a aussi été comparée à une police qui contient plus de détails et dont les courbes constituant les lettres sont plus nombreuses (une police de type médiéval : Old English) : les vitesses de lectures sont ralenties pendant la lecture de la police médiévale. Les formes appliquées aux lettres ont donc un impact direct sur la lisibilité de l’affichage.

Nous avons montré dans la sous-section *2.1.1 La visibilité* comment le contraste peut influencer la lecture électronique. Le contraste est fonction des caractéristiques et paramétrages propres au dispositif d’affichage. Mais le contraste dépend aussi des paramètres de couleur définis pour le texte.

La vitesse de lecture est également impactée par la lisibilité des phrases affichées. La comparaison de la vitesse de lecture entre papier et support digital a été l’objet de bon nombre d’études. Les premières études ont montré un ralentissement de la lecture sur écran par rapport au papier (Gould et al., 1987; Dillon, 1992). Cette différence de performance pourrait s’expliquer par plusieurs facteurs : le rafraîchissement de l’écran, la moindre résolution, le contraste trop faible... Mais, à l’heure l’actuelle, les performances tendent à être équivalentes, notamment sur les liseuses électroniques (Siegenthaler, Wurtz, Bergamin, & Groner, 2011). Des études comparatives entre plusieurs dispositifs électroniques ont également été menées : Zambarbieri & Carniglia (2012) ont comparé un écran d’ordinateur avec une tablette (rétro-éclairée) et une liseuse électronique (ainsi qu’avec un texte sur papier). Leurs analyses montrent que les durées de fixations sont plus longues sur écran d’ordinateur que sur tablette et liseuse. Aucune différence de durée de fixation n’a été observée entre la tablette et la

liseuse, ce qui a été confirmé par une autre étude (Siegenthaler, Bochud, Bergamin, & Wurtz, 2012). Dans ces deux études comparatives, les paramètres d’affichage des caractères étaient contrôlés de la même manière pour chaque écran : la taille de caractère était la même, ou presque équivalente pour chaque support et la même police était utilisée. Si les paramètres liés à la police d’écriture sont équivalents, les éventuelles différences observées peuvent être dues aux caractéristiques intrinsèques des écrans (résolution, contraste) ou aux formats de présentation des textes.

Le format de présentation des textes affecte la vitesse de lecture, Dyson (2004) a notamment proposé une revue critique des études tentant de mesurer l’impact de différentes mises en page (les paramètres présentés dans *2.1.2.1.3 Les phrases*). Les éléments principaux de la revue sont l’espace inter-ligne et la longueur des lignes. L’espace inter-ligne doit être suffisamment important pour faciliter une saccade vers la ligne suivante. La longueur des lignes doit être suffisamment importante pour éviter un trop grand nombre de saccades de changement de ligne. La présence de perturbateurs visuels autour du texte lu peut aussi impacter les mouvements oculaires. C’est le cas des publicités animées que l’on peut trouver en marge des textes sur certains sites web (Pasqualotti & Baccino, 2014). Mais le format de présentation n’affecte pas uniquement la lisibilité, il peut également altérer le processus de compréhension (Schmid & Baccino, 2003). Nous reviendrons sur ces aspects dans la partie *2.1.3 La compréhension*.

2.1.2.4 La mesure de la lisibilité des dispositifs d’affichage

Différents types de tests s’attachent à évaluer la lisibilité d’un dispositif d’affichage, chacun s’adressant à un des niveaux d’organisation du texte. Les tests vont ainsi mesurer la performance pour identifier une lettre ou un mot, ou pour lire une phrase.

La capacité à identifier des lettres peut être mesurée à la manière d’une acuité visuelle (l’acuité a été présentée dans la sous-section *2.1.1 La visibilité*). Les lettres sont souvent employées pour mesurer l’acuité (ce sont alors des « optotypes »), c’est notamment le cas pour l’échelle de Snellen, l’échelle LogMAR. La tâche consiste pour le sujet à citer la plus petite lettre qu’il parvient à identifier. Néanmoins, cette mesure est limitée pour apporter de l’information sur la lisibilité d’un affichage. En effet, si elle permet de connaître la plus petite taille de caractère qui peut être perçue par le sujet, elle ne permet pas de faire une différence en termes de performances entre les tailles de caractères : il n’y a pas de distinction entre une taille où les lettres peuvent être identifiées très facilement et une taille où l’identification est

coûteuse en temps et en concentration pour le sujet. On ne connaît pas non plus le temps de traitement pour identifier la lettre, des tests montrent que l'acuité mesurée n'est pas la même si la durée d'observation est limitée (Maissa, Guillon, Wong, & Lebois, 2010).

D'autres tâches permettent d'obtenir plus d'information sur l'identification ou la détection de lettres. La tâche de détection peut consister à soit afficher une lettre, soit ne rien afficher, et demander au sujet s'il a vu une lettre (Pelli, Burns, Farell, & Moore-Page, 2006). Pour l'identification, il est possible d'afficher une lettre à l'écran pendant une durée contrôlée, puis de demander au sujet de déterminer quelle lettre était affichée (Pelli et al., 2006), il s'agit alors d'une tâche de dénomination. Une autre technique consiste à afficher deux caractères, et à demander au sujet si ces caractères étaient les mêmes (Podgorny & Garner, 1979). Il est possible de faire rechercher une lettre dans un ensemble de lettres (Lee et al., 2008), ou à l'inverse, on peut afficher une série de lettres, puis afficher une lettre pour laquelle le sujet doit déterminer si elle était présente ou absente (Sternberg, 1966). L'avantage commun à toutes ces méthodes est de pouvoir obtenir des données sur la précision (en relation au nombre d'erreur) et sur le temps nécessaire au traitement (par la durée d'affichage du stimulus, ou par le temps de réponse du sujet).

Des tests de la même forme peuvent être construits pour évaluer la lecture des mots. La tâche de prononciation consiste à faire citer à voix haute le mot affiché (Forster & Chambers, 1973). Néanmoins cette tâche fournit un temps de réponse qui intègre les phénomènes de production de la parole. Une solution alternative consiste à contrôler la durée d'affichage du mot, afin de mesurer un temps d'affichage critique nécessaire à l'identification du mot, plutôt qu'un temps de réponse. La tâche de catégorisation permet de mesurer le traitement sémantique d'un mot, par exemple en demandant au sujet si le mot affiché comme « mouton » ou « table », correspond à une catégorie sémantique donnée comme « animal » (Cohen, Dehaene, Vinckier, Jobert, & Montavont, 2008). Enfin, une tâche régulièrement utilisée en psycholinguistique est la tâche de décision lexicale (Rubenstein, Garfield, & Millikan, 1970). On présente au sujet des groupes de lettres, qui forment soit des mots, soit des pseudo-mots. Le sujet doit déterminer si le stimulus affiché est un mot de la langue.

Plus un dispositif d'affichage est lisible, plus le sujet pourra identifier rapidement (temps de réponse ou d'affichage court) et précisément (taux d'erreur faible) les lettres et les mots affichés. Mais l'affichage électronique consistant rarement à la présentation d'une lettre

ou d'un mot isolé, il est nécessaire d'évaluer la lisibilité pendant le processus de lecture à proprement parler.

La vitesse de lecture est un indicateur qui répond à ce besoin : il correspond à une mesure qui englobe l'identification des mots isolés, ainsi que les mouvements des yeux et la construction du sens de la phrase. De nombreuses recherches visant à comparer des vitesses de lecture se basent sur la lecture de textes rédigés ou sélectionnés spécialement pour une étude donnée (Siegenthaler et al., 2011; Zambardi & Carniglia, 2012; Benedetto, Draiz-Zerbib, Pedrotti, Tissier, & Baccino, 2013). Plusieurs problèmes sont posés par l'utilisation de textes ponctuels : la mesure de la vitesse sur du matériel différent ne permet pas de comparaison précises entre les valeurs obtenues à travers l'ensemble des études ; au sein d'une même étude, l'emploi de textes non équivalents en mesure répétées peut ajouter un effet du texte sur la performance qui doit être contrebalancé. Inversement, des tests standardisés visant à mesurer la vitesse de lecture existent (Rubin, 2013). Le test MNREAD (Legge, Ross, Luebker, & LaMay, 1989; Ahn, Legge, & Luebker, 1995) sert à mesurer la vitesse de lecture pour différentes tailles de caractères. Il a d'abord été proposé sous format informatique, puis a été adapté sous forme de deux planches. Chaque planche du MNREAD est constituée de 19 phrases de 60 caractères, chacune étant affichée sur trois lignes. La vitesse de lecture est calculée pour chaque phrase en chronométrant la durée de lecture orale et en tenant compte des erreurs. Une version du test a été éditée en français (Senécal, Gresset, & Overbury, 2006). Le test iRest (Trauzettel-Klosinski, Dietz, & the IReST Study Group, 2012) permet la mesure de la vitesse de lecture sur des textes plus longs (10 textes de 830 caractères en moyenne). Ces deux tests présentent une limitation : le nombre de mesures répétées est limité par le matériel, soit les deux planches du MNREAD et les 10 textes de iRest. En effet, si le lecteur était amené à lire plusieurs fois les mêmes textes, il pourrait s'habituer aux mots ou les apprendre, et augmenter artificiellement sa vitesse de lecture. Pour palier cette limite, un algorithme de génération automatique de phrases a été développé (Crossland, Legge, & Dakin, 2008). Plusieurs milliers de phrases peuvent être générées, ce qui permet de procéder à des mesures répétées sans limitation. Nous décrirons plus spécifiquement ce test dans la sous-section 5.2 *La mesure de la vitesse de lecture maximale*.

La vitesse de lecture peut être alternativement testée dans un contexte plus artificiel. La technique de RSVP (*Rapid Serial Visual Presentation*, présentation visuelle sérielle rapide) de Forster (1970) consiste à afficher séquentiellement les mots d'une phrase au même endroit de l'écran. Le lecteur peut alors lire la phrase sans déplacer son regard. En accélérant

ou en ralentissant la durée d'affichage de chaque mot, le seuil de vitesse de lecture maximal peut être mesuré. La mesure ainsi obtenue renseigne sur les temps de traitement de lecture indépendamment des durées consacrées habituellement aux mouvements oculaires.

Enfin, la qualité de lecture sur un support donné peut être évaluée à l'aide des paramètres liés aux mouvements des yeux. Les principales variables utilisées pour mesurer la performance sont : la durée moyenne de fixation, le nombre de fixations, la taille moyenne des saccades et le pourcentage de saccades régressives.

La lecture est donc un ensemble de phénomènes complexes. La performance du lecteur peut être modifiée par les propriétés du dispositif d'affichage et de l'interface. L'effet de ces propriétés doit être comptabilisé à plusieurs niveaux –les niveaux de traitement de la lecture– : de l'identification de lettres à la lecture des phrases, en passant par le traitement des mots.

2.1.3 La compréhension

Les dispositifs électroniques peuvent afficher des pixels, les ordonner de manière à former des caractères, des mots et des phrases qui peuvent être traités par le lecteur. Mais la finalité du processus de lecture ne s'arrête pas au traitement visuel du langage écrit : une fois les mots lus, ils doivent être compris. En effet, le lecteur procède à la lecture du texte avec un but (s'informer à propos d'une actualité, connaître l'avis d'un internaute au sujet d'un film, savoir le temps qu'il fera demain,...). La compréhension du texte est donc l'aboutissement de la lecture, en ce sens qu'elle permet de répondre au besoin du lecteur.

Nous verrons dans un premier temps les aspects sémantiques des textes, et les formes que ces textes peuvent prendre dans les dispositifs d'affichage. Ensuite, nous nous intéresserons au processus de compréhension chez le lecteur, et nous verrons comment les formes prises par l'information électronique peuvent impacter la compréhension. Enfin, nous passerons en revue les méthodes d'évaluation de cette compréhension.

2.1.3.1 Le texte : sémantique et affichage électronique

2.1.3.1.1 Aspects sémantiques des textes

Le but d'un texte est de transmettre des idées aux lecteurs. Nous avons déjà abordé la quantification des mots au regard des autres mots (fréquence lexicale, voisinage orthographique, etc. Voir 2.1.2 *La lisibilité*). Il est également possible de quantifier les mots

par rapport aux concepts qu'ils désignent : sur des aspects sémantiques. La technique de l'analyse sémantique latente (LSA, de l'anglais *Latent Semantic Analysis*) mesure le nombre de cooccurrences de mots dans un corpus de documents (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990). A partir de ces données, des catégories de mots émergent, celles-ci correspondent à des champs sémantiques. De cette manière, les rapports entre les mots sont quantifiés. Les travaux utilisant l'analyse sémantique latente ne sont pas les seuls à chercher à réaliser ces quantifications, il faut notamment souligner l'exemple de WordNet (Miller, 1995). Cette base lexicale référence les mots de la langue anglaise, et les concepts qui y sont associés. Un mot peut être associé à un ou plusieurs sens (polysémie), et plusieurs mots peuvent désigner le même concept (synonymie). La base WordNet mesure ces aspects : la relation entre la forme écrite et le sens, mais aussi des relations plus complexes entre les concepts, d'ordre ontologique ; les concepts y sont classés et hiérarchisés (par exemple, un chat est un mammifère, un mammifère est un animal, etc.).

Les mots de la langue peuvent aussi être quantifiés selon des variables liées à des aspects psychologiques. Un mot signifie un concept, qui peut évoquer des images mentales au lecteur avec plus ou moins de facilité. Par exemple, le mot « chien » peut facilement évoquer un grand nombre d'images, alors que la tâche est plus difficile pour le mot « cent ». La variable qui quantifie cet aspect des mots est appelée « valeur d'imagerie ». Les concepts désignés sont plus ou moins concrets, c'est-à-dire que le lecteur peut, -avec un certain niveau de facilité-, imaginer les percevoir (les voir, mais aussi les sentir, les toucher, les goûter...). La variable associée à ce niveau est la valeur de « concrétude » des mots. Enfin, un mot peut évoquer au lecteur des émotions plus ou moins plaisantes ou déplaisantes (ou être émotionnellement neutre). Les mots « guerre », « monstre », « tuerie » sont liés à des émotions négatives (peur, colère), alors que les mots « joie », « amitié », « rire » sont liés à des émotions positives. Pour quantifier cette dimension, les psycholinguistes utilisent la « valence émotionnelle » ; une valence négative correspond à un concept déplaisant, alors qu'une valence positive correspond à un concept plaisant (une valence nulle est associée à un mot émotionnellement neutre). Ces différentes variables sont mesurées en demandant à des sujets d'évaluer des mots en termes de valeur d'imagerie, de concrétude, et de valence, en général sur des échelles. Des bases existent et contiennent les valeurs moyennes pour un corpus de mots (Bonin et al., 2003), ou pour plusieurs corpus croisés (Leveau, Jhean-Larose, & Denhière, 2011).

Ce ne sont pas les mots pris isolément qui font le sens du texte, mais leurs combinaisons au sein de phrases. En effet, les mots ont chacun une catégorie grammaticale (nom, verbe, adjectif), ils sont liés entre eux (ce qui peut se matérialiser par des marques grammaticales, accords,...). Les mots remplissent alors une fonction : par exemple, le sujet de la phrase est souvent représenté par un nom, qui peut être précédé par un déterminant, un adjectif, l'ensemble ainsi formé constituant un groupe nominal. Ce groupe nominal peut être suivi d'un verbe, lui-même suivi par un complément. Cet ensemble d'éléments forme une proposition (une partie de phrase contenant un verbe, par exemple : « le chat mange la souris »). Une phrase peut être plus ou moins complexe en fonction du nombre de propositions qui la composent. Une phrase simple contiendra une seule proposition, alors qu'une phrase complexe en contiendra plusieurs (exemple : « Aujourd'hui, j'ai vu que le chat a mangé une souris, bien que celle-ci ait couru le plus vite possible pour s'enfuir. »). La complexité des phrases à plusieurs propositions peut être accentuée par la présence de pronoms, qui remplacent les noms des sujets. Il est nécessaire alors pour le lecteur d'entretenir une mémoire des antécédents textuels pour savoir à quoi ou à qui le pronom fait référence (on parle d'antécédents anaphoriques).

Les textes diffèrent selon leur contenu, et peuvent être classés dans des types. La principale typologie des textes a été proposée par Werlich (1975), qui distinguait 5 types de textes :

- le texte descriptif : description d'un objet, d'un lieu, d'un animal...
- le texte narratif : raconte une suite d'évènement,
- le texte expositif : explication sur les liens de cause à effet entre plusieurs concepts,
- le texte argumentatif : valorise et défend un point de vue, une opinion,
- le texte instructif : donne une marche à suivre, une suite d'ordres,....

La puissance descriptive des typologies de texte reste limitée, puisque tous les textes ne s'accordent pas forcément parfaitement avec un type de texte donné. Un texte peut mélanger plusieurs types, ou être trop à la marge pour justifier un type de texte spécifique. La notion de types de textes est donc remise en cause (Adam, 2005). Cette notion ne doit donc

pas être prise comme un moyen parfait de classification des textes, mais elle reste utile pour avoir une idée générale de la forme des idées contenues dans les textes.

En résumé, les textes peuvent prendre une multitude de forme, en fonction des concepts qui y sont désignés, de l'agencement des mots et des phrases, ainsi que des buts du texte.

2.1.3.1.2 Affichage électronique des textes : l'interface

Nous venons de voir les dimensions qui font qu'un texte peut exprimer un nombre immense d'idées. A partir des 26 lettres de l'alphabet, l'auteur du texte peut générer un nombre presque infini d'idées, grâce aux mots, et à leur organisation. Mais le fond, le texte en soi, n'est pas le seul élément à prendre en considération dans la lecture électronique : la forme y joue également un rôle prépondérant. Nous avons vu dans la sous-section précédente (2.1.2 *La lisibilité*) que les textes affichés électroniquement peuvent prendre une multitude de formes en fonction de leur mise en page. La mise en page du texte n'est pas le seul paramètre qui va modifier son organisation spatiale : l'affichage électronique du texte n'étant pas figé, la disposition spatiale, mais aussi temporelle, peut évoluer au cours du temps et en fonction des interactions de l'utilisateur. Le texte électronique ne se matérialise donc pas au sein d'un document immuable et ancré dans un support physique, mais au sein d'une interface, qui est le résultat de l'état interne d'un dispositif électronique à un moment donné, et qui va pouvoir s'altérer.

Les interfaces électroniques peuvent être multimédias, c'est-à-dire qu'elles intègrent plusieurs types de contenus. Si l'ajout d'images au texte était déjà possible sur un support papier, les interfaces électroniques peuvent aussi contenir des animations (notamment des vidéos) et du son⁷. Les interfaces sont plus ou moins riches : elles peuvent ne comprendre que peu d'éléments visuels (la page d'accueil du moteur de recherche Google ne comporte par exemple qu'un logo, une barre de recherche, et quelques éléments textuels), ou alors être constituées d'un grand nombre d'éléments de différentes formes (par exemple, le réseau social Facebook propose des images, des photos, des vidéos, des publicités, des menus...). La

⁷ Si le texte, l'image, la vidéo et le son sont les principales modalités des interfaces, elles ne sont pas les seules. On peut également évoquer les interfaces « haptiques », où des forces physiques sont mises en œuvre pour donner un retour à l'utilisateur dans une autre gamme sensorielle.

forme de l'interface dépend des fonctionnalités proposées, ainsi que des choix esthétiques et graphiques des concepteurs.

La quantité d'information qui peut être fournie au moyen d'une interface est variable : elle peut aller d'une information minimale (par exemple : une horloge digitale qui n'afficherait que l'heure) à une quantité d'information immense (c'est le cas d'une encyclopédie en ligne, telle que Wikipedia qui contient plus d'un million et demi d'articles). Lorsque toute l'information textuelle n'est pas contenue dans un seul écran, il est nécessaire de proposer à l'utilisateur des solutions pour faire évoluer le contenu affiché. Nous avons présenté la technique du défilement (*scrolling*) dans 2.1.2 *La lisibilité*, qui permet de modifier la position du texte. Mais les éléments de contenus peuvent aussi être distribués dans plusieurs documents, que l'utilisateur doit atteindre pour accéder à l'information. L'analogie au déplacement physique est alors largement utilisée pour décrire cette variation d'information affichée (« aller dans un dossier », « naviguer vers un site ») : on parle de « navigation ». La navigation est généralement rendue possible grâce aux « liens hypertextes ». Un lien hypertexte est un élément cliquable d'un document qui permet d'afficher le contenu d'un autre document, ou de déplacer la position dans le document courant. Le lecteur électronique est donc amené, au cours de sa lecture, à ouvrir de nouveaux documents, revenir sur des documents précédents. En comparaison à la lecture d'un texte sur papier, la lecture électronique est donc moins linéaire, car elle associée à une navigation qui dépend des ouvertures de liens hypertextes.

Les interfaces peuvent donc prendre une multitude de formes, néanmoins ces possibilités sont contraintes par les supports. Le type de technologie employé peut permettre ou non certains types de contenu : si un écran LCD peut afficher des animations, ce n'est pas le cas d'un support utilisant l'encre électronique, qui n'offre pas une vitesse de rafraîchissement suffisamment élevée pour afficher rapidement une série d'images. Ce constat est le même pour l'affichage d'éléments en couleurs. La taille de l'affichage peut aussi limiter le contenu, un écran de smartphone ne pourra pas afficher autant d'information qu'un écran d'ordinateur de 24 pouces. Cet aspect est particulièrement évident pour les sites internet conçus avec la technique de « responsive web » : ces sites permettent de faire varier leur contenu (en disposition et en quantité) en fonction de la taille d'affichage. Ainsi, un même site

pourra être consulté depuis un ordinateur, une tablette ou un smartphone, mais la forme de l'interface va pouvoir s'adapter à chaque support⁸.

Le texte subit donc d'un côté les contraintes de l'interface, qui limite l'espace qui peut lui être alloué, mais bénéficie d'un autre côté des possibilités offertes par la dématérialisation. La lecture électronique donne accès à une masse de données considérable, et cet accès s'inscrit dans la lecture électronique.

2.1.3.2 La compréhension en lecture

La lecture est un processus séquentiel, les mots ou groupes de mots sont traités les uns après les autres, comme l'attestent les mouvements des yeux, qui se déplacent par saccades sur le texte. Ce traitement progressif du texte implique l'action de mémorisation pour permettre la compréhension. En effet, la compréhension correspond à l'extraction des informations du texte pour inférer des connaissances générales répondant aux buts de lecture. L'aspect séquentiel de la lecture rend donc nécessaire la mise en mémoire des informations locales pour être traitées de manière à former une connaissance globale.

Néanmoins, la compréhension d'un texte ne se résume pas à la mémorisation séquentielle des éléments du texte : plusieurs modèles cherchent à décrire le processus de compréhension en lecture, et les différents niveaux de traitements qui y sont associés.

L'un des modèles fondateurs de la compréhension en lecture de texte est celui de Kintsch et Van Dijk (1978). Leur modèle présuppose que pour que le lecteur comprenne les aspects sémantiques du texte, il doit construire une représentation du contenu de ce texte à plusieurs niveaux. Le lecteur prend d'abord connaissance du contenu textuel. Il s'agit de traduire les phrases du langage naturel en propositions symboliques, composées d'un prédicat et d'un ou plusieurs arguments (« Le chien ronge un os » devient [RONGER, CHIEN, OS]). Les phrases du texte sont ainsi converties en mémoire au fur et à mesure de la lecture. Les auteurs dénomment ce construit la « micro-structure ». La micro-structure permet de créer les liens de référence entre les différents éléments du texte. Néanmoins, les capacités mémorielles du lecteur n'étant pas illimitées, il est nécessaire de procéder à des opérations pour réduire la

⁸ Il est possible de tester des sites « responsive », comme <http://seesparkbox.com/> en faisant varier la taille du navigateur afin d'observer les changements dans l'interface.

quantité d'informations à retenir. Le lecteur procède donc, parallèlement à sa lecture, à des opérations sur les propositions en mémoire. Les propositions peuvent être supprimées, généralisées, remplacées par un construit ou rester inchangées. Ces opérations sont appliquées selon les buts du lecteur : si une proposition est jugée pertinente, elle sera conservée et traitée, alors que si ce n'est pas le cas, elle sera supprimée. Le résumé des propositions ainsi extraites est appelée la « macro-structure ».

Les auteurs ont par la suite enrichi leur modèle en considérant d'autres grandes fonctions mises en œuvre durant la compréhension (Van Dijk & Kintsch, 1983). La première fonction mise en œuvre est l'étape d'encodage des mots. A partir de là, le lecteur peut construire la micro et la macro-structure. Parallèlement à la lecture, le lecteur renouvelle ses attentes sur les éléments qu'il espère trouver dans le texte, à partir de ses connaissances du monde, ainsi que de ses expériences personnelles (par exemple : si le lecteur lit un conte, il s'attend à trouver un dénouement à la fin...). Cette interaction entre le contenu textuel et les connaissances du lecteur génère un « modèle de situation », le lecteur créant alors une représentation mentale de la situation décrite par le texte, des intentions de l'auteur, le style de texte présenté. Ces différents processus, allant du traitement des mots, au modèle de situation, en passant par le traitement des propositions, ne sont pas unidirectionnels. Ces processus vont pouvoir rétroagir les uns sur les autres : par exemple, la compréhension des enjeux du texte dans le modèle de situation va pouvoir influencer les stratégies de construction de cohérence entre les propositions. Les auteurs reconnaissent que ces travaux présentent plutôt un cadre de travail pour envisager la compréhension, plutôt qu'un modèle qui définit matériellement le traitement des informations du texte.

Se basant sur ces travaux, Kintsch (1988) a complété la théorie en proposant le modèle de « Construction-Intégration ». L'élaboration des différents niveaux de compréhension du texte s'effectue d'abord par une étape de construction. Le lecteur récupère l'information linguistique du texte. Il élabore alors, à partir de ses connaissances, des inférences afin de rendre l'information cohérente. Les inférences construites peuvent être contradictoires entre elles, le lecteur doit en valider certaines, et rejeter les autres pour que sa compréhension du texte soit cohérente. Intervient alors l'étape d'intégration, les inférences jugées non cohérentes au regard de l'information du texte et des connaissances du lecteur sont rejetées. Ce cycle construction-intégration est répété tout au long de la lecture afin de construire une représentation cohérente du texte.

Bien que les auteurs spécifient qu'il ne s'agit que d'un choix pratique, et pas d'une assertion sur la nature de l'information traitée (Van Dijk & Kintsch, 1983), les modèles de Kintsch et Van Dijk mettent en avant un traitement symbolique de l'information textuelle. D'autres modèles postulent une forme différente de l'information. Le modèle « landscape » (van den Broek, Risdén, Fletcher, & Thurlow, 1996; Rapp & van den Broek, 2005) est proche des modèles présentés précédemment, mais considère que la compréhension se fait à l'aide de l'activation de faisceaux. Ces faisceaux correspondent aux concepts mentionnés dans le texte. Durant la lecture, les faisceaux correspondant aux personnages, aux objets, aux actions s'activent. L'activation des faisceaux va se renforcer si ces concepts sont à nouveau mentionnés dans le texte, et s'amenuiser si ceux-ci ne sont plus mentionnés. Mais l'activation ne dépend pas uniquement du contenu du texte, elle est également liée aux inférences réalisées par le lecteur. Les auteurs distinguent les inférences analogiques et les inférences causales. Les inférences analogiques permettent de maintenir la cohérence sur le sujet de l'action (comprendre à quoi ou à qui fait référence un pronom, par exemple). Les inférences causales permettent d'assurer la cohérence entre les événements si des informations ne sont pas mentionnées dans le texte, mais qu'elles sont nécessaires pour que les phrases soient cohérentes entre elles (les auteurs donnent l'exemple suivant : « (1) La dame donne 10\$ au serveur. (2) Le serveur lui apporte sa monnaie. » Il est nécessaire pour le lecteur d'inférer que le prix de ce que la dame a consommé était inférieur à 10\$ et que le serveur devait en conséquence lui rendre de la monnaie. Cette construction du lecteur permet de maintenir la cohérence du texte.). Les faisceaux les plus activés au cours de la lecture vont s'inscrire en mémoire, et les faisceaux activés simultanément vont être liés, créant ainsi des relations entre les concepts. La compréhension du texte prend donc la forme d'un « paysage » (*landscape*) de faisceaux représentant des concepts, qui s'activent et se désactivent au cours de la lecture.

Au-delà des aspects pratiques sur le transfert en mémoire de l'information du texte, ces modèles s'accordent sur le fait que la compréhension implique, de la part du lecteur, la mobilisation de ses connaissances antérieures, la construction d'inférences pour mettre en place une cohérence du texte et la définition de buts de lecture. La compréhension va mettre en jeu ces aspects, et le contenu du texte pour construire une représentation du texte qui va être mémorisée par le lecteur. Ces modèles mettent également en avant les aspects motivationnels et attentionnels de la lecture : la poursuite des buts de lecture, la volonté du lecteur à produire des inférences pour comprendre le texte sont liées à sa motivation ; la quantité d'information qui va être traitée dépendra des ressources attentionnelles du lecteur.

2.1.3.3 Comment les interfaces agissent sur la compréhension

La compréhension varie selon les textes et les lecteurs. Mais les possibilités offertes par la lecture électronique vont également influencer la compréhension : nous présenterons ici le rôle des interfaces dans ce processus.

L'aspect multimédia des interfaces peut renforcer les connaissances issues de la lecture de textes. L'accompagnement du texte par des images peut améliorer la compréhension, à condition que celles-ci proposent une information cohérente avec le contenu du texte (Carney & Levin, 2002). D'autres illustrations peuvent aussi apporter cet effet bénéfique, c'est le cas des diagrammes qui représentent l'objet du texte (Hegarty & Just, 1993). A l'inverse, si le texte est accompagné d'illustrations qui apportent des informations qui ne coïncident pas avec les buts de lecture, la compréhension est affectée négativement (Harp & Mayer, 1998). Les auteurs soutiennent que les illustrations non adaptées peuvent faire mobiliser au lecteur des connaissances antérieures qui ne sont pas utiles pour l'accomplissement de l'objectif de lecture. Cet effet négatif issu de stimuli visuels non pertinents pour la lecture de texte peut être très présent dans le contexte de la lecture électronique : l'exemple de la navigation sur internet, et des publicités qui entourent les contenus a également montré une baisse de la compréhension (Simola, Kuisma, Öörni, Uusitalo, & Hyönä, 2011).

Nous avons vu dans la sous-section 2.1.2 *La lisibilité* que la mise en page du texte peut influencer sa lisibilité, ainsi que les mouvements oculaires. La mise en page est un des constituants d'une interface, et cette mise en page peut également influencer la compréhension. Dans l'expérience de Schmid et Baccino (2003), les auteurs ont montré que le taux de rappel pouvait globalement diminuer lorsqu'un texte explicatif est présenté en bloc, par rapport au même texte présenté en paragraphes distincts, chaque paragraphe correspondant à une étape de l'explication.

L'aspect non-linéaire de la lecture électronique, du à l'utilisation des liens hypertextes, peut également interagir avec la compréhension du texte. La présence des liens hypertextes permet un accès rapide à un grand nombre d'informations supplémentaires. Le revers de cette fonctionnalité peut se décliner sous deux formes pour le lecteur. D'une part, il est obligé de construire une mémorisation spatiale de son avancée dans le document, mais ce dernier n'étant pas inscrit physiquement dans un support, il ne peut y avoir d'organisation spatiale à proprement parler (Baccino, 2004). La lecture d'un document comportant des liens

hypertextes affaiblit la construction d'une représentation mentale du texte (Miall & Dobson, 2006). Lorsque le lecteur doit revenir sur un document précédemment visionné pour vérifier une information, il peut le faire avec plus ou moins d'efficacité selon son modèle de l'organisation du document qu'il a su construire. D'autre part, si les liens hypertextes peuvent apporter une information complémentaire essentielle à la compréhension, ils peuvent également s'avérer être des distracteurs, qui peuvent perturber les buts de lecture. Le lecteur peut débiter la lecture d'un document électronique avec un objectif, mais en sélectionnant des liens, le premier objectif va s'estomper pour laisser la place à de nouveaux objectifs (Baccino, 2004). Le lecteur ne lira plus pour répondre à son objectif initial, et sa compréhension du contenu sera limitée.

Néanmoins, les aspects négatifs des hypertextes sont à relativiser, car ils peuvent être influencés par plusieurs facteurs. L'effet de la présence de liens hypertextes sur la compréhension n'est pas le même selon le type de texte : si, pour un texte narratif, la présence de liens perturbe la compréhension, ce n'est pas forcément le cas pour un texte de type encyclopédique (Zumbach & Mohraz, 2008). L'effet négatif des liens sur la compréhension serait également corrélé positivement au nombre de liens présents (Zhu, 1999), un grand nombre de liens impliquant une désorientation plus importante qu'un nombre de liens limité. L'utilisation parcimonieuse des liens hypertextes peut donc limiter leur effet néfaste sur la compréhension. En revenant sur l'aspect multimédia des documents électroniques, l'intégration de la structure des documents sous forme d'hypertextes par le lecteur peut être facilitée en affichant la structuration de ces documents sous forme d'un plan schématique au début de la lecture (Salmerón, Baccino, Cañas, Madrid, & Fajardo, 2009), et ainsi permettre une meilleure compréhension. Enfin, dans une revue de travaux sur l'effet des documents hypertextuels, DeStefano et LeFevre (2007) ont montré que plusieurs études s'accordent sur le fait que les connaissances à priori des lecteurs peuvent limiter l'effet néfaste des liens.

La structure de l'interface a donc une influence importante sur la compréhension. Les interfaces peuvent guider la compréhension du lecteur par une mise en forme mettant en avant les idées du texte, et par une structuration des documents adaptée aux buts de lecture. En revanche, une interface présentant une information désorganisée, et jonchée de distracteurs peut gravement atteindre la compréhension du texte.

2.1.3.4 Evaluation de la compréhension

Nous venons de voir que la compréhension est un processus complexe, et la variété des méthodes qui servent à l'évaluer reflètent cette complexité. Rapp et van den Broek (2005) expliquent que deux approches principales sont à identifier dans les recherches sur la compréhension. D'un côté, des recherches se concentrent sur le produit de la compréhension, le modèle de structuration de l'information qui résulte de l'activité de lecture. De l'autre côté, il est possible de s'intéresser aux processus actifs durant la lecture, c'est-à-dire aux fluctuations de l'activité cognitive phrase après phrase. Chaque approche repose sur des mesures spécifiques. Néanmoins, les auteurs montrent que les deux approches peuvent, et ont tout intérêt à être combinées pour pouvoir considérer l'activité de compréhension dans son ensemble.

2.1.3.4.1 Les mesures du produit de la lecture

Les mesures de la compréhension peuvent porter sur le produit de la lecture, à savoir les informations acquises par le lecteur à l'issue de cette activité. Dans ce cas, on s'intéressera d'une part à la mémorisation des éléments constitutifs du texte : il s'agit de la compréhension de « bas niveau », qui correspond à la micro-structure du modèle de Kintsch et Van Dijk (1978). D'autre part, on s'intéressera aux inférences construites par le lecteur, qui sont liées à une compréhension de « haut niveau », en lien avec la macro-structure (Kintsch & van Dijk, 1978) et avec le modèle de situation (Van Dijk & Kintsch, 1983).

2.1.3.4.1.1 Techniques de mesure de la mémorisation

Les techniques ne manquent pas en psychologie cognitive pour évaluer la mémorisation de l'information. La tâche de rappel libre est très simple à mettre en place, elle consiste à demander au sujet de citer le maximum de termes qu'il a rencontré dans le texte lu. Une mesure résultante peut être le taux de mots lus par rapport au nombre total de mots présentés. Une alternative à cette tâche est le rappel indicé. Dans ce cas, l'expérimentateur donne au sujet un ou des indices pour susciter le rappel. Pour l'évaluation de la compréhension, cette tâche peut prendre la forme d'un test de complétion : des fragments de phrases sont proposés au sujet, celui-ci doit les compléter avec les mots lus précédemment. On mesure alors le taux de bonnes réponses.

Une alternative aux tâches de rappel consiste à procéder à des tâches de reconnaissance. On présente au sujet des mots qu'il a pu rencontrer dans le texte, et des mots

absents dans le texte (distracteurs). Il peut être demandé au sujet de déterminer si le mot présenté fait partie du texte (réponse « oui » ou « non »), ou alors on peut présenter plusieurs mots, et le sujet doit déterminer lequel était présent. Dans ce dernier cas, on mesurera le taux d'erreur. Pour les tâches oui/non, il est nécessaire de s'appuyer sur la théorie de détection du signal pour extraire les indicateurs de performance. En effet le sujet adopte une stratégie libérale ou conservatrice (il a tendance à donner des réponses même s'il n'en est pas sûr, ou bien il préfère s'abstenir plutôt que de risquer une mauvaise réponse). Il est donc nécessaire de calculer le nombre de bonnes réponses (en anglais *hits*, représenté par la lettre H), le nombre de fausses alarmes (FA) : le sujet dit avoir vu le mot alors qu'il s'agit d'un distracteur, et le nombre de rejets incorrects (en anglais *miss*, M) : le sujet dit ne pas avoir vu le mot, alors qu'il s'agissait d'un mot présent dans le texte.

Les tâches de reconnaissance peuvent être raffinées de manière à observer quel processus mnésique a été utilisé pour la mémorisation du mot. On emploie alors le paradigme « *Remember versus Know* » (R-K, « souvenir contre savoir ») (Tulving, 1985). Ce paradigme permet de distinguer l'information qui est connue, de l'information dont le sujet se souvient. Si le sujet sait qu'il a vu un mot, mais qu'il est incapable de se rappeler du contexte dans lequel il a été confronté à ce mot, alors ce mot serait stocké dans sa mémoire sémantique. A l'inverse, si le sujet se souvient, c'est-à-dire qu'il est capable de décrire le contexte qui lui a permis de retenir le mot, celui-ci serait contenu dans la mémoire épisodique. La tâche R-K prend la forme d'une tâche de reconnaissance, on présente au sujet des mots issus du texte, ou des distracteurs. Le sujet doit déterminer si le mot présenté est un mot faisant partie du texte, et si c'est le cas, il doit dire s'il s'en souvient (contexte d'apprentissage), ou s'il sait juste qu'il l'a vu. Il est conseillé de poser d'abord la question sur la nouveauté du mot (oui/non) avant de poser la question de la forme de conscience associée à la connaissance ; si les deux questions sont posées en même temps (trois choix : non vu, savoir, souvenir), cela peut entraîner un biais libéral (Hicks & Marsh, 1999). Plusieurs modèles de traitement des résultats existent pour déterminer des indicateurs de compréhension à partir des bonnes réponses, fausses alarmes et rejets incorrects (Snodgrass & Corwin, 1988).

2.1.3.4.1.2 Techniques de mesure des inférences

La technique la plus courante pour évaluer les inférences faites par le lecteur consiste à employer des questionnaires de compréhension. Les questions peuvent être ouvertes (le sujet construit sa réponse lui-même) ou fermées (le sujet répond « oui » ou « non », « vrai » ou

« faux »). Les questionnaires peuvent également être à choix multiples (QCM) : plusieurs possibilités de réponses sont proposées au sujet, et celui-ci doit choisir celle qui est correcte. L'utilisation de questionnaire nécessite donc de définir les questions, ainsi que les réponses (qui serviront à vérifier la performance du sujet pour une question libre ou fermée, et qui seront proposées au sujet dans le cas d'un QCM). Ces questions et réponses sont généralement rédigées directement par les auteurs du test ou les chercheurs réalisant l'étude. Cela pose une limite aux tests de compréhension, dans la mesure où les tests ne sont pas généralisables au-delà des textes choisis et pour lesquels les questions ont été construites. Il est donc nécessaire de reconstruire du matériel de test à chaque fois qu'une mesure de compréhension doit être effectuée.

Les questions et réponses peuvent néanmoins être générées automatiquement. L'algorithme de Mostow et al. (2004) génère automatiquement plusieurs choix de mots pour des textes à compléter. La mesure permet donc de vérifier que le lecteur a bien réussi à construire les inférences référentielles dans le texte. L'utilisation du traitement automatique des langues (TAL) peut également être mise à profit pour générer des questions complètes. Le TAL permet de transposer dans un langage symbolique le contenu d'un texte en langage naturel (Gates, 2008; De Viron, Bernhard, Moriceau, & Tannier, 2011). Une fois le contenu du texte incorporé, l'algorithme peut choisir un élément significatif de la phrase (par exemple : le sujet de l'action, le lieu, la date,... qui sont chacun identifiés comme tels par l'algorithme). L'algorithme n'a plus qu'à générer la question correspondante (Qui ? si la question porte sur le sujet, Où ? si la question porte sur le lieu,...).

L'évaluation du produit de la compréhension porte donc sur deux niveaux : sur la mémorisation des éléments du texte, et sur la construction des inférences. Ces deux niveaux doivent être évalués, car ils gardent chacun un certain degré d'indépendance. En effet, la compréhension de bas niveau peut être correcte, sans aucune compréhension de haut niveau (le lecteur se rappelle du texte, mais n'a pas compris le sens). A l'inverse, le lecteur peut avoir bien intégré les enjeux du texte, mais oublié les détails.

2.1.3.4.2 Les mesures concomitantes à la lecture

La compréhension d'un texte étant un processus dynamique, elle ne peut pas se résumer au construit final réalisé par le lecteur. Des techniques permettent de contrôler la compréhension en temps réel, c'est-à-dire au cours de la lecture. En outre, la dimension

attentionnelle étant importante dans le phénomène de compréhension, nous ferons également un point sur sa mesure.

2.1.3.4.2.1 Les mesures « on-line » de la compréhension

La mesure comportementale la plus courante pour mesurer le processus de compréhension consiste à chronométrer la durée de lecture. Toute chose étant égale par ailleurs, un texte facile sera lu plus rapidement qu'un texte difficile (Young & Greig Bowers, 1995). Il convient tout de même de rester vigilant quant à cette mesure, en effet, la durée de lecture n'est pas strictement liée à la facilité de compréhension. Une durée courte peut certes être due à une compréhension efficace, mais elle peut aussi être le résultat d'une mauvaise compréhension, dans le cas où le lecteur délaisse la compréhension profonde pour une stratégie de lecture superficielle (Dyson & Haselgrove, 2001), auquel cas la compréhension est faible.

Les étapes de la compréhension peuvent être verbalisées au cours de la lecture par le lecteur lui-même, la méthode utilisée est le « *Think-aloud* », pensée à voix haute (Wade, 1990). La méthode consiste à demander au lecteur de verbaliser ce qu'il est en train de faire, de penser, et éventuellement à l'inciter à le faire durant la lecture s'il ne respecte pas la consigne. Le think-aloud permet donc un accès privilégié aux traitements réalisés par le lecteur, mais les données issues de cette méthode sont volumineuses, et longues à analyser.

Magliano et Millis (2003) ont proposé d'étoffer la méthode think-aloud pour simplifier le traitement des données. Les auteurs comparent les verbalisations des lecteurs avec les différentes phrases du texte à l'aide de la technique de l'analyse sémantique latente (LSA). Les phrases sont classées selon leur statut : phrase courante, phrase précédente liée causalement et autre phrase. Les auteurs montrent que chez les lecteurs les plus efficaces, les pensées à voix haute sont sémantiquement proches des phrases reliées causalement, alors que chez les lecteurs moins efficaces, les verbalisations sont sémantiquement proches de la phrase en cours. L'utilisation conjointe du think-aloud et de LSA permet donc une évaluation simplifiée de l'efficacité de la compréhension.

La technique LSA peut également servir dans d'autres contextes d'évaluation de la compréhension. Elle peut servir notamment à mesurer l'efficacité des stratégies de navigation à travers des liens hypertextes. Salmerón et García (2011) ont observé le comportement de navigation d'élève devant répondre à des questions en utilisant un document hypertextuel.

Les auteurs utilisent la mesure de proximité sémantique (issue de LSA) entre les différentes pages du document et les termes de la question posée. La mesure permet ainsi de vérifier la cohérence entre le but de lecture défini à l'aide de la question et les choix de navigation effectués par les lecteurs. Ils montrent que cette méthode s'avère être un bon indicateur de la compréhension, car la cohérence de navigation est corrélée aux mesures de compréhension finales.

Enfin, une dernière mesure permet d'accéder en continu au processus de compréhension, sans pour autant demander d'effort supplémentaire au lecteur (comme la verbalisation dans le cas du think-aloud). Il s'agit de l'utilisation de l'oculométrie pour l'évaluation de la compréhension. Les durées de fixation, ainsi que leur nombre sont corrélées à la difficulté de compréhension (Rayner, Chace, Slattery, & Ashby, 2006). La prise en compte des mouvements oculaires, dans leur agencement spatial et temporel, permet d'obtenir un aperçu des processus utilisés par le lecteur pour construire la compréhension de l'information. On observe par exemple des différences, entre des experts d'un domaine et des sujets n'en ayant pas connaissance, dans les parcours oculaires en présence de schémas qui viennent étayer le contenu d'un texte portant sur le domaine. Des lecteurs ayant une connaissance faible vont avoir tendance à se référer visuellement plus fréquemment au schéma que les lecteurs experts du domaine (Strahm & Baccino, 2006). L'analyse des données oculaires peut également servir à classifier automatiquement les stratégies de lecture mises en œuvre (Pasqualotti & Baccino, 2014) : les auteurs utilisent un classifieur bayésien pour déterminer si le lecteur est en train de réaliser une lecture profonde, ou s'il est en train de réaliser une lecture superficielle (écrémage).

2.1.3.4.2.2 La mesure de la charge attentionnelle

La compréhension étant un processus actif qui mobilise les ressources mémorielles et l'attention du lecteur, il est nécessaire de mesurer sa charge attentionnelle pour pouvoir avoir une vue globale de ce phénomène.

Le niveau d'attention est un paramètre important à mesurer lors de l'interaction entre un sujet et un système, en psychologie et particulièrement en ergonomie. En effet, l'utilisation d'un système requiert de la part du sujet qu'il mobilise plus ou moins de ressources pour pouvoir effectuer la tâche qui lui est demandée. Lorsque la demande est très forte, le sujet doit mobiliser une grande partie de son attention, ce qui peut aller jusqu'à une surcharge. Cette surcharge peut avoir des effets plus ou moins graves selon le contexte. Dans les domaines de

la conduite automobile, de l'aéronautique ou de la conduite de machine dangereuse, la surcharge peut aboutir à des accidents, voilà pourquoi de nombreuses méthodes ont été développées dans le domaine de l'ergonomie pour pouvoir mesurer la charge attentionnelle.

Les méthodes pour mesurer la charge attentionnelle (ou la charge cognitive) peuvent prendre la forme d'une évaluation subjective par le sujet. Il décrit le niveau d'attention qu'il pense ressentir, c'est notamment le cas avec le questionnaire NASA-TLX (Hart & Staveland, 1988). L'évaluation subjective ne permet pas une mesure continue, et n'est pas toujours fiable, des mesures instrumentées ont donc été développées. Ces mesures peuvent passer par l'observation de données physiologiques : le rythme cardiaque peut par exemple augmenter avec le niveau d'attention (Dehais, Causse, Vachon, & Tremblay, 2012). Des marqueurs de l'activité cérébrale peuvent également être utilisés : l'activation du lobe frontal peut indiquer une hausse de la charge attentionnelle, on la mesure à l'aide de la spectroscopie proche infrarouge (Gateau, Durantin, Lancelot, Scannella, & Dehais, 2015). L'oculométrie permet également de mesurer la charge à partir de plusieurs types de relevés : avec la hausse de la charge attentionnelle, le diamètre de pupille augmente (Jainta & Baccino, 2010), la durée des clignements de paupière diminue (Benedetto et al., 2011), et la distribution spatiale des mouvements oculaires peut être affectée (Dehais et al., 2012). Cette liste de mesures n'est pas exhaustive puisque d'autres techniques existent (EEG, mesure du rythme respiratoire, potentiel électrodermal, ...). Enfin, la charge attentionnelle peut également être évaluée en mesurant la performance du sujet à une seconde tâche qui lui est demandée de réaliser pendant l'exécution de la tâche principale ; si les ressources du sujet sont fortement mobilisées pour réaliser la tâche principale, ses performances vont baisser à la seconde tâche, et inversement.

De nombreuses techniques permettent donc de mesurer la charge attentionnelle. En utilisant certaines d'entre elles pendant la lecture de textes, il est possible de quantifier les ressources mobilisées par le lecteur.

2.2 La dimension posturale de la lecture électronique

L'activité de lecture se matérialise par de multiples successions de mouvements des yeux. Le regard se porte sur un mot, puis chemine vers le suivant, revient parfois sur une partie du texte... C'est donc par ces déplacements des yeux que la lecture d'un texte est

rendue possible. Néanmoins, les capacités des mouvements oculaires ne sont pas illimitées. Ces mouvements sont contraints par la disposition des muscles oculaires et par les cavités orbitales. Ce sont les mouvements de la tête qui vont permettre d'augmenter l'espace accessible par le regard. La tête est elle-même contrainte par le tronc. Pour pouvoir atteindre l'information visuelle dans l'environnement, c'est donc le corps dans son ensemble qui se met en action, pour changer de point de vue, ou pour interagir avec l'environnement, par exemple en saisissant le support d'un texte pour le rendre visible.

La posture est donc un aspect à prendre en compte dans la lecture. Voilà pourquoi nous poserons quelques bases sur cette notion dans un premier temps. Ensuite, nous verrons comment envisager la posture dans le contexte de la lecture électronique.

2.2.1 La posture

La notion de posture réfère à deux aspects : la position du corps dans l'espace ainsi que son maintien. On ne considère donc pas les phénomènes où le corps est en déplacement : marche, course, sauts, ... Mais ce n'est pas pour autant que la posture correspond à une station strictement immobile du corps. La posture est en permanence régulée par des micromouvements ; par exemple, dans le cas de la station debout, le centre de masse du corps suit le mouvement d'un pendule inversé (Gage, Winter, Frank, & Adkin, 2004). La posture est donc le résultat de l'action simultanée des muscles du corps.

Concernant l'activité de lecture, les paramètres posturaux qui nous intéresseront le plus seront ceux de la tête. Elle peut se mouvoir selon trois axes (les mesures résultantes s'exprimant en degrés) :

- la tête peut s'élever et s'abaisser, tout en restant dans le plan sagittal du corps (voir *Figure 5*), on parlera de « l'élévation » ou de « l'abaissement » de tête (par exemple : lorsque l'on fait « oui » de la tête),
- la tête peut pivoter vers la gauche, ou la droite, tout en restant parallèle au plan transverse du corps (voir *Figure 5*), on parlera alors de « rotation latérale » (exemple : le « non » de la tête),
- la tête peut fléchir à gauche ou à droite, en restant dans le plan frontal (voir *Figure 5*), on parlera de « flexion latérale » ou de « roulis de tête ».

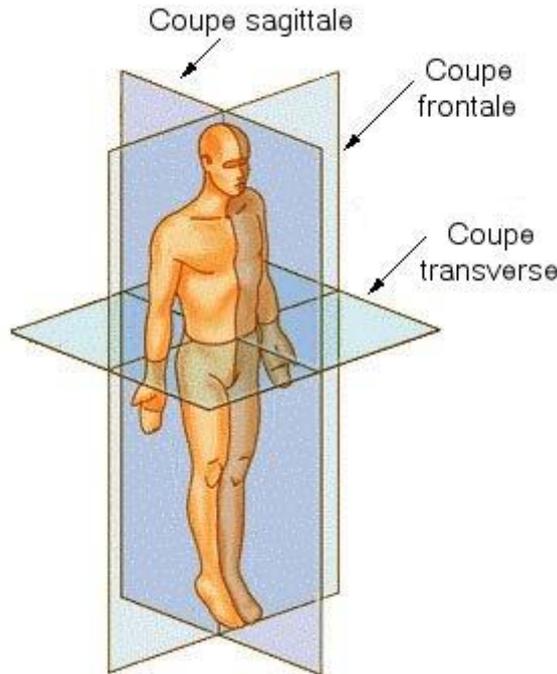


Figure 5. Schéma descriptif des trois plans de l'anatomie humaine (issu de l'article Wikipedia « Système de référence en anatomie »).

Pour pouvoir être mesurées, ces positions doivent être décrites relativement à un référentiel. Il est possible d'effectuer les mesures relativement à une position de référence du sujet. Pour définir cette référence, le sujet doit alors se mettre dans une position statique, être relaxé ; une possibilité parmi d'autres pour définir une position de tête contrôlée est de demander au sujet d'élever ou d'abaisser sa tête jusqu'à ce que le plan de Francfort (plan passant par les planchers des orbites et les extrémités hautes des méats acoustiques externes) soit parallèle au sol. Les paramètres de position de la tête peuvent alors être calculés comme étant les écarts angulaires à cette position de référence.

Si la mesure ne porte pas uniquement sur la tête, mais également sur un objet de l'environnement (un écran, une tablette,...), les données peuvent être exprimées relativement à cet objet. On s'intéressera particulièrement à la distance séparant les yeux de l'objet, parfois appelée « distance de lecture », « distance œil/écran »,... Il s'agit de la distance séparant le centre de la face visible de l'objet d'un point défini par rapport aux yeux. Pour des raisons pratiques, le point utilisé est à mi-distance des deux yeux, les points utilisés pour définir les yeux pouvant être les centres des pupilles, ou les centres de rotation des yeux. On appellera alors ce point central « l'œil cyclope ».

Si les mesures de posture sont effectuées par rapport à un objet, il est également possible de définir des mesures angulaires décrivant la position de l'objet relativement à la tête. Si l'on connaît l'angle d'élévation de la tête, ainsi que l'angle d'inclinaison du support, il est possible de calculer la différence relative d'inclinaison entre la tête et le support. Le même principe peut être utilisé pour calculer une différence relative de roulis ou de rotation latérale.

Enfin, dans le cas où le support de lecture est une tablette, une liseuse ou un smartphone, la position des mains peut également être pertinente à mesurer. En effet, ces appareils sont prévus pour être manipulés par le lecteur. La distance œil/écran va donc être en partie déterminée par la manière dont le sujet tient l'appareil.

Sur le plan technique, plusieurs solutions existent pour procéder aux mesures de la posture. La position du corps peut notamment être mesurée dans le temps à l'aide des techniques de suivi du mouvement (*motion capture* en anglais). La position des membres du sujet peut alors être recueillie à l'aide de capteurs optiques. Pour cela, il faut placer des marqueurs sur le corps du sujet, selon les besoins les marqueurs peuvent être placés sur la tête, et/ou sur différentes parties du corps pour définir des segments (bras, avant-bras, tronc, etc.) qui seront pertinentes pour l'analyse. Plusieurs caméras vont enregistrer la position des marqueurs. Un système central, qui connaît la position relative de chaque caméra suite à une phase de calibrage, va alors pouvoir interpoler la position de chaque marqueur dans l'espace. Les marqueurs peuvent être passifs, c'est-à-dire qu'ils réfléchissent la lumière émise par les caméras ; ou bien actifs, dans ce cas ils émettent leur propre lumière qui va être traitée par des cellules photosensibles. Des techniques n'utilisant pas de marqueurs sont également disponibles, c'est le cas de la Kinect de Microsoft : le système récupère directement la lumière infrarouge renvoyée par le corps.

Les systèmes optiques ne sont pas la seule solution pour mesurer le mouvement : dans le cas d'un système se basant sur les modifications du champ magnétique, les marqueurs réfléchissants peuvent être remplacés par des éléments métalliques perturbant un champ généré par des bobines. Les modifications du champ vont alors être mesurées, et utilisées pour inférer la position et l'orientation des éléments métalliques. Enfin, d'autres techniques peuvent compléter ou remplacer les solutions précédemment citées : il s'agit des plateformes de force, des capteurs de tension musculaires, des télémètres, des accéléromètres, des gyroscopes...

Ces techniques permettent donc de connaître la position de différents segments du corps à un instant précis ou bien au cours du temps. L'évolution temporelle des données de position donne des indications sur le maintien postural. Il s'agit donc de considérer des distributions de positions, de distances, d'angles... A minima, il est donc nécessaire de calculer au moins la variance de ces mesures pour obtenir un indice de la stabilité de la posture. Il est également possible de traiter le signal de manière à dégager les différentes fréquences qui le composent : on peut ainsi distinguer un mouvement de fond des micromouvements qui peuvent l'accompagner.

2.2.2 La posture pendant l'utilisation d'appareils électroniques

Bien que l'information qu'ils affichent soit dématérialisée, les dispositifs électroniques restent des objets physiques, qui peuvent être manipulés. Ces dispositifs ayant été rendus essentiels pour le travail de nombreuses personnes, l'ergonomie physique s'est rapidement intéressée à leur impact sur la posture, dans le but de réduire les risques encourus lors de leur utilisation. Cependant, les postures engendrées par l'utilisation de ces dispositifs n'ont pas qu'un impact physique, elles peuvent également être liées à des phénomènes visuels et cognitifs chez le lecteur.

2.2.2.1 Caractéristiques des dispositifs d'affichage électronique

Comme tout objet physique, un dispositif électronique peut être défini par plusieurs caractéristiques. Il a une position dans l'espace, une taille, un poids, une forme et une orientation. Les caractéristiques qui seront pertinentes à considérer dépendent du type de support. En effet, si le poids d'une tablette ou d'un smartphone est un paramètre à retenir dans l'étude de la lecture électronique, cette caractéristique est peu pertinente si l'on considère un écran d'ordinateur, ou un projecteur. Les caractéristiques sont donc à prendre en compte ou non selon le contexte d'utilisation : la manipulation d'un appareil portatif tel que le smartphone ne mobilisera pas de la même manière le corps que l'interaction avec un appareil placé dans l'environnement comme l'ordinateur portable.

Le dispositif d'affichage n'est cependant pas le seul objet physique qui joue un rôle sur la posture durant la lecture électronique. En effet, le dispositif d'affichage est généralement accompagné de périphériques permettant l'interaction. Si parfois l'objet d'affichage et l'objet d'interaction sont confondus, notamment pour les écrans tactiles, ce n'est pas toujours le cas et un écran d'ordinateur est généralement accompagné d'un clavier et d'une souris, qui vont eux même contraindre la posture de l'utilisateur. D'autres outils

d'interaction peuvent aussi être utilisés : un stylet, une manette de jeu, une webcam... Les postures des utilisateurs vont donc devoir s'adapter à ce système au cours de la lecture électronique.

2.2.2.2 Approche de l'ergonomie physique

La présence importante des dispositifs d'affichage électronique sur le lieu de travail, ainsi que les contraintes qu'ils amènent sont un enjeu fort pour les recherches en ergonomie physique. En effet, le travail sur écran peut engendrer des pathologies portant sur les articulations, les muscles et l'ossature des opérateurs. Les symptômes douloureux observables sont regroupés sous le terme de troubles musculosquelettiques (TMS). Ces troubles peuvent toucher le poignet et la main (utilisation répétitive de la souris ou du clavier), le cou (positionnement de l'écran), les épaules ou le dos. Les positions engendrées par le travail sur écran, ainsi que les aspects temporels associés (cadence, rigidité...) sont des facteurs qui peuvent augmenter le risque de TMS (Cail & Aptel, 2006).

Les risques amenés par l'utilisation des supports électroniques ont donc poussé à la mise en place d'environnement de travail plus adaptés. Le positionnement des différents éléments (écran et périphériques d'interaction) peut être défini par des spécifications visant à limiter ces risques. Des normes, telles que celles édictées par l'AFNOR (Association française de normalisation, 2002) spécifient les distances à l'écran, les angles de vision à respecter, dans le but d'améliorer le confort de travail à l'écran.

Se basant sur l'expérience négative de l'impact des écrans sur la posture, les ergonomes physiques ont rapidement transposé cette question aux tablettes lorsque celles-ci se sont répandues. Straker et al. (2008) ont réalisé une étude comparant la posture de jeunes enfants pendant l'utilisation d'une tablette, d'un ordinateur et de papier. Leurs résultats ont montré un impact négatif de l'utilisation de la tablette sur la posture car cette dernière engendrait un abaissement de tête plus important, et une position plus asymétrique comparée à la posture utilisée face à un ordinateur. L'utilisation de la tablette entraîne toutefois un effet positif : le corps est plus mobile, ce qui pourrait contrebalancer l'effet des postures à risque. Néanmoins, cette étude a été réalisée chez des enfants de 5 à 6 ans, avec une tablette de grande taille, posée sur une table ; il n'est donc pas aisé d'étendre ces résultats à la population générale. Le sur-abaissement de la tête lié à l'utilisation des tablettes a depuis été confirmé (Young, Trudeau, Odell, Marinelli, & Dennerlein, 2012), mais varie selon l'activité et l'environnement (regarder un film avec une tablette posée sur une table serait moins

problématique que d'écrire un message avec la tablette posée sur les genoux). Les mêmes auteurs ont également montré un risque quant à l'effet des tablettes sur l'utilisation des poignets (Young, Trudeau, Odell, Marinelli, & Dennerlein, 2013). Enfin, le fait de pouvoir plus librement orienter la tablette peut accentuer ou diminuer l'effet de sur-abaissement de tête lié aux tablettes (Albin & McLoone, 2014).

2.2.2.3 Posture et dimensions visuelle et cognitive de la lecture électronique

Les postures résultantes de la lecture électronique n'ont pas uniquement rapport à des aspects physiques. Celles-ci peuvent également être en lien avec la vision, ainsi qu'avec des aspects cognitifs. Ce sont ces deux points qui seront abordés ici.

2.2.2.3.1 Posture et vision

Comme nous l'avons vu précédemment, la posture joue un rôle dans la lecture, en effet, pour lire, les yeux réalisent des saccades, et ces déplacements sont souvent accompagnés par des mouvements de la tête (Lee, 1999). Pendant la lecture d'une ligne, la tête réalise une rotation latérale de gauche à droite, c'est-à-dire dans le sens du texte, pour ensuite revenir en position initiale, et recommencer le mouvement à la ligne suivante.

Dans le cas de la lecture électronique, l'orientation de la tête relative au support peut parfois être préjudiciable. C'est le cas de certains écrans LCD où la couleur et le contraste sont optimaux lorsque l'utilisateur est placé face à l'écran, et où la qualité de l'image diminue lorsque l'utilisateur s'éloigne de cette position idéale (Thomas, 2009). Dans d'autres cas, le changement d'orientation de l'écran peut être utile à une lecture confortable, par exemple lorsqu'il est réalisé dans le but de limiter les reflets de la lumière environnante sur l'écran (Albin & McLoone, 2014).

Nous avons vu dans la sous-section *2.1.1 La visibilité* que la capacité à discerner un objet est fonction de sa taille angulaire, c'est-à-dire à la fois de sa taille physique et de sa distance aux yeux du sujet. La distance de lecture est donc un élément important à considérer pour comprendre les aspects visuels de la lecture électronique. Durant l'utilisation de tablettes, la taille des caractères choisie influence la distance de lecture (Maniwa, Kotani, Suzuki, & Asao, 2013) : plus les caractères sont de petites taille, plus la distance entre les yeux et l'appareil est courte. D'après cette expérimentation, la distance aurait également tendance à diminuer avec la durée d'utilisation. Enfin, les auteurs ont montré que la distance

augmentait avec la taille de l'appareil (à taille de caractères égale) ; néanmoins, si les auteurs attribuent cet effet à la taille de l'écran, il pourrait être dû au poids de l'appareil, qui est un facteur confondu.

La distance de lecture impacte également la convergence. Plus un objet est proche des yeux, plus ceux-ci devront converger, pour assurer la fusion des images rétiniennes. La proximité de l'écran entraîne donc une contrainte musculaire pour les yeux. Cette contrainte est plus ou moins bien supportée par les sujets (Jaschinski, 2002). Le système visuel de certains sujets s'adapte efficacement aux distances proches, alors que cela peut être plus difficile pour d'autres. Pour ces derniers, on notera la présence de disparités de fixation plus importantes, c'est-à-dire que leurs axes visuels ne se croisent pas parfaitement au niveau du texte. Ce phénomène différencié est un facteur qui pourrait expliquer en partie les différences interindividuelles de distances de lecture. La distance de lecture, en lien avec la convergence est l'une des causes de la fatigue visuelle associée à la lecture sur écran (Rempel, Willms, Anshel, Jaschinski, & Sheedy, 2007).

2.2.2.3.2 *Posture et aspects cognitifs*

Le lien entre la posture et l'utilisation de dispositifs électroniques ne se limite pas à des aspects visuels. Des expérimentations ont pu montrer que la distance à l'écran peut diminuer lorsque le niveau d'attention du sujet augmente. Ces études ne portent pas sur la lecture, mais sur d'autres types de tâches. Une première étude a porté sur une tâche de simulation de défense aérienne, où le sujet doit identifier les avions à l'écran, et les éliminer s'ils sont hostiles (Balaban et al., 2004). Les auteurs ont montré que lorsque le nombre d'avions augmente, les sujets ont tendance à se rapprocher de l'écran. L'augmentation de la difficulté de la tâche entraîne donc un rapprochement à l'écran. Des résultats similaires ont été observés pour des tests cognitifs associés à différents niveaux de complexité (Qiu & Helbig, 2012). Dans cette étude, la charge attentionnelle liée à la réalisation de chaque tâche a été évaluée à l'aide du questionnaire NASA-TLX (Hart & Staveland, 1988). Les tâches sont associées à des niveaux attentionnels plus ou moins élevés (de la détection d'un stimulus visuel à des calculs mentaux). Les résultats montrent que la distance à l'écran est inversement corrélée au niveau de demande mentale, et que la variabilité de cette distance diminue avec cette demande. Ces résultats sont intéressants, néanmoins, dans la mesure où ils ne portent pas sur une activité de lecture électronique, la possibilité de les transposer à cette activité reste à questionner.

Dans la mesure où certains dispositifs d'affichage électronique sont devenus tactiles, et facilement manipulables, il est légitime de s'interroger sur le rôle de la position des mains sur les traitements cognitifs liés à la lecture électronique. Davoli, Du, Montana, Garverick et Abrams (2010) se sont intéressés à l'effet de la proximité des mains à l'écran sur le traitement sémantique de texte. Leurs expérimentations ont montré une baisse de la performance au traitement sémantique lorsque les mains sont apposées à l'écran. C'est le cas pour une tâche de classification de phrases cohérentes/incohérentes : les phrases incohérentes sont moins bien identifiées lorsque les mains sont contre l'écran. Une seconde expérience a porté sur la tâche de Stroop : les sujets doivent citer la couleur d'un mot. Les mots présentés sont des noms de couleurs et, de manière générale, les sujets répondent plus rapidement si le nom et la couleur du mot sont identiques, que lorsque le mot présenté ne correspond pas à sa couleur (par exemple « JAUNE » de couleur bleue). Les sujets traitent simultanément les informations sur la couleur du stimulus et l'information sémantique (le nom de la couleur). Lorsque ces deux informations ne sont pas congruentes, on observe un effet d'interférence. Les auteurs ont montré que cet effet d'interférence était réduit lorsque les mains des sujets sont contre l'écran, et attribuent cette amélioration de la performance à une diminution du traitement sémantique du texte présenté.

L'effet négatif de la position des mains a également été montré dans une tâche de détection de lettres (Le Bigot & Jaschinski, 2011). Néanmoins, un effet positif de la position des mains sur l'écran a été montré pour une tâche de plus bas niveau cognitif : la détection d'un stimulus visuel (Le Bigot & Grosjean, 2012). La présence d'une main, ou des deux contre l'écran peut améliorer la sensibilité aux stimuli, cette amélioration dépendant de la latéralité manuelle du sujet. La performance est la meilleure pour les gauchers lorsque leur main gauche tient l'écran, l'amélioration est un peu plus faible lorsque ce sont les deux mains ou la main droite qui sont dans cette situation. Pour les droitiers, la performance est la meilleure lorsque les deux mains sont contre l'écran. La performance est plus faible pour la main droite seule, encore plus faible pour la main gauche, et enfin la performance est la moins élevée lorsque les mains ne sont pas apposées à l'écran (mais l'effet est moins important que chez les gauchers). Les performances des sujets gauchers ne sont pas symétriques à celles des droitiers. Ce résultat n'est pas surprenant dans la mesure où l'activation du cortex prémoteur n'est pas strictement symétrique entre les gauchers et les droitiers dans une autre tâche visuelle : la lecture de caractères (Longcamp, Anton, Roth, & Velay, 2005). Cela pourrait être une piste pour expliquer pourquoi l'action des mains n'a pas le même effet chez

les droitiers et chez les gauchers. L'ensemble de ces résultats ne permet pas de tirer une conclusion générale sur un éventuel effet positif ou négatif de la position des mains pour la lecture électronique. Ils montrent néanmoins que la posture, à travers la position des mains, peut avoir des conséquences sur les traitements cognitifs des textes.

Chapitre 3 : Problématique

Comme nous venons de le voir dans le chapitre précédent, la lecture est un processus complexe mobilisant simultanément plusieurs niveaux de traitement. Nous partons donc du principe que pour être évaluée exhaustivement, la lecture électronique doit être mesurée pour chacune des dimensions qui lui sont associées.

Or, la plupart des études portant sur la lecture électronique ne considère généralement qu'un seul niveau de la lecture : certaines études portent sur les mouvements oculaires (Siegenthaler et al., 2011), d'autres sur la compréhension (Salmerón & García, 2011) et d'autres sur la posture (Young et al., 2012),... Ces études n'offrent donc pas une vision globale des performances de lecture associées à l'utilisation d'un dispositif électronique donné.

Les performances de lecture varient avec le matériel et la méthode utilisés : le texte employé par exemple va modifier la vitesse de lecture (2.1.2 *La lisibilité*) et la qualité de la compréhension (2.1.3 *La compréhension*). Comme les auteurs construisent généralement un matériel expérimental dédié à chaque étude, les résultats obtenus par une étude spécifique ne peuvent donc pas être comparés directement avec ceux d'une autre étude. L'absence d'outils réutilisables pour effectuer des mesures de performances de lecture est donc préjudiciable à l'aboutissement de conclusions claires sur les études comparatives entre les supports.

Si des tests permettant d'évaluer globalement la lecture de manière standardisée existent, ceux-ci sont à chercher dans le contexte des sciences de l'éducation, dans des programmes tels que PISA (Schleicher, Zimmer, Evans, & Clements, 2009). Ces tests standardisés permettent de comparer le niveau d'aptitude à la lecture des sujets, mais ne sont pas conçus pour permettre des mesures répétées sur les mêmes sujets afin d'isoler l'effet d'un support.

La problématique de cette thèse peut se résumer sous la forme d'une question : « comment comparer la performance de lecture entre plusieurs supports numériques ? »

De cette problématique découlent plusieurs questions sous-jacentes :

- Comment évaluer globalement la lecture électronique ? Nous tenterons de répondre à cette question en adoptant une approche multidimensionnelle de la lecture électronique. En évaluant la lecture électronique pour chaque dimension

qui y est associée (les trois dimensions classiques (Baccino, 2004) et la dimension posturale), nous essayerons d'apporter une solution au problème opposant la complexité de la lecture électronique aux approches unidimensionnelles qui sont généralement mises en œuvre.

- Comment pouvoir comparer les performances de lectures associées à plusieurs dispositifs d'affichage ? Cette seconde question rend compte de la nécessité de pouvoir effectuer des mesures répétées à partir des tests de lecture électronique. Pour y répondre, nous chercherons à adopter une approche s'inspirant de la psychophysique. La psychophysique est une branche de la psychologie visant à définir des seuils perceptifs associés à des grandeurs physiques de stimuli, par exemple en déterminant à partir de quel volume en décibels un son est perceptible. L'intérêt de cette approche est d'offrir un contrôle fort du matériel expérimental, et de permettre grâce à ce matériel contrôlé de pouvoir effectuer des mesures répétées, en ne faisant varier que les facteurs d'intérêt. S'il est assez facile d'imaginer comment l'approche psychophysique pourrait permettre d'évaluer la visibilité d'un écran (variation du niveau lumineux des pixels), cela est moins évident pour l'évaluation de processus de plus haut niveau tels que la compréhension. Néanmoins, nous nous attacherons dans les chapitres suivants à proposer des tests dont le matériel est suffisamment contrôlé pour permettre d'effectuer des mesures répétées.
- Comment rendre compte synthétiquement de la performance de lecture électronique sur un support donné ? Cette question est liée au fait qu'une approche multidimensionnelle implique un nombre important de mesures et de données. Pour pouvoir jauger simplement la performance de lecture associée à un support, que cela soit au niveau global ou au niveau de chaque dimension, il est nécessaire de construire des indicateurs. Nous mettrons donc en place des indicateurs qui permettent de résumer les résultats issus des différents tests pour chaque dimension de la lecture électronique. Enfin, un indicateur global, le « Quotient de Lecture Electronique » sera construit pour synthétiser ces indicateurs unidimensionnels.

Les quatre prochains chapitres porteront donc sur les quatre dimensions que nous avons sélectionnées pour définir la lecture électronique. Nous y proposerons des tests qui

devront pouvoir être utilisés conjointement, afin d'évaluer globalement le sujet de cette thèse, et qui seront répétables de manière à permettre la comparaison entre différents supports. L'avant-dernier chapitre présentera une mise en application de ces tests qui permettra de valider leur utilisation vis-à-vis de notre problématique.

Chapitre 4 : L'évaluation de la visibilité

Nous avons expliqué dans la sous-section 2.1.1 *La visibilité* que la visibilité correspond à la capacité dont dispose un système d'affichage pour transmettre une information visuelle avec efficacité. Les caractéristiques qui sont liées à la visibilité d'un dispositif d'affichage doivent être comprises en tenant compte du traitement perceptif humain. Ce nouveau chapitre vise à décrire concrètement la méthode que nous avons mise en place pour mesurer ces caractéristiques.

4.1 Les méthodologies d'évaluation de la visibilité

La caractérisation d'un écran consiste dans le relevé de ses caractéristiques colorimétriques précises. Classiquement, elle est réalisée à l'aide d'un luminancemètre. Un colorimètre peut également être nécessaire si on souhaite étudier la restitution des couleurs de l'écran. Dans notre cas, nous ne nous intéresserons qu'aux niveaux de gris, nous nous passerons donc du colorimètre.

La caractérisation consiste en un ensemble de mesures physiques visant à évaluer pour chaque niveau d'entrée la luminance de sortie. Ces mesures sont parfois réalisées sans tenir compte de la luminosité ambiante, c'est le cas lorsque le luminancemètre utilisé est à courte portée, directement apposé à l'écran, ou lorsque la mesure est effectuée sans autre éclairage que celui propre à l'écran, comme préconisé par l'association de normalisation VESA (Video Electronics Standards Association, 2001) afin de procéder à une mesure très reproductible et qui n'est pas impactée par les variations d'ambiance lumineuse⁹. Nous avons montré précédemment comment l'éclairage ambiant et les éventuels reflets peuvent impacter les contrastes et la luminance de sortie des dispositifs d'affichage. L'utilisation habituelle des écrans a lieu dans des contextes où un éclairage artificiel ou naturel est présent, nous avons choisi de réaliser une mesure qui se rapproche le plus des conditions auxquelles les lecteurs sont coutumiers : dans des situations où l'éclairage est présent. Nous avons également précisé

⁹ La mesure de luminance dans le noir n'est pas recommandée par tous les organismes de normalisation, l'IEC (International Electrotechnical Commission) recommande notamment de réaliser les mesures avec une source de lumière externe, tout en mesurant l'ambiance lumineuse à l'aide d'un luxmètre (voir IEC 62563-1:2010 Medical electrical equipment – medical image display systems – part 1: evaluation methods. International Standards Organisation, Geneva).

que les capacités du système visuel humain à détecter les contrastes sont limitées et ne dépendent pas linéairement de l'intensité lumineuse. Un des buts de notre mesure sera donc, en plus de connaître la fonction de transfert physique de l'écran (du niveau RGB d'entrée, vers la luminance de sortie), de pouvoir tenir compte des facteurs propres au système visuel de l'observateur.

Afin de connaître les fonctions de transfert des différents dispositifs d'affichage que nous testons, nous utilisons une procédure classique : pour chaque niveau de gris RGB pouvant être affiché, nous effectuons une mesure de luminance à l'aide d'un luminancemètre.

Pour mesurer les aspects de visibilité en relation avec le système visuel, des tâches de détection peuvent être utilisées (Diarra et al., 2011). Il s'agit d'afficher à l'écran un fond uni ainsi qu'un éventuel stimulus moins lumineux que le fond. La tâche consiste pour l'observateur à déterminer si un stimulus est apparu ou non. Une procédure psychophysique est utilisée de manière à ce que tant que le sujet donne des bonnes réponses, on réduit l'écart de luminance entre le fond et le stimulus (diminution du contraste), et inversement, lorsque le sujet donne une réponse erronée, l'écart est augmenté (augmentation du contraste). La procédure tend à converger vers le seuil de détection du stimulus. Néanmoins les auteurs n'ont mesuré ce seuil que pour une seule valeur de luminance de fond, alors que les seuils de détection varient avec la luminance du fond, comme nous l'avons expliqué dans la sous-section 2.1.1 *La visibilité*. L'avantage de la méthode tient dans le fait qu'elle rend compte des capacités du système visuel afin d'évaluer la visibilité du point de vue de l'observateur.

D'autres auteurs ont proposé des méthodes pour effectuer une caractérisation d'écran en intégrant la perception de l'utilisateur (Gille et al., 2004). Leur mesure consiste à afficher en arrière-plan chaque niveau de gris qui peut être généré (soit un total de 256 arrière-plans) ; sur chaque arrière-plan, ils affichent 8 disques dont les niveaux de gris vont du niveau de fond + 1 pour le premier disque jusqu'au niveau du fond + 8 pour le dernier. Les sujets doivent indiquer le premier disque qu'ils parviennent à détecter (donc le moins contrasté). Les auteurs obtiennent ainsi les seuils de discriminabilité par des observateurs pour l'ensemble des niveaux de gris affichables. Leur but est de réaliser la caractérisation sans instrumentation particulière, uniquement au moyen des réponses des observateurs. Enfin, ils cherchent à inférer, à partir des données recueillies, la fonction de transfert (physique) de l'écran, mais cette procédure n'a pas permis de recréer la fonction pour l'ensemble des niveaux testés. En effet, une objection peut être apportée à cette dernière étape : la méthode permet de

déterminer les niveaux de gris consécutifs indiscernables, i.e. on sait pour quels niveaux consécutifs de gris l'observateur ne voit pas de différence (généralement le noir et les gris très foncés, le blanc et les gris très clairs). Cependant la méthode ne permet pas de déterminer à quel point deux niveaux consécutifs sont discernables. Si un sujet parvient à discerner un niveau x d'un niveau $x+1$, on sait que ces niveaux sont distincts, mais on ne peut pas savoir l'amplitude de la différence ressentie (les deux niveaux consécutifs peuvent paraître très similaires, avec une très faible différence, ou extrêmement différents). Cette amplitude de différence ressentie peut pourtant affecter la qualité d'affichage d'un écran. Dans le cas où deux niveaux sont très distincts, cela va avoir un effet sur le rendu d'une image en créant une coupure nette entre les deux niveaux consécutifs : un artéfact graphique. L'absence de cette information peut expliquer pourquoi les auteurs ne sont pas parvenus à recréer des fonctions de transfert pour l'ensemble des niveaux de gris. Cette limite impacte surtout l'affichage d'images. Pour la lecture électronique et l'affichage de caractères, elle ne s'applique pas, car notre objectif est de savoir si les niveaux de gris sont discernables (par exemple pour des pixels utilisés dans l'anti-aliasing).

Une seconde limite peut être problématique pour l'évaluation de la visibilité : à la manière d'une tâche d'ajustement, le sujet détermine librement l'écart RGB nécessaire pour détecter un niveau de gris, et cette mesure peut être impactée par les stratégies du sujet. En effet, si le sujet adopte une stratégie libérale, il aura tendance à donner une réponse, même s'il n'est pas sûr de voir le disque. Au contraire, si le sujet adopte une stratégie conservatrice, il ne donnera des réponses que lorsqu'il voit des disques très contrastés, même s'il pourrait parvenir à détecter un contraste moins élevé. En donnant plus de réponses, les sujets adoptant une stratégie libérale vont obtenir des scores meilleurs que ceux des sujets adoptant une stratégie conservatrice. A cause de ces stratégies, il est donc possible d'observer des différences qui ne sont dues ni à l'écran testé, ni au système visuel de l'observateur. Nous allons voir dans la prochaine section des solutions qui peuvent être apportées pour palier cette limite.

4.2 Technique de caractérisation psychométrique

Cette partie présente une méthode d'évaluation de la visibilité que nous proposons en nous basant majoritairement sur la technique proposée par Gille et al. (2004). L'objectif est de connaître, pour chaque niveau de gris de l'arrière-plan, l'écart RGB (différence absolue entre le niveau d'un stimulus et le niveau de fond) pour que le stimulus soit détectable.

Pour éviter les risques de variabilité liés à des stratégies différenciées des observateurs, nous avons pris le parti d'adopter une approche psychophysique. Un disque dont le niveau RGB est choisi relativement au niveau d'arrière-plan est affiché aléatoirement en haut, en bas, à gauche ou à droite de l'écran. L'observateur doit déterminer la position de ce disque. Pour éviter les stratégies conservatrices, nous avons opté pour un paradigme à choix forcé (paradigme 4-SAFC, de l'anglais *Four Spatial Alternative Forced Choice*, choix forcé à 4 alternatives spatiales), où l'observateur doit donner une réponse même s'il ne voit pas le disque ou s'il n'est pas certain de sa réponse. Tous les observateurs sont ainsi amenés à adopter une stratégie plutôt libérale. La position est choisie aléatoirement entre les 4 possibilités. Afin d'éviter que le sujet ne puisse retenir les 4 positions possibles, les positions associées aux 4 directions sont légèrement variables (cf. *Figure 6*). Le diamètre du disque est fixé à 1° d'angle visuel.

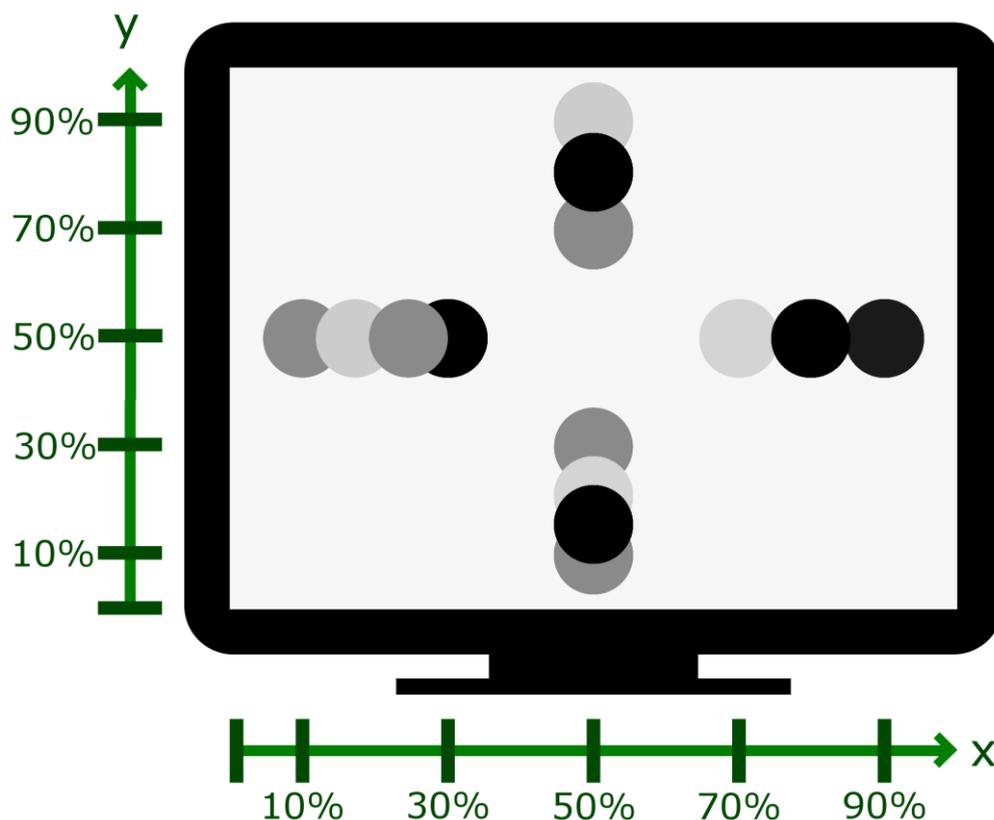


Figure 6. Représentation des positions possibles des disques à l'écran. Le centre de chaque disque est toujours placé au milieu de la hauteur de l'écran pour les directions gauche et droite, et au milieu de la largeur pour haut et bas. Pour la direction à gauche, l'abscisse du centre est située entre 10 et 30% de

la largeur de l'écran, entre 70 et 90% pour la droite. Pour la direction en bas, l'ordonnée du centre est située entre 10 et 30% de la hauteur de l'écran, entre 70 et 90% pour la direction en haut. Les niveaux de gris sont exagérés et donnés uniquement à titre d'illustration.

Le niveau de gris du disque est choisi relativement au niveau de celui de l'arrière-plan. Ce niveau est déterminé par une procédure adaptative qui module le niveau en fonction des réponses précédentes de l'observateur : si l'observateur donne la direction où le disque se situe effectivement, le disque va être moins contrasté par rapport au fond, et inversement si l'observateur donne une mauvaise réponse. Nous avons implémenté la procédure adaptative bayésienne Ψ de Kontsevich & Tyler (1999). Cette dernière sert à déterminer quelle fonction psychométrique correspond au mieux aux réponses données par le sujet (i.e. la fonction qui associe à l'intensité du stimulus x , la probabilité de bonne réponse $f(x)$). On stipule d'abord l'ensemble des fonctions psychométriques possibles, chacune étant définie par deux paramètres : a le seuil et b la pente de la fonction. Nous avons utilisé la loi de Weibull pour définir ces fonctions (Weibull, 1951) :

$$f_0(x) = 1 - (1 - h) * e^{-\left(\frac{k * x}{a}\right)^b}$$

avec h : la probabilité de bonne réponse au hasard (étant donné que le sujet doit choisir une réponse parmi 4, cette probabilité est de 0,25), a le seuil, b la pente de la fonction et k définit par :

$$k = \left(-\log\left(\frac{1-h}{2}/(1-h)\right) \right)^{\frac{1}{b}}$$

Afin de tenir compte des éventuelles mauvaises réponses alors que le sujet est parvenu à voir le stimulus, un correctif a été appliqué :

$$f(x) = 0,99 * f_0(x) + 0,005$$

$$\lim_{x \rightarrow +\infty} f(x) = 0,995$$

De cette manière,

$$\lim_{x \rightarrow +\infty} f_0(x) = 1$$

alors que

Dans notre étude, nous ne nous intéressons qu'au seuil, et pas à la pente¹⁰. Nous avons donc fixé pour chaque fonction la pente à 10. Les seuils testés vont de 1 à 20 avec un pas de 1 (soit un total de 20 fonctions testées).

La méthode Ψ attribue à chaque fonction une probabilité d'être la vraie fonction psychométrique du sujet. En début de test, nous distribuons uniformément ces probabilités (1 / 20 pour chaque fonction). La méthode détermine quel niveau du stimulus doit être testé pour obtenir le plus d'information. On effectue la mesure pour le niveau déterminé, et les probabilités de chaque fonction sont recalculées à partir de la réponse du sujet. La procédure est répétée de manière à ce que la probabilité d'une fonction (et donc d'un seuil) soit suffisamment élevée relativement aux autres pour pouvoir décider où se situe la vraie fonction psychométrique. Pour notre test, nous affichons 8 niveaux de gris par arrière-plan. On estime alors l'écart RGB en calculant la moyenne des seuils testés, pondérée par les probabilités de chaque seuil.

On utilise la méthode Ψ pour déterminer le seuil de détection RGB pour chaque niveau d'arrière-plan testé. Pour limiter la durée du test, nous avons choisi de ne pas évaluer les 256 niveaux de gris. Le test de Gille et al. (2004) permet de détecter les niveaux de gris non discernables, et, d'après leurs données, ces niveaux se situent surtout pour les gris foncés, et pour les gris clairs. Nous n'évaluons donc pas les niveaux moyens, et nous limitons aux intervalles [0 ; 45] et [210 ; 255]. Les intervalles sont parcourus avec un pas de 5. Les écarts RGB correspondant aux niveaux moyens (entre 50 et 205) sont artificiellement fixés à 1 (on suppose que chaque niveau de gris est discernable d'un autre pour les valeurs moyennes). Les données sont ensuite décrites par un polynôme de degré 2 à l'aide de la méthode des moindres carrés. Pour synthétiser les données obtenues sous forme d'un score unique, nous calculons l'aire sous la courbe de la fonction polynomiale par la méthode de Simpson (voir *Figure 7*).

¹⁰ Nous avons fait ce choix car le nombre de seuils évalués était élevé, et l'estimation de la pente nécessite un grand nombre d'essais. La mesure aurait donc nécessité trop de temps pour estimer les deux paramètres en même temps.

Caractérisation psychométrique

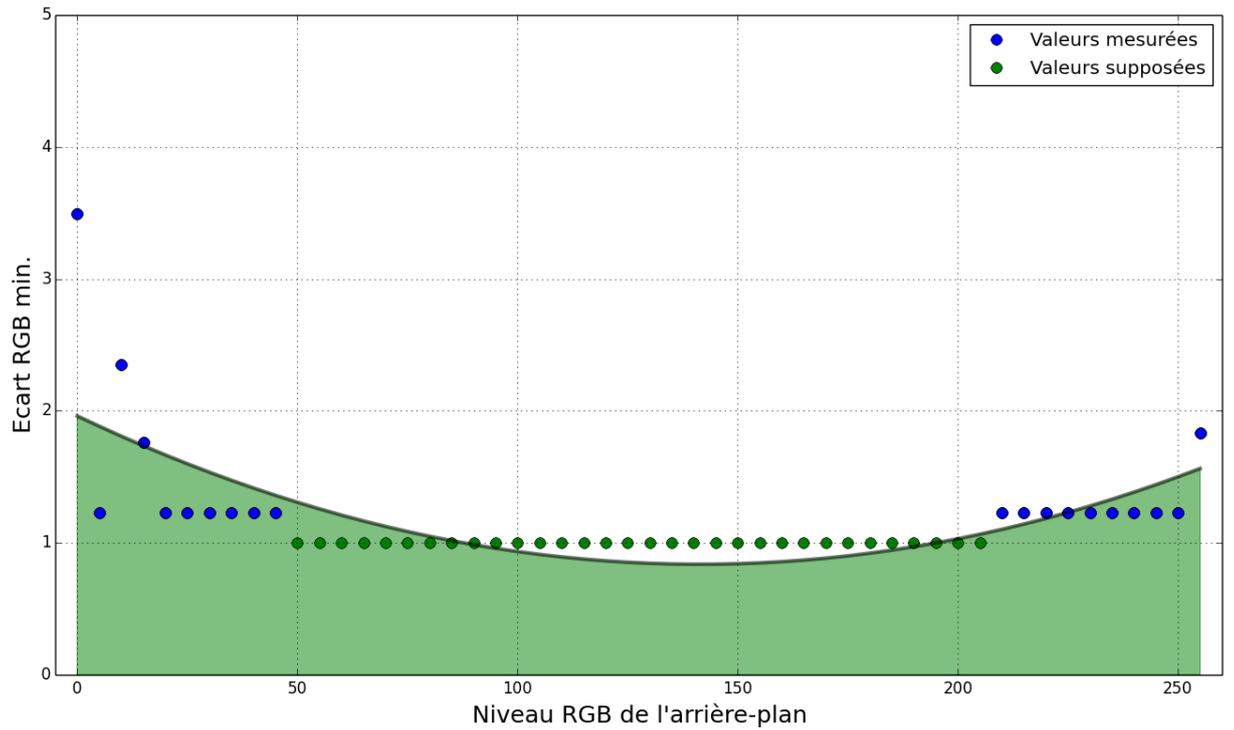


Figure 7. Illustrations de données pouvant être recueillies à l'aide du test de caractérisation subjective. Les données mesurées et supposées sont décrites par une fonction polynomiale dont on calcule l'aire sous la courbe (correspondant à la surface en vert).

La technique proposée ici permet donc de fournir une mesure liée au profil de luminance d'un dispositif d'affichage électronique, tout en intégrant la capacité de perception des contrastes propre au sujet.

Chapitre 5 : L'évaluation de la lisibilité

Ce chapitre décrit les tests que nous avons adaptés ou conçus pour rendre compte de la lisibilité d'un dispositif d'affichage électronique. Comme il l'a été mentionné dans 2.1.2 *La lisibilité*, l'acquisition de l'information textuelle par le lecteur s'effectue à plusieurs niveaux : sont traités les lettres, les mots et les phrases. Nous présenterons dans une première section deux tests permettant d'évaluer l'identification des lettres et celle des mots, puis une seconde section concernera l'évaluation de la lecture des phrases par la mesure de la vitesse de lecture. La technique oculométrique, bien qu'étant une manière d'évaluer la lisibilité, n'est pas présentée dans ce chapitre, car elle ne forme pas un test en soi. Néanmoins, nous utiliserons cette technique par la suite dans des expériences 7.1 *L'effet de la posture sur la lecture* et 7.2 *L'effet de la lecture électronique sur la posture*.

5.1 Test d'identification de lettres et test de décision lexicale

Cette partie présente deux méthodes visant à évaluer la lisibilité des écrans quant à l'affichage des lettres, et des mots. Nous avons mutualisé la description de ces deux tests, car leurs formats sont proches. Le but de ces tests est de mesurer avec quelle rapidité et quelle efficacité le lecteur peut identifier des lettres et des mots affichés. Le test d'identification de lettres consiste en une tâche où le sujet doit déterminer le plus rapidement possible si une lettre cible est présente ou non dans un groupe de lettres. Le test de décision lexicale consiste en une autre tâche où le lecteur doit déterminer le plus rapidement possible si le groupe de lettres est un mot qui existe ou non en langue française.

5.1.1 Sélection des stimuli

Pour chacun des deux tests, les stimuli sont sélectionnés de manière à garantir une certaine homogénéité entre les passations successives, et de manière à générer à chaque fois deux conditions de difficultés différentes.

5.1.1.1 Stimuli pour le test d'identification de lettres

L'objectif du test d'identification de lettres est de mesurer, sur chaque support, la capacité du sujet à identifier rapidement et efficacement des lettres affichées à l'écran dans un non-mot (une suite de lettre qui ne forme ni un mot, ni un pseudo-mot, c'est-à-dire qui ne suit pas les règles de construction de mot en français). Pour construire ce test, il est nécessaire de définir les lettres et groupes de lettres qui vont être présentés.

Afin de pouvoir définir des conditions où la difficulté est élevée ou faible, nous nous sommes appuyés sur la notion de « distance perceptuelle » des caractères alphabétiques : les 26 caractères de l'alphabet sont visuellement plus ou moins similaires les uns aux autres. Courrieu, Farioli et Grainger (2004) appellent « distance perceptuelle » le degré de différence visuelle entre deux lettres. Cette distance est calculée à partir des durées de réponses recueillies sur une tâche de discrimination : deux lettres sont présentées au sujet, celui-ci doit déterminer le plus rapidement possible si ces deux lettres sont identiques ou distinctes. Plus les lettres sont visuellement différentes, plus la réponse est rapide. En soustrayant une durée de réponse de base, et en calculant l'inverse de cette durée de réponse corrigée, les auteurs obtiennent la valeur de la distance perceptuelle entre les différents caractères.

Pour notre test, nous définissons plusieurs groupes de lettres à partir des données de distance perceptuelle. Dans un premier temps, les groupes de lettres similaires sont définis comme étant les groupes d'au moins trois lettres dont la distance perceptuelle est inférieure à 90. Nous obtenons :

- Groupe de lettres similaires 1 : [p, q, g],
- Groupe de lettres similaires 2 : [f, i, l, t].

Dans un second temps, nous définissons un ensemble de lettres visuellement éloignées des lettres précédemment sélectionnées. Nous sélectionnons toutes les lettres dont la distance à chacune des lettres des groupes similaires est strictement supérieure à 100, et obtenons :

- Groupe de lettres éloignées : [a, c, e, m, n, o, r, s, u, v, w, x, y, z].

Tous les non-mots que nous allons construire seront composés uniquement de lettres appartenant à l'un des trois groupes.

Le processus de création des non-mots décrit dans ce paragraphe et les suivants est résumé dans la *Figure 8*. Pour créer les non-mots, nous sélectionnons d'abord une lettre cible parmi les deux groupes de lettres similaires (chaque lettre des groupes similaires est sélectionnée le même nombre de fois au total). La moitié des non-mots contiendra cette lettre cible, l'autre ne la contiendra pas.

La longueur du non-mot est sélectionnée aléatoirement entre 6 et 8 lettres. On compose le non-mot avec des lettres appartenant au groupe de lettres éloignées. Afin d'éviter

que par le hasard des sélections de lettres, le groupe de lettres forme un pseudo-mot, voire un mot, nous n'insérons que des paires de lettres successives dont la probabilité d'apparition en français est très faible. Nous avons utilisé les valeurs de Lexique 2 (New, Pallier, Brysbaert, & Ferrand, 2004) : la base « bigramme » recense le nombre d'occurrences de chaque paire de lettres dans un ordre donné (chaque bigramme) par million de mots. Nous ajoutons une lettre au non-mot uniquement si le nombre d'occurrence du nouveau bigramme formé est strictement inférieur à 50 pour un million.

Pour constituer les non-mots correspondant aux deux conditions (facile et difficile), nous insérons ou non des lettres similaires à la lettre cible dans le non-mot. La moitié des non-mots est destinée à la condition facile : ils ne contiennent que des lettres éloignées (et une fois sur deux la cible). L'autre moitié contient des distracteurs : des lettres similaires à la lettre cible (qu'elle soit présente ou non). Le nombre de lettres similaires est choisi aléatoirement entre 1 et la longueur du non-mot – 1. Les lettres similaires sont alors choisies aléatoirement dans le groupe de lettres similaires de la lettre cible, et placées aléatoirement dans le non-mot (en remplacement des lettres éloignées précédemment sélectionnées). Enfin, si le non-mot doit contenir la cible, celle-ci est placée aléatoirement dans le non-mot (en remplacement d'une lettre éloignée).

5.1.1.2 Stimuli pour le test de décision lexicale

L'objectif du test de décision lexicale est de mesurer sur chaque support la capacité du sujet à identifier rapidement et efficacement des mots affichés à l'écran. Des mots et des pseudo-mots (suite de lettres qui respectent les règles de construction de la langue, mais qui ne sont pas des mots du lexique) vont être présentés au sujet, il devra déterminer si le stimulus présenté est un mot de la langue. Pour construire ce test, il est nécessaire de définir les groupes de mots et pseudo-mots qui vont être présentés.

Ces mots et pseudo-mots doivent également permettre de mettre en place deux conditions : une condition où le choix est rapide (condition facile), et une condition où le choix est plus long (condition difficile). En général, les chercheurs construisent ces deux conditions en utilisant comme moyen de catégorisation les fréquences lexicales des mots affichés (Balota & Chumbley, 1984; Grainger, 1990). Mais ce choix peut présenter deux inconvénients :

- la fréquence lexicale des mots est la variable qui explique majoritairement les variations de temps de réaction dans une tâche de décision lexicale, mais la contribution de cette variable est $R^2 = 37,4\%$ (Ferrand et al., 2010) : une grande partie de la variation reste inexpliquée,
- la fréquence lexicale est une variable qui s'applique uniquement aux mots (la fréquence lexicale d'un pseudo-mot est de 0 par définition), il n'est donc pas possible de construire une condition facile et une condition difficile pour les pseudo-mots à partir de cette variable.

Pour pallier ces inconvénients, nous avons utilisé directement les valeurs de temps de réaction recueillies par Ferrand et al. (2010). Les auteurs ont mesuré les temps de réponses et les taux d'erreurs dans une tâche de décision lexicale sur plus de 75 000 mots et pseudo-mots, ces données sont regroupées dans la base appelée « French Lexicon Project » (FLP). Pour limiter l'effet de longueur du mot, nous n'avons sélectionné que des mots de 6 à 8 lettres. En effet, d'après les auteurs, ce sont ces longueurs qui aboutissent aux temps de réponse les plus courts et les moins variables.

Chaque mot ou pseudo-mot du FLP a été testé par au moins 25 sujets. Néanmoins, si un sujet mettait plus de 4 secondes pour répondre, ses données n'étaient pas traitées. Ainsi, les valeurs de temps de réponse et d'erreur ne sont pas toutes obtenues pour 25 sujets. Nous

avons limité notre sélection aux mots ayant reçu au moins 20 réponses, et aux pseudo-mots en ayant eu au moins 16, afin que les moyennes de temps de réaction et d'erreurs soient issues d'un échantillon suffisamment fourni. Pour limiter les mots ou pseudo-mots pour lesquels l'appartenance à la langue pourrait être ambiguë, nous avons rejeté les mots pour lesquels le taux d'erreur était supérieur à 22,72% (par exemple : « varech », « prolo ») et les pseudo-mots pour lesquels le taux d'erreur était supérieur à 22,22% (par exemple : « congré », « lisée », « hôta »). Pour déterminer ces deux seuils, nous avons procédé à une analyse en k-moyennes : les taux d'erreurs ont été classés dans deux groupes, et seuls les groupes associés aux plus faibles taux d'erreurs ont été conservés.

Afin de construire les deux conditions expérimentales, nous avons décidé de sélectionner des mots et des pseudo-mots associés à des temps de réponse courts pour la condition facile, et à des temps de réponse longs pour la condition difficile. Pour cela, nous avons classé d'une part les mots, et d'autre part les pseudo-mots du FLP à partir des temps de réponse centrés-réduits. En élaborant nos groupes de mots et pseudos-mots à partir de ce classement, nous pouvons contrôler la difficulté ; les groupes faciles sont constitués de mots et pseudos-mots correspondant au second décile du classement, et les groupes difficiles correspondent à l'avant-dernier décile :

- Pour les mots, le groupe « facile » comprend les mots dont le temps de réaction centré-réduit appartient à l'intervalle : [D1 = -0,593579 ; D2 = -0,481909] (second décile ; exemples : « acteurs », « vieillir ») ; pour le groupe « difficile », ce sont les mots de l'intervalle : [D8 = 0,141735 ; D9 = 0,341947] (avant-dernier décile ; exemples : « chairs », « disions »),
- Pour les pseudo-mots, le groupe « facile » correspond à l'intervalle : [D1 = -0,536748 ; D2 = -0,419244] (second décile ; exemples : « urmelors », « cetivers ») ; le groupe « difficile » correspond à l'intervalle : [D8 = 0,214556 et D9 = 0,420520] (avant-dernier décile ; exemples : « resclame », « éganer »).

A partir des mots sélectionnés, nous avons constitués 10 sets composés de 30 mots (dont 15 du groupe « facile » et 15 du groupe « difficile ») et de 30 pseudo-mots (15 faciles et 15 difficiles). Chaque mot et pseudo-mot est sélectionné aléatoirement dans le groupe correspondant. L'ordre des mots et pseudo-mots dans chaque set est défini aléatoirement.

5.1.2 Procédure

La séquence d’affichages qui constitue chacun des deux tests est détaillée dans la *Figure 9*.

Pour le test d’identification de lettres, la lettre cible est d’abord affichée pendant une seconde. Vient ensuite une croix de fixation affichée au centre de l’écran pendant une demi-seconde. La croix est remplacée par le non-mot, qui reste affiché jusqu’à ce que le sujet donne une réponse. Une fois la réponse donnée, un écran blanc s’affiche pendant une demi-seconde. La consigne donnée au sujet est de répondre le plus rapidement « oui » si la cible fait partie du groupe de lettres affiché, et « non » si la cible est absente (la modalité de réponse dépend du support testé, voir 8.2.4 *Interfaces de réponse*).

La séquence d’affichage est la même pour le test de décision lexicale ; à la distinction près qu’il n’y a pas de lettre cible et que le stimulus est un mot ou un pseudo-mot. . La consigne donnée au sujet est de répondre le plus rapidement « oui » si le mot affiché existe, et « non » si le mot n’existe pas¹¹ (la modalité de réponse dépend du support testé, voir 8.2.4 *Interfaces de réponse*).

Un test comporte 60 essais au total. Pour les deux tests, un écran de pause d’une durée de 10s apparaît tous les 15 essais. Le sujet a donc trois courtes pauses par test. Un message prévient le sujet de la reprise du test trois secondes avant celle-ci.

¹¹ Les termes utilisés dans la consigne ne sont pas tout à fait corrects, puisque l’écran n’affiche des « mots » que la moitié du temps (et des pseudo-mots l’autre moitié). Nous avons opté pour une consigne facile à comprendre, plutôt que de faire apprendre au sujet la notion de pseudo-mots.

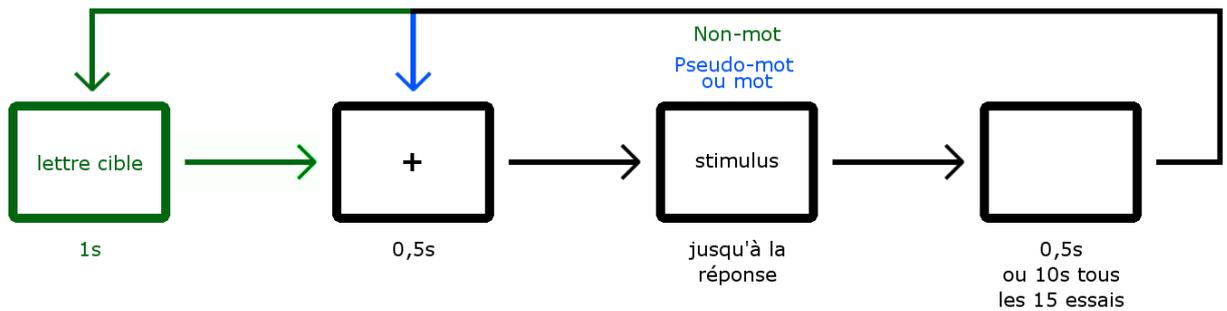


Figure 9. Schéma du déroulement des deux tests, les parties en vert correspondent au test d'identification de lettres, les parties en bleu au test de décision lexicale, les parties en noir sont le tronc commun.

5.1.3 Mesure du temps de réponse de base

Lorsque l'on compare plusieurs supports, les temps de réponse peuvent varier avec plusieurs facteurs :

- la qualité visuelle de l'information affichée (lisibilité),
- la facilité d'interaction avec le support (réponse au clavier, tactile,...),
- le traitement logiciel des durées de réponses propre au support et à son environnement logiciel.

Afin de pouvoir obtenir un indicateur de performance lié uniquement à la qualité visuelle des dispositifs d'affichage, nous mesurons également un temps de réponse de base. Cette mesure consiste à obtenir un temps de réponse moyen sur une tâche très simple. Un écran blanc est présenté au sujet, puis devient aléatoirement orange ou noir, dans un délai choisi aléatoirement entre 1 et 2,5 secondes. Le sujet doit répondre le plus vite possible « orange » ou « noir » en fonction de la couleur prise par l'écran. Une fois la réponse donnée, l'écran devient blanc à nouveau. Une mesure comprend 11 changements de couleurs de l'écran, et la mesure est répétée 5 fois. Le temps de réponse de base correspond à la moyenne des temps de réponse. Cette mesure permet de quantifier la durée de réponse liée à la facilité d'interaction et au traitement logiciel ; elle va nous permettre de pouvoir isoler la durée liée à la qualité visuelle de l'information affichée.

5.1.4 Traitement des données

Nous procédons à la même analyse pour le test d'identification de lettres et pour le test de décision lexicale. Chacun de ces deux tests contient 2 facteurs expérimentaux croisés : un

facteur lié à la difficulté (présence ou absence de lettres similaires dans le groupe de lettres ; mot et pseudo-mots associés à une décision lente ou rapide) et un facteur lié à la réponse demandée au sujet (présence ou absence de la lettre cible ; affichage d'un mot ou d'un pseudo-mot). Les deux facteurs étant croisés, nous avons donc 4 conditions pour chaque test.

Pour chaque test et pour chaque condition, nous obtenons 15 réponses et temps de réponse. On calcule pour chaque condition la moyenne des temps de réponse, à laquelle on soustrait le temps de réponse de base (mesuré à l'aide du test présenté dans 5.1.3 *Mesure du temps de réponse de base*). Le nombre d'erreurs est divisé par 15 afin d'obtenir un taux d'erreur.

5.2 La mesure de la vitesse de lecture maximale

Cette section de la thèse est une traduction d'un article que nous avons publié sous le titre « *A new sentence generator providing material for maximum reading speed measurement* » (Perrin, Paillé, & Baccino, 2014a), pouvant se traduire par « Un nouveau générateur de phrases pour mesurer de la vitesse maximale de lecture ». L'article original est accessible depuis le site de l'éditeur Springer, dans la revue *Behavior Research Methods*.

Résumé :

Nous proposons une nouvelle méthode pour générer du matériel textuel afin de mesurer la vitesse de lecture maximale de lecteurs adultes. La procédure décrite permet de générer un grand nombre de courtes phrases équivalentes. Ces phrases peuvent être affichées pour différentes durées afin de déterminer la vitesse de lecture maximale en utilisant un algorithme psychophysique de recherche du seuil. Chaque phrase est construite de manière à être vraie ou fausse, du point de vue des connaissances générales. La lecture effective est vérifiée en demandant au lecteur de déterminer la valeur de vérité de la phrase. Nous avons basé ce générateur sur le concept de Crossland et al. et l'avons amélioré. Le nouveau générateur manipule des concepts organisés dans une ontologie, ce qui permet une détermination facilitée de la valeur de vérité ainsi qu'un contrôle accru des paramètres psycholinguistiques. De cette manière, un grand nombre de phrases équivalentes peut être généré et affiché afin de réaliser la mesure. Les scores de vitesse de lecture obtenus à l'aide d'une sélection pseudo-aléatoire des phrases du générateur sont bien corrélés avec les valeurs obtenues par l'utilisation du MNREAD ($r = ,836$). De plus, la quantité importante de phrases rend possible la mesure répétée puisque la possibilité pour le lecteur d'apprendre les phrases

est éliminée. Ce générateur peut donc être utile aux chercheurs s'intéressant à la variabilité de performance intra-sujet.

Mots-clés :

Vitesse de lecture, mesure, générateur de phrases, psychophysique.

5.2.1 Introduction

Cet article porte sur une méthode pour mesurer la vitesse de lecture maximale du lecteur adulte. Les chercheurs de différents domaines étudient la lecture à plusieurs niveaux : des aspects perceptuels, tels que la reconnaissance de lettres et de mots (Pelli et al., 2006; Sereno & Rayner, 2003), aux processus cognitifs de plus haut niveau, tels que la compréhension (RAND Reading Study Group & Snow, 2002). Une manière d'adresser la question de la mesure de la lecture est d'évaluer la vitesse de lecture maximale. Elle reflète l'efficacité de la prise d'information et de son traitement. Cette efficacité varie avec différents facteurs. La vitesse de lecture varie entre les sujets : les vitesses sont mesurées par exemple en sciences de l'éducation et en psychologie du développement en relation au niveau d'éducation et au niveau de développement cognitif (Landerl & Wimmer, 2008). La vitesse de lecture varie également pour un même sujet. Les scientifiques de la vision, les optométristes et les ophtalmologues peuvent ainsi mesurer l'évolution de la fonction visuelle sur la durée de patients affectés par des pathologies de basse vision, ou mesurer le résultat d'une intervention thérapeutique (Ahn et al., 1995; Seiple, Szlyk, McMahon, Pulido, & Fishman, 2005; Trauzettel-Klosinski et al., 2012). De plus, les données de vitesses de lecture peuvent être collectées pour observer des effets du texte ou de son format de présentation. Les psychologues et psycholinguistes ont montré que le nombre de mots lus par minute varie avec des aspects textuels comme la fréquence lexicale (Kliegl et al., 2004) et la difficulté du texte (Just et al., 1996). La manière dont un texte est affiché (Dyson, 2004), ainsi que son support (Dillon, 1992) sont d'autres facteurs affectant la vitesse. L'utilisation massive de la vitesse de lecture comme une métrique pour la recherche expérimentale rend nécessaire le développement d'outils efficaces pour obtenir une mesure valide et fiable.

L'objectif principal de cet article est de proposer un moyen d'évaluer la lecture à travers la vitesse maximale. A terme, cela devrait offrir des mesures fiables pour quantifier l'effet de la manipulation d'un facteur expérimental (par exemple, la comparaison de la vitesse de plusieurs lecteurs, l'effet d'une pathologie sur la lecture dans le temps, les

différences de performance de lecture entre deux interfaces...). Pour y parvenir, la méthode doit répondre à quatre points :

(1) elle doit permettre l'évaluation de la lecture en dehors d'autres phénomènes. Il est important d'exclure des interférences telles que la production phonologique en ne considérant que la lecture silencieuse ;

(2) le matériel de lecture doit refléter l'expérience de lecture habituelle du sujet. Il doit donc être constitué de phrases complètes et structurées. Pour évaluer la lecture avec un matériel contrôlé, il est possible d'utiliser un corpus de mots isolés. Cette solution est néanmoins associée à l'identification de mots isolés, qui est différente de la lecture en situation écologique (Latham & Whitaker, 1996) ;

(3) la lecture d'une phrase issue du corpus doit être équivalente à la lecture de n'importe quelle autre. Les processus de compréhension complexes doivent être évités car ils peuvent altérer la séquence de lecture selon la difficulté (Staub, 2010) et selon les connaissances a priori du lecteur (Kendeou & van den Broek, 2007) ;

(4) enfin, le matériel doit contenir suffisamment de texte pour permettre les mesures répétées tout en empêchant un apprentissage qui modifierait la vitesse.

Ayant identifié les mêmes besoins, Crossland, Legge et Dakin (2008) ont développé un générateur automatique de phrases pour fournir des milliers de phrases destinées à la mesure de la vitesse de lecture maximale. Leur algorithme sélectionne aléatoirement un quantificateur, un objet et un trait pour former une phrase grammaticalement correcte. Quelques exemples sont les suivants : « Aucun chimpanzé n'a de plume », « Certains comédiens sont sans emploi »... Le processus de génération des phrases peut être répété pour obtenir beaucoup de matériel. Ces phrases peuvent être utilisées comme stimuli psychophysiques : en les affichant pour une durée contrôlée puis en vérifiant à chaque durée la lecture effective. Crossland et al. (2008) ont utilisé une méthode des stimuli constants, en affichant les phrases pour des durées allant de 0,016 à 0,5 secondes par mot. Le pourcentage de phrases lues correctement est tracé en fonction de la durée d'affichage, et une courbe est appliquée aux données pour les décrire. Un seuil de performance est déterminé (par exemple 80%) et le score qui correspond à l'intersection de la courbe et de ce seuil est calculé pour obtenir la vitesse de lecture maximale. La lecture effective est contrôlée par une tâche de compréhension. Chaque phrase est vraie ou fautive en fonction des éléments qui la composent.

Le lecteur doit simplement décider si la phrase lue était vraie ou fausse. Cela permet de mesurer la lecture silencieuse sans que le lecteur ne doive lire à voix haute. Comme le contenu de chaque phrase est plutôt facile à comprendre, une mauvaise réponse devrait dénoter l'incapacité du lecteur à lire la phrase à une vitesse donnée.

Leurs travaux ont été une avancée importante pour notre capacité à réaliser des mesures répétées de la vitesse de lecture. C'est un aspect essentiel pour les plans d'expériences intra-sujet, ou pour l'évaluation des fonctions visuelles dans le temps. Cependant leur méthode peut être encore améliorée. Crossland et al. (2008) ont demandé à des lecteurs de déterminer la valeur de vérité de phrases générées. De mauvaises réponses ont été observées, même lorsque la durée d'affichage était illimitée (entre 2 % et 17 % de mauvaises réponses pour chaque lecteur). Ces erreurs ne peuvent pas être dues à la durée d'affichage, elles doivent donc être dues à des difficultés pour les lecteurs à inférer les valeurs de vérité telles qu'elles sont déterminées par le générateur. Ces difficultés peuvent être le résultat de la sémantique ou de l'attribution automatique des valeurs par le générateur. Ces deux aspects sont traités dans les sections suivantes (voir respectivement « Choix des objets » et « Gestion des valeurs de vérité »). Durant l'utilisation des phrases de manière psychophysique, les erreurs peuvent biaiser l'estimation de la vitesse de lecture : une courbe psychométrique devrait idéalement atteindre les 100% de bonnes réponses un peu après avoir dépassé le seuil perceptuel. La mauvaise détermination des valeurs de vérité des phrases, bien qu'elles aient été lues, peut donc distordre la courbe psychométrique et biaiser l'estimation du seuil.

Le but de notre travail a été d'améliorer la méthode de mesure en perfectionnant le générateur. L'amélioration principale de notre nouveau générateur est l'augmentation de l'homogénéité des phrases. Cela est rendu possible par un contrôle plus fort sur les variables psycholinguistiques. Puisqu'il a été montré que la fréquence lexicale a un effet sur la vitesse de lecture, et que la valence et la concrétude peuvent altérer la vitesse et la compréhension (Egidi & Gerrig, 2009; Sadoski, 2001), nous avons supposé que le contrôle de ces variables pourrait améliorer la fiabilité. La seconde amélioration est une nouvelle méthode de détermination des valeurs de vérité. Elle permet au lecteur de déterminer les valeurs de vérité des phrases sans ambiguïté, et ainsi d'éviter les mauvaises réponses lorsque la phrase est lue. Un bénéfice supplémentaire de la méthode est qu'elle facilite le développement du corpus de phrases. Finalement, il a été nécessaire de valider le générateur de phrase en tant que test de lecture fiable en l'évaluant sur un grand nombre de lecteurs adultes. La version améliorée du

générateur a été conçue pour fournir des phrases en français pour des raisons pratiques, mais son principe devrait pouvoir être facilement appliqué pour n'importe quel langage syntaxiquement proche du français.

5.2.2 Principes de fonctionnement du générateur de phrases

Tout comme dans la méthode de Crossland et al. (2008), le but de notre nouveau générateur est de générer des phrases composées d'un quantificateur, d'un objet et d'une propriété. Chaque phrase doit être soit vraie soit fausse. Cette valeur de vérité doit pouvoir être déterminée sans effort par le lecteur, puisque nous ne sommes intéressés que par la capacité de lecture, et non par les aspects de compréhension. Si le lecteur est parvenu à lire la phrase en entier, il doit pouvoir donner la bonne réponse ; si la phrase n'est pas lue, la réponse doit être donnée au hasard. En outre, les phrases doivent être équivalentes lexicalement et en termes de compréhension, afin que la capacité à lire la phrase ne soit affectée que par sa durée d'affichage.

5.2.2.1 Gestion des valeurs de vérité

Chacune des phrases générées doit avoir une valeur de vérité (« Vraie » ou « Fausse »), du point de vue des connaissances générales. La valeur de vérité de chaque phrase pourrait être assignée manuellement. Ce serait cependant extrêmement long et répétitif, car le générateur a pour vocation de créer un grand nombre de phrases. Nous proposons une méthode qui permet de déterminer les valeurs de vérité assez aisément.

Crossland et al. (2008) ont proposé d'ajouter un « trait » qui suit le quantificateur et l'objet. Ce trait est une description de deux mots. Le premier mot est un verbe, le second un adjectif, ou un autre mot qui donne du sens à la phrase en suivant le verbe (« être démocrate », « dessiner des immeubles »...). Notre version du générateur est différente. Un corpus constitué de noms communs a été utilisé. Tous les noms de cette base ont été structurés dans une ontologie (voir la *Figure 10* pour un exemple de visualisation) en utilisant le logiciel Protégé 4.1 (Stanford Center for Biomedical Informatics Research, Stanford, Etats-Unis). Si un objet est considéré comme faisant partie d'une classe représentée par un autre nom, on le spécifie dans l'ontologie (par exemple : « chien » fait partie de la classe « animal »). La structure de l'ontologie va servir à déterminer les valeurs de vérité, elle doit donc être universellement acceptée.

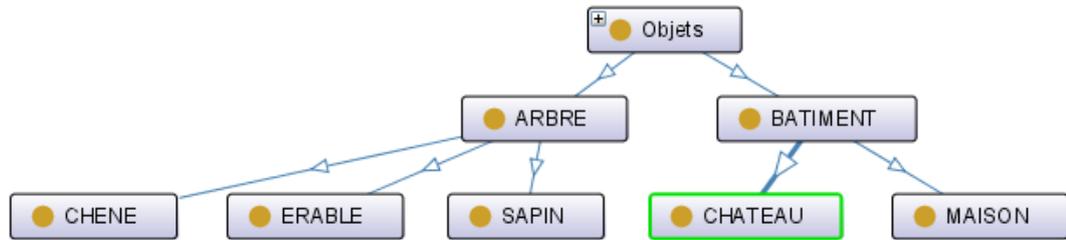


Figure 10. Exemple d'une partie de l'arbre sémantique¹². Les concepts du bas correspondent aux entités, ceux qui sont placés plus haut correspondent aux classes. Le concept supérieur "Objets" inclut tous les éléments de l'ontologie. Cette visualisation ainsi que celle de la figure suivante sont générées à l'aide du logiciel Protégé 4.1.

Nous utilisons le terme de « classes » pour des concepts qui incluent au moins un autre concept ; et « entités » pour des concepts qui sont inclus dans une classe, mais qui n'incluent pas eux-mêmes d'autres concepts. La différence principale entre classe et entité est expliquée dans la sous-section suivante. Pour construire des phrases, nous considérons des couples d'éléments dans la base. Chaque couple doit contenir au moins une entité. Ces couples serviront à construire des phrases sur le schéma suivant : « *Elément 1* est un *Elément 2* ». Ces couples ont une « proto-valeur de vérité », qui n'est pas un simple booléen (vrai ou faux). Certains couples sont toujours vrais (« V ») : « un chien est un animal » (quel que soit le chien, c'est toujours un animal). D'autres sont toujours faux (« F ») : « un chien est un oiseau ». Enfin, d'autres couples sont parfois vrais et faux dans les autres cas (« V/F ») : « un animal est un chien », c'est vrai pour certains animaux, mais pas pour tous. Nous utilisons la position des deux éléments dans l'ontologie pour déterminer automatiquement la proto-valeur de vérité. Si le premier élément est un descendant du second, le couple est toujours vrai. Si le deuxième élément est descendant du premier, le couple est V/F. Enfin, si les éléments ne sont pas liés par un lien de descendance, le couple est toujours faux.

En raison de la structure en arbre de l'ontologie, une entité ne peut avoir qu'un seul ascendant direct. Cette structure ne peut pas gérer les mots qui ont plusieurs usages (un mot qui dénote un concept, mais ce concept peut être intégré dans plusieurs classes). Nous devons spécifier dans l'ontologie les relations qui sont interdites. Ces deux éléments ne seront pas présents conjointement dans une même phrase (voir la *Figure 11* pour un exemple).

¹² Une flèche allant du concept A vers le concept B signifie : « A inclut B ».

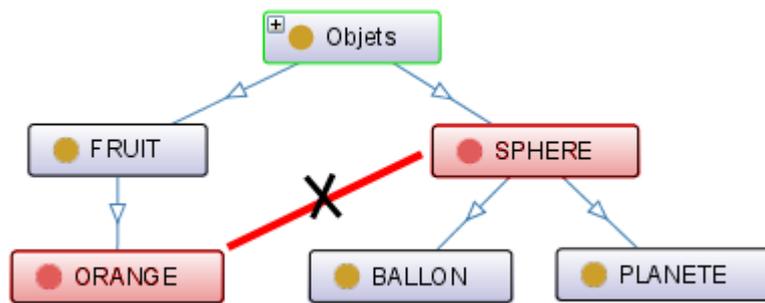


Figure 11. Exemple d'une ontologie contenant un concept ayant deux usages. Le concept "orange" est considéré dans l'ontologie comme étant un "fruit". La phrase "une orange est un fruit" est vraie. Mais dans le savoir commun, "une orange est une sphère" est également vraie. Comme l'ontologie ne peut manipuler qu'un seul usage pour chaque mot, nous devons spécifier au générateur de ne pas créer de phrase contenant ces deux mots.

En couplant chaque élément de l'ontologie avec tous les autres éléments autorisés, nous générons beaucoup de couples ainsi que leur proto-valeur de vérité. Nous devons nous assurer que les phrases finales auront une valeur de vérité binaire (V, ou F). Un quantificateur doit donc être ajouté pour avant le couple pour désambiguïser la proto-valeur de vérité. Nous avons utilisé les mêmes quantificateurs que Crossland et al. (2008) : « Aucun », « Certains » et « Tous les ». Nous proposons de déterminer la valeur de vérité en fonction du quantificateur et de la proto-valeur de vérité choisie. Avec trois quantificateurs, et trois proto-valeurs de vérité, il devrait y avoir neuf combinaisons possibles. Cependant, nous ne les considérerons pas toutes pour deux raisons.

Il n'y a pas d'équivalence absolue entre les langages naturel et logique. Un exemple est la combinaison du quantificateur « certains » (traduit par \exists en symbolisme logique) associé à un couple avec une proto-valeur de vérité V (« un chien est un animal »). La phrase « Certains chiens sont des animaux. » est vraie dans le sens de la logique du premier ordre. $\exists x (\text{Chien}(x) \wedge \text{Animal}(x))$ est une proposition vraie, car il y a au moins une chose dans le monde qui est à la fois un chien et un animal. Mais dans le langage naturel, cette phrase pourrait signifier implicitement qu'il existe des « chiens » qui ne sont pas des « animaux ». De telles phrases peuvent, au mieux, compliquer la compréhension, et, au pire, pousser le lecteur à l'erreur, même s'il a eu le temps de lire la phrase. Pour éviter ce problème, les phrases contenant ce quantificateur et cette proto-valeur de vérité ne doivent pas être générées.

Certaines combinaisons, dont les sens logique et naturel s'accordent, ont également été supprimées afin d'équilibrer les valeurs de vérité finales. Grâce à ces suppressions, la probabilité pour la phrase d'être vraie s'élève à 50%, quel que soit le quantificateur ou la proto-valeur de vérité utilisée (voir la *Table 1* pour toutes les combinaisons possibles). Ainsi, le lecteur ne peut pas augmenter ses chances de deviner la valeur de vérité en n'ayant vu qu'une partie de la phrase.

Table 1

Table de vérité des phrases générées, selon le couple d'éléments et le quantificateur sélectionnés

	Proto-valeur de vérité du couple			p(Vraie)
	V	V/F	F	
Quantificateur				
"Tous les"	V	F	(f)	0,5
"Certains"	x	V	F	0,5
"Aucun"	F	(f)	V	0,5
p(Vraie)	0,5	0,5	0,5	

Note. Les valeurs de vérité entre parenthèses représentent les combinaisons non sélectionnées afin d'équilibrer la probabilité pour une phrase d'être vraie. La valeur représentée par un « x » n'a pas été utilisée car ambiguë.

Un avantage de l'attribution automatique des valeurs de vérité à l'aide de l'ontologie consiste dans le fait que le processus de génération est plus facile et plus rapide : tous les couples possibles sont sélectionnés, les quantificateurs appropriés sont appliqués, aboutissant à des phrases non ambiguës sur leurs valeurs de vérité. La non-ambiguïté est importante pour prévenir les difficultés de détermination par le lecteur. Ainsi la présence de mauvaises réponses lorsque la phrase est lue devrait être limitée. Tout changement dans l'ontologie peut facilement être pris en compte pour la génération des phrases.

5.2.2.2 Choix des objets

La sous-section précédente a expliqué comment combiner deux mots de la base. Ici nous décrivons le processus de sélection pour remplir cette base de noms communs. Comme la fréquence lexicale a un effet important sur la vitesse de lecture (Kliegl et al., 2004) et que les phrases doivent être équivalentes, les mots sélectionnés doivent avoir une fréquence

lexicale élevée. Le nombre moyen d'occurrences de chaque mot pour un million était de 144,9, ce qui correspond à une valeur moyenne en log10 de 1,80 (écart-type : 0,56), selon la base « Lexique 3 » (New, Pallier, Ferrand, & Matos, 2001). Chaque mot fait partie des 7000 mots les plus courants en français. Les mots composés et les indéénombrables ont été exclus.

Dans l'ontologie, les « entités » sont les mots qui correspondent à des concepts avec au moins un ascendant et sans descendant. Ces mots doivent être faciles à comprendre. Dans le cas contraire, le lecteur pourrait mal comprendre la phrase, et donner une mauvaise réponse même si la phrase est complètement lue. Les mots abstraits doivent aussi être évités afin que le lecteur ne soit pas amené à analyser et interpréter des concepts ambigus. Certaines variables psycholinguistiques sont connues et permettent de contrôler ces propriétés. Seuls les mots avec une forte valeur de concrétude et une valence émotionnelle neutre ou positive ont été sélectionnés, à partir des données de Bonin et al. (2003).

Pour les classes – les mots qui ont au moins un descendant – seule la fréquence lexicale a été contrôlée. Comme ces mots dénotent plus de concepts, ils peuvent être moins spécifiques (par exemple le concept « Animal »). Une classe peut éventuellement être considérée comme une entité, à condition qu'elle respecte les valeurs de concrétude et de valences appliquées aux entités et qu'elle ait au moins un ascendant. Notre base a été constituée d'un total de 65 mots.

5.2.2.3 Génération des phrases et lecture

Pour générer les phrases, tous les couples possibles sont sélectionnés en tenant compte des règles précédemment décrites. La proto-valeur de vérité de chaque couple est connue. Ensuite, tous les quantificateurs appropriés sont choisis ; une application de la table de vérité (table 1) donne la valeur de vérité finale de la phrase. La phrase est composée d'un quantificateur, d'un nom commun, du verbe « être » et d'un second nom commun. La dernière action est d'ajouter l'article approprié au second nom commun, et d'appliquer les règles de grammaire et de conjugaison pour obtenir une phrase bien structurée avec une valeur de vérité connue. A partir de la base des 65 mots précédemment décrite, et étant données les relations non autorisées spécifiées, nous avons pu générer environ un millier de phrases. Une quantité modérée de concepts dans l'ontologie peut donc générer un grand nombre de phrases, ce qui limite le risque qu'une phrase soit affichée plusieurs fois. Une phrase devrait se répéter une fois sur mille, et ce taux pourrait être encore diminué en incluant plus de concepts dans l'ontologie.

En plus des mots ayant plusieurs usages décrits précédemment, certains mots peuvent avoir plusieurs sens (un seul mot qui dénote plusieurs concepts) : les homonymes et les mots polysémiques. Ceux pour lesquels les différents sens sont couramment utilisés doivent être évités (par exemple « avocat » désigne à la fois le fruit et la profession). Mais biens d'autres mots ont un sens majeur et un ou plusieurs sens mineurs ; il serait inutile d'essayer de tous les exclure. Par exemple, dans notre système la phrase « Toutes les chemises sont des vêtements » est vraie. Mais si le mot « chemise » dénote l'habit, il correspond aussi à la fourniture de bureau (« chemise cartonnée »).¹³ Ce double sens ne peut pas être géré directement par l'ontologie. Nous conseillons donc d'inclure dans la consigne le fait de devoir considérer le sens premier et le plus commun des mots durant l'administration du test.

Un exemple de chaque quantificateur et de chaque proto-valeur de vérité (pvv) est donné dans les phrases suivantes avec sa valeur de vérité finale (vv) : « Un train n'est pas un animal » (pvv : F, vv : V), « Certains trains sont des animaux » (pvv : F, vv : V), « Certaines fleurs sont des roses » (pvv : V/F, vv : V), « Toutes les fleurs sont des roses » (pvv : V/F, vv : F), « Une orange n'est pas un fruit » (pvv : V, vv : F), « Toutes les oranges sont des fruits » (pvv : V, vv : V).

5.2.3 Validation des phrases générées dans l'évaluation de la vitesse de lecture

Le générateur est capable de produire un grand nombre de phrases. Il est nécessaire de vérifier qu'elles peuvent constituer une mesure valide de la vitesse de lecture maximale. L'équivalence de la méthode de vérification à travers la réponse « vrai » ou « faux » avec une restitution orale des phrases lues silencieusement doit aussi être validée.

5.2.3.1 Protocole

5.2.3.1.1 Participants

Pour pouvoir prendre part à l'étude, les participants devaient avoir une bonne acuité binoculaire pour la distance à l'écran : au moins 0 LogMAR (10/10), en portant leur correction visuelle si nécessaire. Le port de verres double foyer ou progressifs n'était pas

¹³ Pour des raisons évidentes de traduction, les exemples de cette partie sont différents de ceux présentés dans l'article original.

autorisé afin d'éviter une altération de la performance visuelle due à la position du mot dans le verre. L'acuité des sujets a été mesurée à l'aide d'un Optoprox (Essilor International, Créteil, France) à la distance du test (44cm du centre de l'écran). Les sujets étaient de langue maternelle française et avaient entre 18 et 45 ans. La limite de 45 ans a été choisie pour éviter au maximum d'inclure des personnes presbytes. Tout antécédent de pathologie visuelle ou risquant une interférence avec l'étude était un critère d'exclusion. Les sujets ayant participé à cette validation ont été recrutés directement par les expérimentateurs. Tous les participants ont signé un formulaire de consentement éclairé. 45 participants ont pris part à l'étude, 23 femmes et 22 hommes. L'âge moyen était de 32 ans (écart-type : 7), allant de 22 à 45 ans.

5.2.3.1.2 Appareillage

Les phrases ont été présentées sur un écran E2311H (Dell, Round Rock, Etats-Unis) avec une résolution de 1280*800. Les lettres étaient affichées en noir sur fond blanc, à un contraste de 98,5% (Michelson) dans la lumière ambiante. La luminance de l'arrière plan a été fixée à 88,6 cd/m², mesurée avec un Cal-SPOT 401 (The Cooke Corporation, Romulus, Etats-Unis). Les sujets étaient positionnés dans une mentonnière de manière à ce que la distance de 44cm à l'écran soit maintenue. La police utilisée était Arial avec une taille de 18 px. Dans ces conditions, une lettre minuscule « o » avait un diamètre de 16,6 minutes d'angle visuel (ce qui correspond à environ 0.5 LogMAR). Le script d'affichage a été écrit en Python 2.7.6 avec le module PyGame 1.9.1.

5.2.3.1.3 Tests de lecture

La validation du test a consisté en la vérification de l'équivalence entre le score de vitesse de lecture obtenu par la lecture des phrases générées avec un score obtenu par un test reconnu dans la littérature. Nous avons choisi le même test que celui utilisé par Crossland et al. (2008) lorsqu'ils ont validé leur générateur, c'est-à-dire le MNREAD (Ahn et al., 1995; Legge et al., 1989). Ce test sert surtout à évaluer la vitesse de lecture orale, mais peut être aussi utilisé pour la lecture silencieuse. Nous avons utilisé une version en langue française du MNREAD (Senécal et al., 2006) puisque les phrases de notre générateur sont destinées à des lecteurs français. Tous les sujets ont du lire des phrases du MNREAD et de notre algorithme. Celles de l'algorithme étaient lues dans deux conditions : l'une avec une restitution orale de la phrase lue, et l'autre avec la tâche vrai/faux. L'ordre des tests était pseudo-randomisé pour chaque sujet. Chaque mesure était réalisée binoculairement.

5.2.3.1.4 Procédure et mesures

Pour chacun des trois tests, plusieurs phrases étaient affichées pour une durée limitée. Chaque phrase était précédée par une croix de fixation à la position de la première lettre de la phrase. Quand la phrase disparaissait, elle était suivie d'un masque. La croix, ainsi que le masque étaient affichés pendant une seconde. Pour vérifier la lecture effective, le lecteur devait citer à voix haute les 4 derniers mots qu'il avait lus dans la phrase du MNREAD ; il pouvait également la citer en entier selon son envie, mais seule la lecture effective des 4 derniers mots était prise en compte. Pour les phrases générées, le lecteur devait citer à voix haute la phrase en entier, ou indiquer si le contenu était vrai ou faux. Si tous les mots demandés étaient cités sans erreur (tests oraux), ou si la bonne valeur de vérité était donnée (v/f), la réponse était considérée comme correcte. Les sujets pouvaient prendre autant de temps que souhaité pour donner leur réponse après l'affichage. Pour le MNREAD, chaque phrase était sélectionnée pseudo-aléatoirement (tirage sans remise) parmi les phrases de ce test. Chaque phrase générée était sélectionnée pseudo-aléatoirement (tirage avec remise) parmi toutes les phrases générées. Un facteur de pondération a été appliqué pour qu'une proto-valeur de vérité (Faux, Vrai/Faux, Vrai) ait la même chance d'être sélectionnée qu'une autre.

La variable utilisée pour moduler le niveau du stimulus était la durée d'affichage par mot en millisecondes. Par exemple, si la phrase contenait 5 mots, et que la variable était fixée à 200ms, alors la phrase entière était affichée pendant 1s. Le score issu du test est la durée minimale d'affichage par mot qui permet au sujet de lire la phrase en entier. Cette mesure peut être facilement traduite dans l'unité classique de mots par minutes, et donc être interprétée comme un score de vitesse de lecture maximale.

Une méthode de type *staircase* (ou en escalier) a été utilisée pour obtenir le score de durée d'affichage minimale : la méthode d'approximation stochastique (Robbins & Monro, 1951; Treutwein, 1995). Pour les deux tests où le sujet devait citer la phrase après l'avoir lue, le niveau de performance vers lequel converger a été fixé à 50% : si la durée d'affichage n'a pas été assez longue, le lecteur ne doit pas pouvoir citer la phrase en entier. Pour le test vrai ou faux, le lecteur a toujours 50% de chances de répondre correctement, même s'il n'a pas lu la phrase ; le seuil a été fixé à 75%. Les autres paramètres du staircase ont été calculés à partir de données expérimentales de 10 sujets ayant passé les tests avec une méthode des stimuli constants : plusieurs staircases avec différents jeux de paramètres ont été simulés sur les données recueillies. Les paramètres qui ont permis d'atteindre les scores les plus proches de ceux obtenus avec la méthode des stimuli constants ont été sélectionnés. Les paramètres des

trois tests sont donnés dans la *Table 2*. La durée d’affichage minimale a été calculée en faisant la moyenne des durées aux 6 dernières inversions du staircase.

Table 2

Paramètres pour calculer le niveau du stimulus à l’essai n;

Test	Paramètres			
	Nombre d’essais	Niveau au premier essai [ms]	Taille initiale du pas [ms]	Valeur de probabilité
	<i>N</i>	X_1	<i>c</i>	Φ
MNREAD	20	200	200	0,5
Phrases générées (oral)	35	500	500	0,5
Phrases générées (vrai ou faux)	45	250	500	0,75

Note. Le niveau du stimulus à l’essai *n* est donné par : $X_n = X_{n-1} - ((c / n) * (Z_{n-1} - \Phi))$. Z_n représente la réponse du sujet à l’essai *n* (1 = bonne réponse, 0 = mauvaise réponse).

Chaque test était précédé d’une phase d’instruction et de familiarisation. Le lecteur était informé qu’une croix qu’il devait fixer allait apparaître, suivie d’une phrase qu’il devait lire le plus vite possible et silencieusement. On expliquait que la durée d’affichage allait varier, et qu’il était normal de ne pas pouvoir lire la phrase en entier. Pour le MNREAD, il lui était demandé de citer les 4 derniers mots qu’il avait pu lire ou la phrase en entier, mais on lui signifiait que seuls les 4 derniers mots seraient pris en compte. Pour les phrases générées avec restitution orale, il était demandé de citer la phrase en entier. Pour l’autre modalité, il était demandé au lecteur de déterminer si la phrase lue était vraie ou fausse, et de répondre au hasard s’il n’avait pas pu lire la phrase. Avant chaque test, deux phrases étaient présentées au lecteur : l’une pendant une longue durée et la seconde pour une durée très courte. Pour la modalité « vrai ou faux », deux phrases supplémentaires étaient montrées : l’une pour laquelle la valeur de vérité était fausse, et l’autre où un homonyme était présent, de manière à expliquer au lecteur qu’il devait tenir compte du sens premier des mots présentés.

5.2.3.1.5 Analyse des données

Une fois les trois scores de chaque sujet obtenus, nous avons comparé les scores obtenus au MNREAD avec ceux obtenus sur les phrases générées avec la modalité orale. Le but est de vérifier que la lecture des phrases générées aboutit à une mesure liée aux performances de lecture obtenues par un test classique. Ensuite, nous avons examiné la

concordance entre les deux modalités (orale et vrai ou faux) de notre test. L'objectif du nouveau générateur est de fournir des phrases dont la valeur de vérité peut être déterminée sans ambiguïté. Le score obtenu à l'oral doit donc être équivalent à celui obtenu avec la tâche vrai ou faux.

Une manière habituelle de déterminer la concordance entre deux méthodes de mesure est l'analyse graphique de Bland-Altman (Bland & Altman, 1986) : il faut obtenir les données issues des deux mesures sur plusieurs objets. On intègre ces données dans un graphique dont l'abscisse est la moyenne des deux mesures et dont l'ordonnée est la différence signée entre les deux mesures. Néanmoins cette technique est critiquée car elle peut présenter un biais qui n'est pas dû à une différence entre les mesures, mais à la technique elle-même : si l'écart-type associé à chaque méthode testée n'est pas égal aux autres, la représentation graphique se verra affectée d'une corrélation négative entre les valeurs moyennes et la différence signée (pour une explication complète de l'origine du biais, et sur les manières de l'éviter, se référer à l'article et aux simulations de (Hopkins, 2004). L'utilisation de l'approche corrélationnelle évite l'émergence du biais artificiel, voilà pourquoi nous l'avons choisie pour effectuer les deux comparaisons. Cette approche se compose de plusieurs étapes : on réalise d'abord une régression pour vérifier la relation entre les deux paires de scores ; les coefficients de corrélation r ont été calculés et des tests de Student ont été réalisés pour vérifier la significativité de ces coefficients. Une fois les relations montrées, il faut comparer les paramètres de l'équation de la régression avec la fonction identité pour vérifier l'absence de biais entre les deux mesures. Finalement, il faut analyser les résidus en vérifiant leur normalité à l'aide d'un test de Lilliefors, et contrôler visuellement l'absence d'hétéroscédasticité sur les graphes de régression : si l'on ne constate pas d'augmentation de la variance des résidus aux extrémités des graphes, cela signifie que l'erreur est globalement la même pour chaque sujet.

5.2.3.2 Résultats

Deux participants ont été retirés de l'analyse car leurs résultats n'ont pas permis d'atteindre un niveau stable au test MNREAD. Deux autres ont été supprimés car ils ont indiqué avoir utilisé une stratégie impliquant une déviation par rapport aux instructions de départ pour améliorer leur score durant le test « vrai ou faux ». Cette stratégie est discutée dans la dernière section. Finalement, les données 21 femmes et 20 hommes ont servi à

l'analyse. Chaque staircase a produit au moins six inversions et a donc permis le calcul du seuil de durée d'affichage minimale.

Les durées d'affichage minimales mesurées à chaque test sont données dans la *Figure 12*. Le score moyen pour le MNREAD était de 115 ms (écart-type : 32,8). Le score équivalent en vitesse de lecture est 521,7 mots par minute. Cette vitesse de lecture est très élevée comparativement aux vitesses classiquement relevées sur des textes. Nous pouvons expliquer cette vitesse par le fait que les textes sont très courts (3 lignes de 20 caractères) et faciles à comprendre. Les moyennes des durées d'affichage minimale pour les phrases générées étaient plus faibles : 60ms pour le test à réponse orale et 58ms pour le test en modalité vrai/faux (écart-types de 25,2 et 26,8 respectivement).

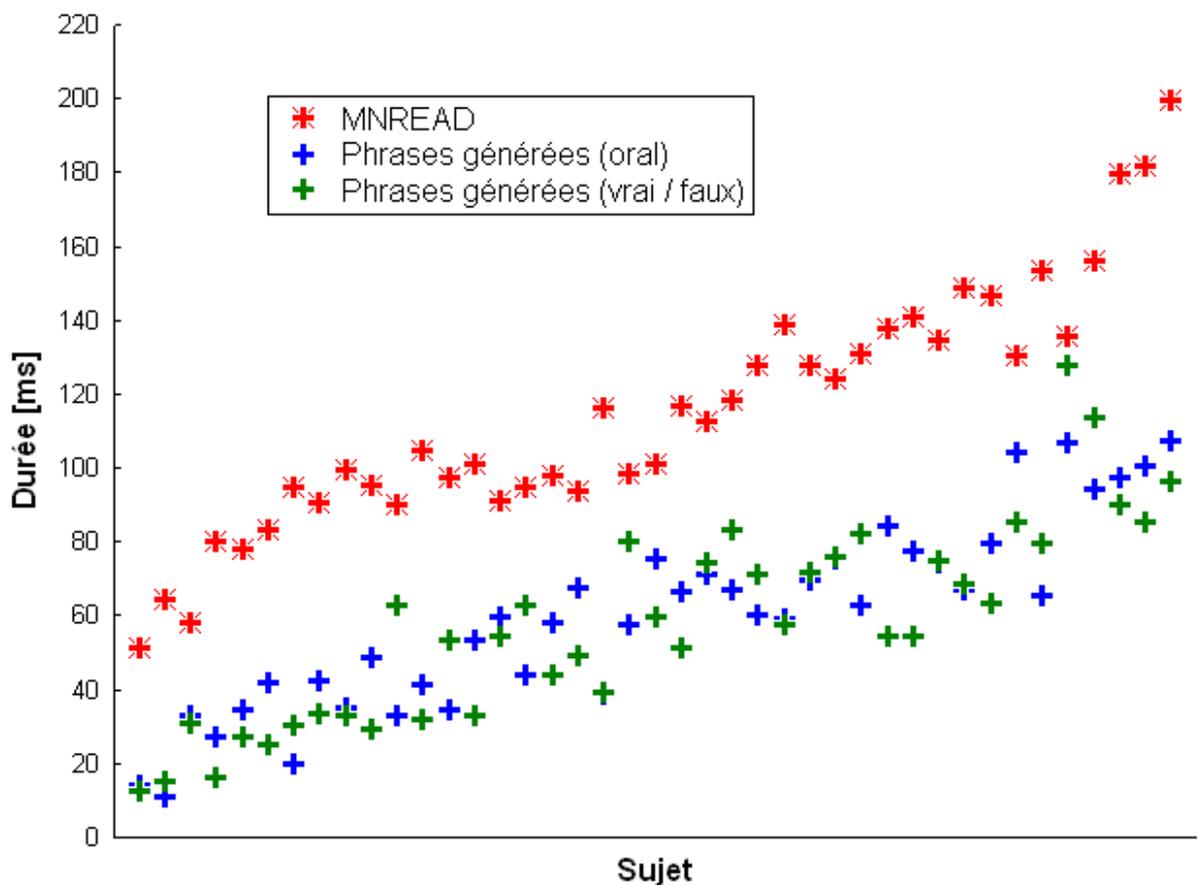


Figure 12. Durée minimale d'affichage par mot permettant la lecture effective. Les participants sont classés dans l'ordre croissant de la moyenne de leurs trois scores.

Il y a une différence importante entre les scores du MNREAD et ceux de nos tests. Cela peut être dû à deux raisons. Dans nos tests, les phrases étaient affichées sur une seule ligne, en comparaison aux trois lignes du MNREAD. Les deux saccades pour aller à la ligne

suiuante dans le MNREAD sont consommatrices de temps. De plus, les phrases g n r es suivent toujours la m me structure. Le lecteur peut rapidement s'habituer   cette pr sentation, et pr dire o  seront plac s les mots importants, ce qui diff re de la structure syntaxique qui est diversifi e dans les phrases du MNREAD.

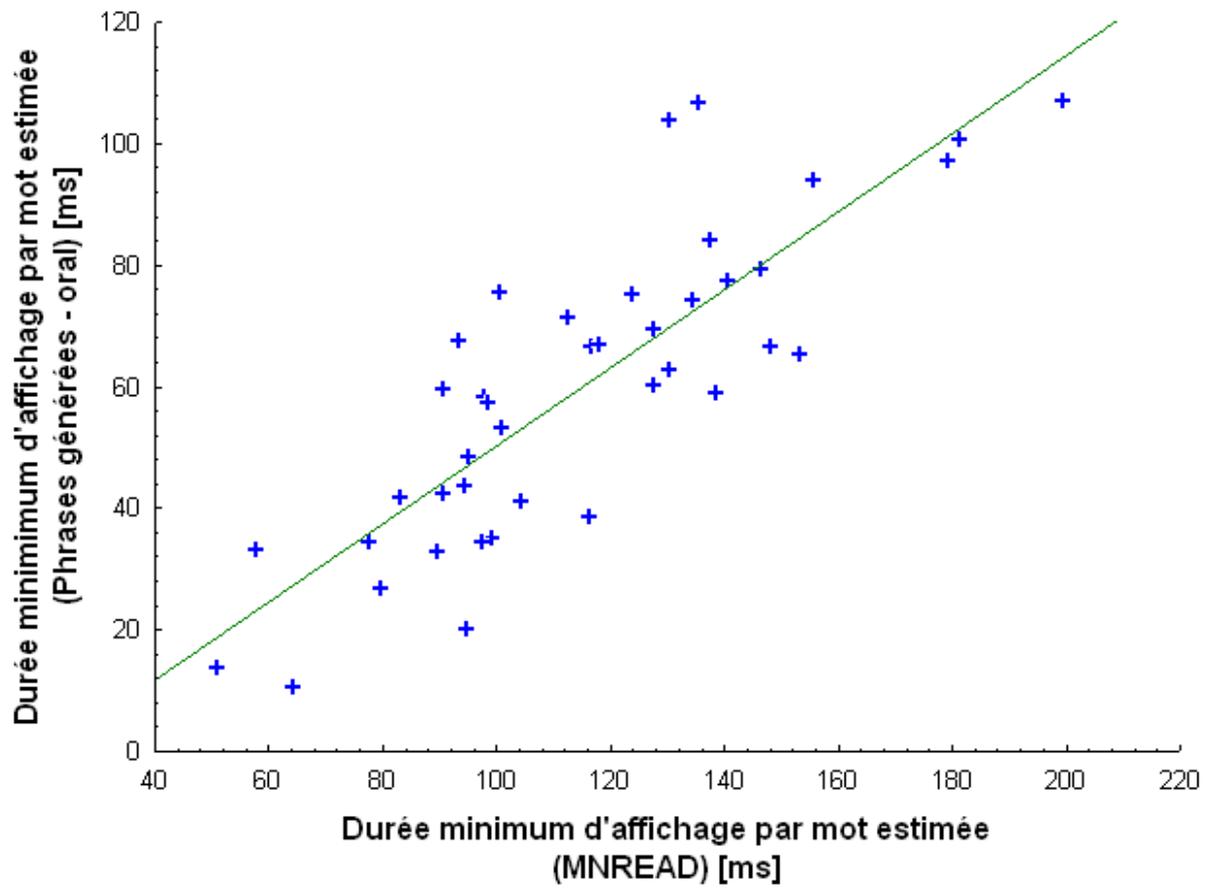
Pour valider notre test, nous auons d'abord effectu  les deux mesures de concordance : (1) la comparaison entre les deux scores en modalit  orale : le MNREAD et les phrases g n r es, et (2) la comparaison entre les deux modalit s de tests avec les phrases g n r es (orale et vrai/faux). Les graphiques de corr lation correspondant aux deux mesures de concordance sont pr sent s dans la *Figure 13*. Pour la comparaison (1), le coefficient de r gression est acceptable : $r = .836$, $t(39) = 9,504$, $p < ,001$, $IC_{95\%} = [0,658 ; 1,014]$, erreur-type de l'estimation $SEE = 14,03$. Nous auons pu faire la m me observation pour la comparaison (2) : $r = ,839$, $t(39) = 9,645$, $p < ,001$, $IC_{95\%} = [0,663 ; 1,015]$, $SEE = 14,77$. Comme les coefficients de corr lation sont  lev s pour chaque paire de donn es (chaque $r > ,80$), nous auons affirm  que les mesures sont fortement li es entre elles.

La r gression (1) aboutit   l' quation suivante : $score_{PhrasesG n r esOral} = A_0 + A_1 * score_{MNREAD}$ avec $A_0 = - 14,04$, $IC_{95\%} = [-30,37 ; 2,30]$ et $A_1 = 0,64$, $IC_{95\%} = [0,51 ; 0,78]$. L'intervalle de confiance de la pente de la r gression indique que cette valeur est significativement diff rente de 1. La fonction de r gression est donc diff rente de la fonction identit  : les scores du MNREAD sont diff rents de ceux observ s avec les phrases g n r es. Cela confirme la diff rence observ e pr c demment.

La r gression (2) donne : $score_{PhrasesG n r esV/F} = A_0 + A_1 * score_{PhrasesG n r esOral}$, avec $A_0 = 4,53$, $IC_{95\%} = [-7,60 ; 16,68]$ et $A_1 = 0,89$, $IC_{95\%} = [0,71 ; 1,08]$. Etant donn s les intervalles de confiance de A_1 et de A_0 , la pente de la droite de r gression ne peut pas  tre consid r e comme significativement diff rente de 1, et l'ordonn e   l'origine ne peut pas l' tre de 0. On ne peut donc pas rejeter l'hypoth se de l' galit  entre la fonction de r gression et la fonction identit  (c'est- -dire que $score_{PhrasesG n r esV/F} = score_{PhrasesG n r esOral}$ ne peut pas  tre rejet e). Ce r sultat est en faveur de l'ad quation entre les mesures des deux modalit s de test.

L'hypoth se de normalit  n'a pas  t  rejet e pour les r sidus des deux r gressions d'apr s les tests de Lilliefors : pour (1) $d = ,092$, $p > ,20$ et pour (2) $d = ,105$, $p > ,20$.

L'observation visuelle des deux graphiques ne révèle pas la présence de problèmes d'hétéroscédasticité.



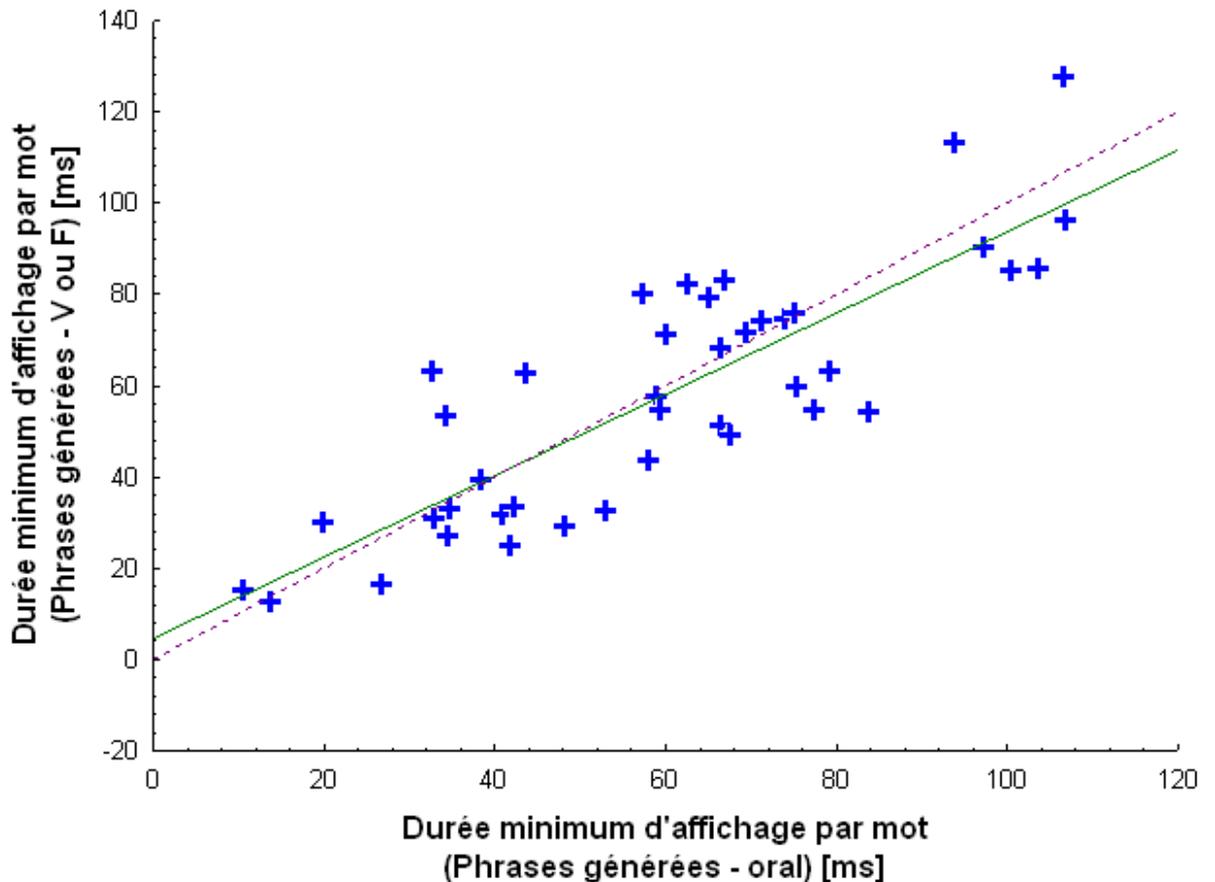


Figure 13. Graphiques de régression pour les comparaisons entre les deux tests et les deux modalités. Les lignes vertes représentent les droites de régression. Dans le deuxième graphique, la ligne violette en pointillés représente la fonction identité.

5.2.4 Discussion / conclusion

Pour les deux comparaisons entre le MNREAD et les phrases générées, nous avons observé des corrélations importantes ainsi que la normalité et l'homoscédasticité des résidus. Cela implique que le test utilisant les phrases générées peut être considéré comme un bon prédicteur du score obtenu avec le test classique. Néanmoins il existe un décalage entre les scores du MNREAD et des phrases générées. Les lecteurs ont été capables de lire plus rapidement les phrases générées que celles du MNREAD. Comme cela a été expliqué, la raison peut être la différence de nombre de lignes (trois pour le MNREAD contre une seule pour les phrases générées) ou alors la structure répétitive des phrases générées, et l'utilisation de mots très fréquents. La vitesse de lecture est plus importante dans notre test, mais tout de même linéairement reliée à la mesure classique. Cela implique que le score de vitesse de lecture maximale obtenu avec notre nouveau test a juste besoin que l'on tienne compte de cette différence linéaire pour être équivalent à un test déjà validé de vitesse de lecture. Les lecteurs les plus rapides que nous avons observés ont obtenu des scores de vitesse

extrêmement rapides : les phrases ont pu être lues, même si elles étaient affichées pendant moins de 100ms, ce qui n'a pas pu leur permettre de réaliser des mouvements des yeux. Ils ont donc été capables de lire des phrases avec seulement une fixation, à la manière du paradigme de présentation visuelle rapide et sérielle (RSVP) de Forster (1970). Ces lecteurs doivent avoir un empan perceptuel suffisamment grand pour couvrir presque toute la phrase, leur permettant de l'intégrer sans bouger les yeux. Des recherches précédentes avaient déjà montré que les lecteurs rapides ont un empan plus grand (Rayner, Slattery, & Bélanger, 2010). Malgré ce comportement inhabituel, où nous aurions pu nous attendre à observer une cassure dans la relation linéaire, les scores des lecteurs rapides au MNREAD sont associés aux scores aux phrases générées, de la même manière que chez les lecteurs « normaux ». Cela montre que les mesures effectuées avec notre nouveau test reflètent bien les capacités liées à celles utilisées durant les mouvements de lecture normaux (c'est-à-dire, qui incluent des mouvements des yeux).

Pour aller plus loin, on peut se poser la question de l'équivalence entre les valeurs obtenues grâce à nos tests et à la vitesse de lecture en situation écologique, étant donné les scores de vitesse très rapides. La vitesse de lecture habituelle n'est pas une valeur fixée ; sa variabilité est due à des changements dans les stratégies de lecture (Carver, 1990; Lemaire, Guérin-Dugué, Baccino, Chanceaux, & Pasqualotti, 2011) et aux aspects sémantiques et visuels du texte lu. Nous avons voulu contrôler ces aspects dans les phrases générées afin d'obtenir un matériel qui soit le plus homogène possible. La stratégie de lecture était aussi contrôlée, en demandant aux lecteurs de lire le plus vite possible tout en lisant de manière à pouvoir citer et comprendre les phrases. Ainsi, le nouveau test permet de mesurer une vitesse de lecture maximale, qui n'est pas la vitesse habituelle que l'on observerait en condition naturelle. Cette mesure permet de limiter la variabilité intra-lecteur afin d'obtenir une valeur stable, et de pouvoir effectuer des mesures répétées. L'effet d'un changement dans le format de présentation et l'évolution des performances de lecture dans le temps peuvent donc être étudiées en minimisant les autres facteurs.

Les scores obtenus à travers les deux modalités de réponse pour les phrases générées sont aussi bien corrélés. Leur relation est proche de l'égalité. Ce résultat est encourageant pour valider la modalité « vrai ou faux ». L'un des buts de notre générateur était de fournir des phrases pour lesquelles les valeurs de vérité peuvent être déterminées sans ambiguïté par le lecteur. Si des phrases ambiguës avaient été montrées aux lecteurs, nous aurions observé des erreurs, même lorsque les durées d'affichage auraient été assez longues pour permettre la

lecture. Le staircase aurait aboutit à des seuils de performance de vitesse de lecture plus faibles pour la modalité « vrai ou faux ». Comme cela n'a pas été le cas, nous pouvons affirmer que cet objectif est rempli. Les deux modalités peuvent donc être utilisées alternativement pour obtenir une valeur de vitesse maximale de lecture. La modalité V/F est plus facile à mettre en place, car elle permet au lecteur de passer le test sans que l'investigateur ne soit obligé de vérifier ses réponses orales. A l'inverse, si le sujet n'apprécie pas la procédure V/F, l'observateur peut basculer sur le test oral. La nouvelle méthode de génération de phrases peut offrir une quantité de matériel substantielle pour la mesure de la vitesse de lecture. Notre méthode proposée permet une production efficace des phrases : les seuls efforts requis par le concepteur sont de sélectionner les noms communs adaptés et de les structurer dans une ontologie. Les étapes suivantes de la production sont automatisées. L'une des améliorations majeures de la méthode est le contrôle des variables lexicales et psycholinguistiques permis par la sélection manuelle des mots. Le contrôle de ces paramètres est crucial pour éviter la variabilité due au texte lui-même. Ce contrôle, associé à un formalisme proche du langage naturel, permet au lecteur de donner des réponses correctes plus facilement.

Dans la phase de validation, chaque phrase était affichée entièrement, et non pas mot après mot comme dans le paradigme RSVP, où chaque mot de la phrase est affiché pendant une durée très courte. La vitesse d'affichage des ordinateurs usuels est limitée par le taux de rafraîchissement des écrans, et ne ceux-ci peuvent pas afficher chaque mot lorsque la fréquence d'apparition est proche de ce taux. Nos phrases pourraient être utilisées pour un paradigme RSVP, mais uniquement sur des systèmes à très haute résolution temporelle.

Nous avons dû exclure quelques sujets car leurs réponses aux tests n'ont pas permis au staircase de déterminer un niveau de performance stable. Cela indique un changement dans leur niveau de performance. Cela pourrait être expliqué par une variation de leur niveau attentionnel, mais aussi par un changement de stratégies utilisées pour la lecture. Certains sujets ont expliqué avoir utilisé une stratégie qui était de ne pas fixer la croix avant l'apparition de la phrase, mais de regarder quelques centimètres plus à droite. Cette stratégie permet à leur empan visuel de couvrir une plus grande partie de la phrase. Pour empêcher cette stratégie, la croix de fixation pourrait être placée plus loin dans la phrase, ou alors un enregistrement des mouvements des yeux pourrait être utilisé pour ne démarrer l'affichage que quand le lecteur fixe effectivement la croix. Néanmoins, la majorité des lecteurs ont pu donner des résultats exploitables.

Nous avons proposé une méthode permettant la mesure de la vitesse de lecture maximale en adoptant une approche psychophysique. Une telle approche permet de multiplier les mesures de manière répétées. La méthode est préconisée pour l'évaluation des différences de vitesse de lecture liées aux facteurs visuels. Néanmoins, la lecture n'est pas limitée à ses aspects visuels. Le but de la lecture est d'extraire de la connaissance incluse dans un texte, ce qui implique de nombreux processus cognitifs agissant sur l'information visuelle acquise. En mettant en place des phrases facilement compréhensibles, qui suivent la même syntaxe, et composées de mots très fréquents qui désignent des concepts concrets, nous avons minimisé l'effet des processus cognitifs autres que visuels sur la mesure résultante. La recherche sur l'évaluation de la lecture ne peut être complète sans adresser ces processus : par exemple, le nouveau test n'est pas sensible à l'acquisition du vocabulaire puisque les mots sont très couramment utilisés. Le rôle de la compréhension est également très limité dans notre test car la tâche de compréhension ne dépendait que des connaissances antérieures. Nous espérons que la communauté de recherche sur la lecture continuera à développer des outils robustes pour étudier la lecture à tous les niveaux, et que notre approche pourra inspirer de futurs outils conçus dans ce but.

Depuis la publication de cet article, nous avons amélioré la mesure en remplaçant la méthode de staircase par l'algorithme adaptatif Psi (Kontsevich & Tyler, 1999), dont le fonctionnement est détaillé dans *4.2 Technique de caractérisation psychométrique*. Le nombre d'essais nécessaires pour déterminer le seuil de vitesse de lecture maximale a ainsi été ramené à 20 essais pour la modalité orale, au lieu des 35 essais initiaux.

Chapitre 6 : Chapitre 6 : L'évaluation de la compréhension

6.1 Objectifs du test de compréhension

Ce chapitre a pour objet de décrire le test de compréhension que nous avons construit pour faire partie de la batterie de tests du Quotient de Lecture Electronique. Le but d'un tel test est de nous permettre de mesurer les différences de compréhension dues à l'interface ou au dispositif d'affichage. Les tests de compréhension contiennent deux volets : un volet de lecture de texte par le lecteur, et un volet de vérification ou de mesure des processus de compréhension par l'expérimentateur. Pour construire un test, il est donc nécessaire de définir dans un premier temps les textes employés puis dans un second temps la tâche de vérification.

La tâche servant à la vérification du produit de la compréhension prend souvent la forme d'un questionnaire de compréhension. Les modèles de la compréhension (Van Dijk & Kintsch, 1983; Kintsch, 1988; van den Broek et al., 1996) montrent qu'il ne s'agit pas d'un phénomène simple, mais que la compréhension d'un texte se construit simultanément à différents niveaux de traitement. La compréhension de bas niveau peut être mesurée à l'aide de questions de mémorisation¹⁴ alors que la compréhension de haut niveau peut être évaluée par des questions d'inférence. Pour être complète, la mesure doit être réalisée au moyen de questionnaires évaluant chacune de ces facettes de la compréhension.

Notre test doit permettre de mesurer les différences de performances liées à l'interface utilisée. Le test doit donc donner la possibilité d'effectuer des mesures répétées. Ce besoin nous pousse encore une fois à nous inspirer des méthodes de la psychophysique. Bien entendu, un test de compréhension ne peut pas être considéré comme étant un test psychophysique, dans la mesure où les textes ne varient pas selon une seule dimension. Les stimuli employés par la psychophysique sont très contrôlés, et ne visent à tester qu'une seule variable à la fois, en général une grandeur physique (on peut faire varier l'intensité lumineuse, le contraste, le volume sonore,...). Les textes, eux, varient à de nombreux niveaux : sémantique, structure, type de texte,... En outre, la complexité est accrue par les interactions entre le lecteur et le texte. En effet, les connaissances antérieures du lecteur peuvent faciliter

¹⁴ Dans ce chapitre, nous utilisons le terme de « mémorisation » en sous-entendant « mémorisation des propositions sémantiques du texte ».

sa compréhension (Kendeou & van den Broek, 2007) : un géologue comprendra plus facilement un texte qui décrit les plaques tectoniques qu'un informaticien ; à l'inverse, l'informaticien sera plus performant pour comprendre un texte traitant de programmation logicielle.

On peut toutefois tenter de s'approcher d'une mesure psychophysique en contrôlant au maximum les textes, et en les faisant varier principalement par rapport à une dimension : leur difficulté. Le choix de l'approche psychophysique rend également nécessaire la création d'un grand nombre de textes équivalents. En effet, le but de l'approche est d'obtenir, pour un niveau de difficulté donné, le même résultat quel que soit le nombre de présentations du stimulus. Or, la lecture d'un texte et sa relecture ne sont pas équivalentes : si le lecteur lit mais ne comprend pas le texte à la première lecture, il pourra peut-être le comprendre à la seconde, et répondre correctement au questionnaire de compréhension sans même lire le texte à la troisième. Il faut donc éviter que le sujet puisse relire deux fois le même texte et pour cela, il est nécessaire de disposer d'un grand nombre de textes afin de pouvoir pratiquer la mesure sans souffrir des effets liés à l'apprentissage.

En résumé, pour que le test puisse répondre à notre problématique de comparaison d'interfaces, il doit :

- permettre d'évaluer la mémorisation et les inférences du lecteur,
- être composé d'un grand nombre de textes, ces textes étant :
- de difficulté variable et connue,
- construits de manière à limiter l'effet des connaissances a priori.

Nous verrons dans ce chapitre comment nous avons tenté de répondre à ces problématiques.

6.2 Un générateur automatique de textes pour l'évaluation de la compréhension

6.2.1 Concepts manipulés par le générateur

Pour remplir les objectifs définis précédemment, nous avons développé un générateur automatique de textes. Les textes générés sont de type informatif : ils contiennent une série d'informations ponctuelles à propos d'un sujet. L'avantage de ce type de texte réside dans le fait que les questions de mémorisation y sont très faciles à construire. En revanche, ils peuvent être très impactés par les connaissances antérieures du lecteur sur le sujet. Si le lecteur a une bonne connaissance du sujet, il pourra lire très facilement le texte, voire

répondre aux questions sans avoir à le lire. L'état des connaissances antérieures peut perturber la mesure en permettant à certains lecteurs d'obtenir de meilleures performances sur certains textes. Il semble difficile de construire un corpus de texte qui contiendrait suffisamment de thématiques, pour qu'aucun lecteur ne puisse être évalué sur un texte pour lequel il détient des connaissances a priori. Cette observation pose un problème plus général : comment étudier la construction des connaissances si celles-ci peuvent être acquises a priori, dans l'environnement naturel ?

Dans un tout autre domaine, Gopnik et Sobel (2000) ont été confrontés au même type de problème. Cherchant à comprendre les processus de construction de la causalité chez les jeunes enfants, il leur était impossible d'étudier des phénomènes causaux qu'un adulte aurait pu leur apprendre auparavant. Les auteurs ont alors créé des objets artificiels, associés à des effets de causalité qui n'existent pas dans l'environnement : des « blickets ». Lorsqu'un blicket était posé sur une machine, dénommée un « blicket detector », la machine se mettait en marche. L'expérimentation a consisté à présenter des objets qui déclenchaient la machine, et d'autres qui ne la déclenchaient pas. La capacité de définir les liens de causalité entre les objets et la machine, et donc la capacité à comprendre quels objets étaient des blickets ont ainsi pu être étudiées. Ainsi les auteurs ont pu tester la manière dont les enfants mettaient en jeu et s'appropriaient le principe de causalité sur ces objets nouveaux, indépendamment de toute connaissance antérieure.

Nous nous sommes inspirés de leurs travaux pour créer des textes pour lesquels le lecteur ne peut pas avoir de connaissances a priori sur l'objet. Nos textes décrivent des concepts imaginaires, correspondant soit à une ville, soit à un animal. Les villes et animaux décrits n'existent pas dans la réalité. Ils sont donc représentés par des mots qui n'existent pas dans la langue (des pseudo-mots). Les pseudo-mots ont été créés à l'aide de l'algorithme de génération à partir de trigrammes disponible avec Lexique 3 (New et al., 2001). Cet algorithme sélectionne les trigrammes d'un ensemble de mots fourni par l'utilisateur, et les combine aléatoirement pour former des pseudo-mots. Pour les noms attribués aux animaux, les mots existants utilisés pour générer les pseudo-mots sont ceux dont la longueur est de 7 caractères dans la base Lexique 3 (exemples de pseudo-mots obtenus : « caleron », « baccail », « gripote »). Pour les pseudo-mots désignant des villes fictives, les noms entrés

dans l'algorithme sont issus d'une liste des villes de France¹⁵. Les pseudo-noms de ville ainsi créés étaient comparés à la liste originale, afin de retirer les noms correspondant à de vraies villes (exemples de pseudo-mots retenus : « Cassoue », « Segonne », « Lozerie »). Dans le cas des animaux, nous attribuons un genre à chaque pseudo-mot.

6.2.2 Attribution de caractéristiques au concept

Les textes générés vont correspondre à une description d'un concept irréel. On attribue à chaque concept un ensemble de caractéristiques. Un concept représentant un animal va ainsi « avoir des griffes », « vivre en groupe », « être de petite taille », etc. Pour les villes, les caractéristiques possibles sont « être boisée », « posséder des édifices culturels »,... Afin de pouvoir faire varier les concepts, chaque caractéristique peut prendre plusieurs valeurs : par exemple, une ville peut être « très boisée », « peu boisée » ou « pas du tout boisée ». Chaque valeur de caractéristique est associée à une ou plusieurs formulations, un exemple est donné dans la *Table 3*. Plusieurs formulations sont proposées afin d'éviter des textes trop répétitifs. Les formulations sont des « proto-phrases » : elles contiennent des parties de phrases qui vont être intégrées à la phrase finale telles quelles, et des champs qui vont être adaptés à la phrase (notamment pour le nom du concept, mais aussi pour les adjectifs qui vont s'accorder, des déterminants à adapter,...).

Table 3

Exemples de valeurs attribuables à une caractéristique et des formulations associées

¹⁵ Cette liste est disponible via l'adresse suivante : <http://sql.sh/736-base-donnees-villes-francaises>

Caractéristique : « être boisée »	
Valeur	Formulation(s)
1	1.1) (<i>Nom_ville</i>) n'est pas du tout boisée 1.2) On ne peut voir aucun arbre à (<i>nom_ville</i>) 1.3) Il n'y a pas d'arbres à (<i>nom_ville</i>)
2	2.1) (<i>Nom_ville</i>) est boisée par endroits 2.2) On peut voir des arbres par endroits dans (<i>nom_ville</i>)
3	3.1) (<i>Nom_ville</i>) est très boisée 3.2) On peut voir des arbres un peu partout dans (<i>nom_ville</i>)

L'attribution de caractéristiques au concept va permettre de pouvoir tester la mémorisation du lecteur. En fonction des caractéristiques présentées, nous pourrons vérifier ce que le lecteur a retenu de l'animal ou de la ville. Nous avons construit une question de mémorisation pour chaque caractéristique. Une proposition est donnée au lecteur, il doit déterminer si celle-ci est vraie ou fausse. La réponse correcte dépend de la valeur de la caractéristique (pour un exemple, voir dans la *Table 4*).

Table 4

Question de mémorisation associée à l'exemple de la Table 3

Proposition : Il y a un grand nombre d'arbres à (<i>nom_ville</i>)		
Valeur de la caractéristique :	1 ; 2	3
Réponse correcte :	Faux	Vrai

6.2.3 Construction de traits généraux liés au concept

L'attribution de caractéristiques à un concept permet de générer un texte pour lequel on peut évaluer la mémorisation du lecteur. Cependant, l'évaluation de la compréhension serait incomplète si l'on évaluait uniquement sa mémorisation. La compréhension de haut niveau quant à elle peut être évaluée à l'aide de questions d'inférences. Pour que les textes puissent permettre l'évaluation des inférences du lecteur, nous avons regroupé les

caractéristiques dans des « traits » : des qualités plus générales attribuées au concept. Par exemple, les caractéristiques « avoir_des_usines », « être_boisée », « avoir_un_trafic_automobile » vont être regroupées sous le trait général « être_polluée ». Le fait que la ville soit polluée ou non n'est pas renseigné directement dans le texte. Néanmoins, le lecteur doit pouvoir l'inférer à partir des caractéristiques. Une ville qui est très boisée, sans usine et avec très peu d'automobiles devrait être considérée comme étant peu polluée. Ce serait l'inverse pour une ville comptant des centaines d'usines et n'ayant aucun arbre. Poser la question du niveau de pollution de la ville permet de tester les inférences du lecteur, puisque la réponse porte sur la construction d'une représentation mentale des qualités du concept. En outre, la valeur du trait est plus informative que la valeur d'une caractéristique, car elle peut être positionnée sur une échelle continue. Une ville peut être « pas du tout polluée », « légèrement polluée », « très polluée », en fonction des valeurs des caractéristiques qui la composent. Contrairement aux questions de mémorisation où le choix est dichotomique, la réponse à la question d'inférence « à quel point la ville est polluée ? » peut permettre de mesurer une réponse continue (voir la *Figure 14*).

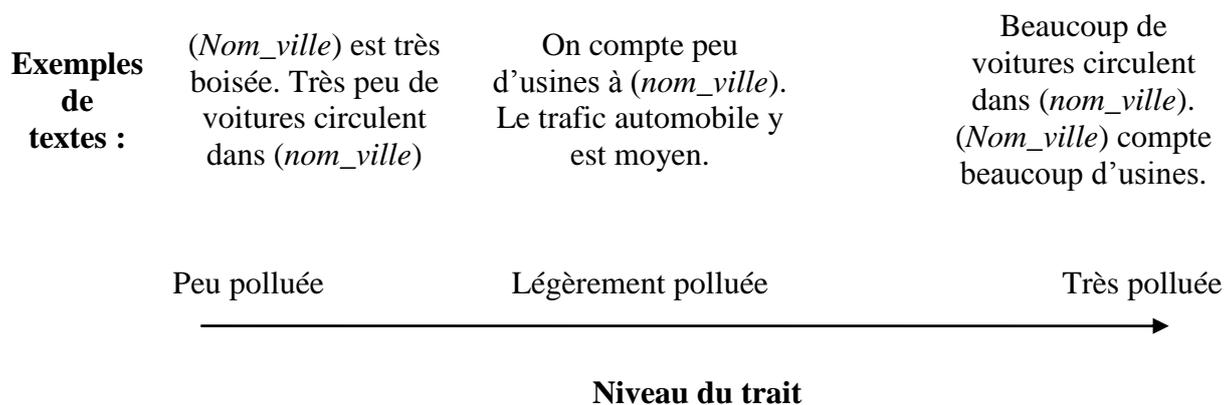


Figure 14. Exemples de textes et niveaux de trait associés.

Pour alimenter le générateur de textes, nous avons donc construit une base de caractéristiques. La base comprend les deux thèmes (« animaux » et « villes »). Le thème « animaux » est constitué de 7 traits généraux (« dangereux », « sociable », « intelligent »,...) et le thème « ville » est constitué de 6 traits (« grande », « polluée », « culturelle »,...). Chaque trait est représenté en moyenne par 4 caractéristiques (écart-type : 1,2), soit un total de 53 caractéristiques. Chaque caractéristiques peut prendre en moyenne 2,7 valeurs (écart-

type : 0,52). Enfin, chaque valeur de caractéristique peut être traduite par en moyenne 1,6 formulations (écart-type : 0.64).

En sélectionnant pour chaque concept des caractéristiques, et attribuant à chaque caractéristique une valeur parmi les valeurs possibles, on peut obtenir un grand nombre de combinaisons de caractéristiques, et donc un grand nombre de textes.

6.2.4 Construction des textes

Au niveau technique, notre générateur n'utilise pas ou très peu de savoirs linguistiques pour produire les phrases en langage naturel. Il ne bénéficie pas des techniques issues du traitement automatique de la langue naturelle (Danlos & Roussarie, 2000). L'algorithme utilise la technique du formulaire (ou « texte à trous »). Un formulaire est composé de proto-phrases où certains éléments sont laissés vides, et remplis selon les paramètres fournis à l'algorithme. Le formulaire peut ainsi modifier le sujet des phrases, ou encore appliquer des règles grammaticales (simples). Cette technique est suffisante pour développer notre générateur automatique de texte, dans la mesure où le contenu des phrases est très contraint par l'attribution de valeurs aux différentes caractéristiques.

L'algorithme procède par génération successive de paragraphes. Le processus de construction d'un paragraphe est décrit dans la *Figure 15*. L'algorithme commence par sélectionner un thème (« animaux » ou « villes »). Un pseudo-mot correspondant au thème est alors choisi pour désigner le concept. L'algorithme attribue au concept un des 6 ou 7 traits du thème. Il sélectionne aléatoirement des caractéristiques du thème : au minimum une caractéristique, au maximum toutes les caractéristiques.

A cette étape, le squelette du paragraphe est défini. Il est nécessaire d'attribuer les valeurs aux caractéristiques qui vont composer le paragraphe. Pour chaque caractéristique sélectionnée, une valeur est attribuée aléatoirement parmi les valeurs possibles. Nous avons vu précédemment, qu'à une valeur peuvent correspondre plusieurs formulations. La formulation qui sera utilisée est choisie aléatoirement. Ne reste alors à l'algorithme qu'à compléter les proto-phrases en insérant le nom du concept, des adjectifs et des déterminants accordés au genre du sujet, ... Afin d'éviter les répétitions, ces éléments peuvent prendre plusieurs formes (surtout les éléments de dénomination du concept : « Le batteck » peut être remplacé par « Cet animal » ; « Vouzeau » peut être remplacé par « La ville », « Elle »...).

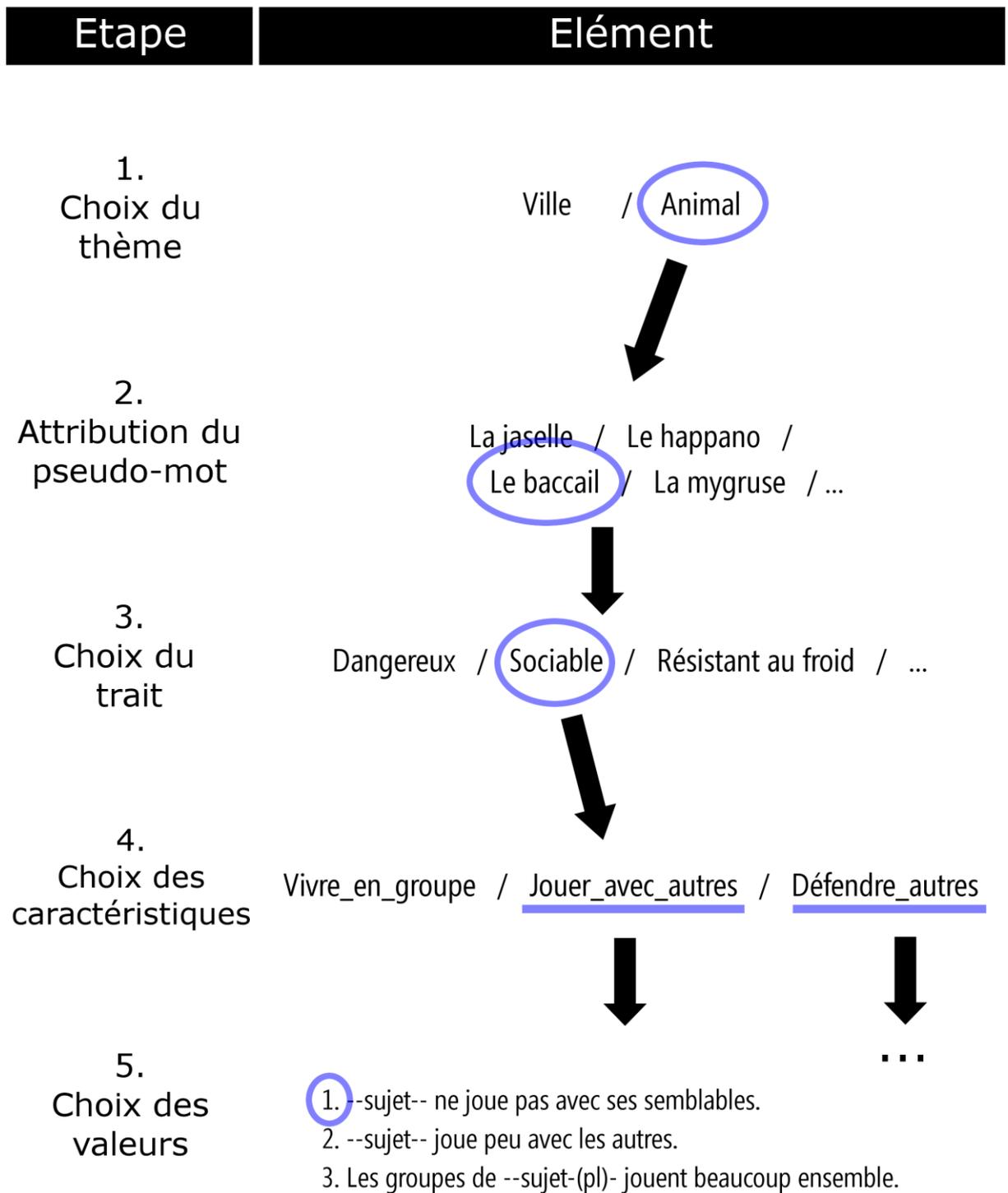


Figure 15. Schéma illustratif du processus de sélection du concept et de ses attributs pour la génération automatique de paragraphes.

Le générateur nous fournit donc des paragraphes décrivant un concept, qui correspondent à un trait, et dont on connaît les caractéristiques mentionnées, ainsi que leurs valeurs. Il est maintenant nécessaire de pouvoir juger de la cohérence des textes, de leur difficulté ainsi que de la valeur des traits pour pouvoir utiliser ces textes dans l'évaluation de la compréhension.

6.3 Attributions des valeurs aux textes et aux questionnaires

6.3.1 Méthodes possibles de calcul des valeurs finales

Les paragraphes sont générés automatiquement, comme décrit précédemment. Néanmoins, rien ne garantit que ces textes soient cohérents : la sélection aléatoire peut faire coïncider deux caractéristiques qui s'opposent (par exemple : un animal qui ne rencontre jamais ses congénères, mais qui joue régulièrement avec eux...). Il est donc nécessaire de pouvoir exclure les textes non cohérents.

Ensuite, pour pouvoir être utilisés dans un test inspiré des méthodes de la psychophysique, il faut pouvoir connaître la difficulté des textes. Une valeur de difficulté doit être calculée pour chaque texte.

Nous avons montré comment les traits (dangerosité de l'animal, pollution de la ville,...) peuvent permettre d'évaluer la construction des inférences du lecteur. Le lecteur devra évaluer le niveau du trait sur une échelle. La valeur du trait est fonction des valeurs des caractéristiques données dans le texte. Il est nécessaire de calculer la valeur du trait à partir des valeurs des caractéristiques, afin de pouvoir la comparer avec la réponse du lecteur.

Nous devons donc trouver un moyen qui permettrait de passer des combinaisons de caractéristiques aux valeurs de cohérence, difficulté et à la valeur du trait. Plusieurs possibilités ont été envisagées pour pouvoir inférer ces trois valeurs dont nous avons besoin.

Une première solution consisterait à utiliser la logique propositionnelle. Il s'agit alors de représenter chaque caractéristique sous forme d'une variable propositionnelle. La combinaison des caractéristiques, liées par des connecteurs logiques, va alors permettre de déterminer les valeurs qui nous intéressent. Par exemple, dans le cas d'un concept représentant un animal imaginaire :

$$(\text{Avoir_griffes} \vee \text{Avoir_cros}) \wedge \text{Attaquer_proies} \rightarrow \text{Dangereux}$$

Ce qui peut se traduire en langage naturel par « Un animal qui a des griffes ou des crocs, et qui s'en sert pour attaquer ses proies est un animal dangereux ». On peut donc inférer la dangerosité de l'animal à partir des valeurs des caractéristiques : un animal qui détient ces caractéristiques est dangereux. A l'inverse, s'il s'agit d'un animal qui a des griffes, mais qui ne s'en sert pas pour attaquer, on déterminera qu'il n'est pas dangereux.

$$(\text{Avoir_griffes} \wedge (\text{Griffes_acérées} \text{ XOR } \text{Griffes_peu_coupantes})) \text{ XOR } (\neg \text{Avoir_griffes} \wedge \neg (\text{Griffes_acérées} \vee \text{Griffes_peu_coupantes})) \rightarrow \text{Cohérent}$$

Cette autre formule propositionnelle permet de vérifier la cohérence de la phrase : si le texte mentionne un animal qui a des griffes qui sont acérées ou peu coupantes, ou un animal qui n'a pas de griffes, le texte est cohérent. Par contre, si le texte mentionne que l'animal n'a pas de griffes, et que ses griffes sont tranchantes, alors le texte n'est pas cohérent.

Enfin, la difficulté du texte peut être déterminée par la suite des caractéristiques, qui confirment toutes la même idée, ou alors qui vont créer des oppositions dans la construction de sens que peut faire le lecteur (caractéristique inattendue par rapport aux autres).

$$\text{Avoir_griffes} \wedge \text{Griffes_acérées} \wedge \text{Capable_couper_acier} \wedge \text{Attaquer} \rightarrow \text{Facile}$$

$$\text{Avoir_griffes} \wedge \text{Griffes_acérées} \wedge \text{Capable_couper_acier} \wedge \underline{\text{Les utiliser uniquement pour grimper arbres}} \rightarrow \text{Difficile}$$

Néanmoins, l'utilisation de la logique propositionnelle entraîne un défaut majeur : toutes les valeurs de vérité –celles des caractéristiques et celles des valeurs finales (trait, cohérence, difficulté) – ne peuvent prendre que deux valeurs : « Vrai » ou « Faux ». Cela pose problème dans la mesure où les caractéristiques que nous avons définies peuvent prendre différentes valeurs. Il serait intéressant également de pouvoir définir les valeurs des traits sur un continuum (ici, un animal avec des griffes sera tout aussi dangereux qu'un animal avec des griffes tranchantes, des crocs et qui attaque ses proies).

Afin de pouvoir manipuler des valeurs continues, il est possible de remplacer l'utilisation de la logique propositionnelle par celle de la logique floue (Zadeh, 1965). La logique floue manipule des valeurs numériques (des degrés de vérité), plutôt que des valeurs de vérité binaires. On peut ainsi écrire des propositions du type :

$$\text{Avoir_griffes} \text{ fz } \vee \text{ Avoir_crocs} \rightarrow \text{Dangereux}$$

Le OU de la logique floue ($fz \vee$ pour « *fuzzy OR* ») permet de tenir compte du degré de vérité de chacune des deux caractéristiques. Le calcul est procédé de la manière suivante : $A fz \vee B = A + B - A*B$. Concrètement : un animal qui a des griffes et des crocs sera plus dangereux qu'un animal qui n'a qu'une seule des deux caractéristiques.

L'utilisation de la logique floue permettrait de pallier le problème de la binarité du formalisme en logique propositionnelle. Mais l'utilisation du langage logique (propositionnel ou flou) pour formaliser nos textes laisse encore apparaître deux défauts :

Il est nécessaire d'écrire manuellement chaque formule, pour chaque trait et pour chacune des trois valeurs recherchées (valeur du trait, cohérence, difficulté). La tâche peut donc s'avérer très longue étant donné le nombre de possibilités du générateur. Ce défaut n'est néanmoins pas insurmontable.

Le second défaut est plus contraignant : c'est au concepteur de définir les formules logiques aboutissant aux valeurs finales. Néanmoins, le point de vue du concepteur n'est pas nécessairement le même que celui de tous les lecteurs. Par exemple, si le concepteur juge qu'un animal est plus dangereux lorsqu'il a des griffes que quand il a des crocs, c'est peut-être l'inverse pour la majorité des lecteurs.

Nous nous sommes donc orientés vers une autre technique pour calculer les valeurs finales : les réseaux bayésiens. Un réseau bayésien se représente sous la forme d'un graphe orienté acyclique probabiliste. Il permet de décrire et d'attribuer des valeurs numériques aux relations causales entre des variables. En ce sens, le formalisme d'un réseau bayésien se rapproche de celui de la logique floue. Néanmoins, les relations ne sont pas de type déterministe, mais sont de type probabiliste. Ils ne représentent pas des degrés de vérité, mais des probabilités d'observation d'une variable étant données d'autres variables. L'intérêt majeur des réseaux bayésiens tient dans leur capacité à généraliser la probabilité d'un état de choses à partir de cas particuliers. Une fois le réseau construit, il peut inférer des valeurs inconnues à partir des informations qui lui sont présentées. Dans le cas des textes générés automatiquement, des réseaux bayésiens pourraient inférer les valeurs finales (inconnues) à partir des données connues : les valeurs des caractéristiques.

6.3.2 Principe de calcul des valeurs finales par réseaux bayésiens

Le concepteur d'un réseau bayésien (ou un expert dans le cas de la description d'un domaine spécifique) peut incorporer dans le réseau des connaissances a priori, en structurant les relations de causalité de manière plus ou moins complexe. A l'inverse, le réseau peut être affranchi de toute connaissance a priori, en lui attribuant une structure causale minimale (voir les exemples de la figure *Figure 16*).

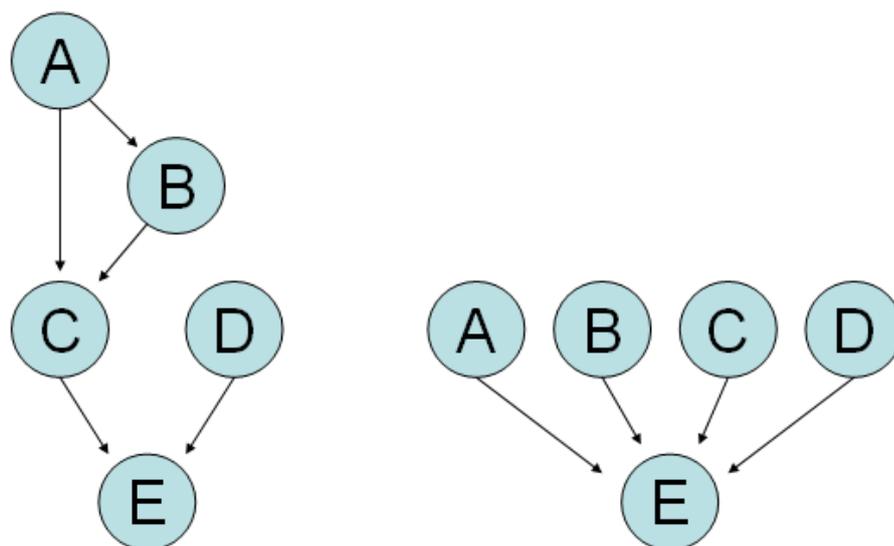


Figure 16. Exemples de graphes représentant des réseaux bayésiens. Les relations causales sont complexes dans le graphe de gauche, et minimales dans le graphe de droite (réseau bayésien naïf).

Il est possible de décrire nos textes par des réseaux bayésiens : en affectant un nœud à chaque caractéristique, et en affectant le nœud terminal (les nœuds « E » dans la *Figure 16*) pour les valeurs que l'on veut inférer à partir des caractéristiques (valeur du trait, difficulté, cohérence). Chaque nœud représentant une caractéristique doit pouvoir prendre les différentes valeurs de la caractéristique. Les nœuds terminaux doivent prendre une valeur continue.

La structure du réseau bayésien en elle-même ne suffit pas pour inférer les valeurs des nœuds terminaux. Il faut également connaître les probabilités conditionnelles du réseau, c'est-à-dire la probabilité d'observer un état d'un nœud (typiquement, la valeur de sortie), étant données les valeurs des nœuds d'entrée. Ces probabilités vont permettre de répondre à des questions du type : « Quelle est la probabilité que l'animal soit dangereux, étant donné qu'il a des griffes et des crocs ? », « Quelle est la probabilité que le texte soit incohérent, sachant que l'animal n'a pas de griffes et que ses griffes sont acérées ? ». En réalité, comme les valeurs de

sortie sont continues, une seule probabilité ne suffit pas pour décrire ces valeurs. Elles doivent être décrites par une distribution de probabilité conditionnelle : c'est cette distribution qui nous permettra de connaître la probabilité associée à chaque valeur de sortie possible, en fonction des valeurs attribuées aux autres nœuds.

Afin de pouvoir calculer ces distributions de probabilités, il est nécessaire pour le réseau bayésien d'apprendre à partir d'observations. Un algorithme d'apprentissage va permettre de modifier les distributions des probabilités en fonction des données observées. On parle « d'apprentissage bayésien ». Les distributions de probabilités avant l'apprentissage sont appelées « distributions a priori », et les distributions obtenues par l'apprentissage sont les « distributions a posteriori ».

En donnant des exemples de textes et des valeurs finales (difficulté, cohérence, valeur du trait) à apprendre à des réseaux bayésiens, nous pourrions obtenir automatiquement les valeurs finales pour n'importe quel texte généré. Mais il est nécessaire dans un premier temps de recueillir des données pour l'apprentissage.

6.3.3 Recueil d'observations pour le corpus d'apprentissage

Le corpus de données pour l'apprentissage a été obtenu à l'aide de lecteurs. Six lecteurs ont accepté de participer au recueil de données. Il leur a été proposé de lire des textes, et de répondre à différentes questions dans le but de construire un test de compréhension de lecture. Nous avons envoyé à tous les participants un logiciel pour lire des textes générés automatiquement, et répondre aux questions. Les lecteurs volontaires pouvaient démarrer le logiciel, et effectuer la tâche à n'importe quel moment de la journée, lorsqu'ils avaient du temps à y consacrer.

En démarrant le logiciel, les lecteurs pouvaient lire des instructions générales. Il leur était rappelé l'utilité de leur participation (construire un test de compréhension), puis la consigne suivante était affichée : « *Vous allez devoir lire un texte court. Celui-ci concerne un concept imaginaire (une ville, un animal). Le texte décrit certaines caractéristiques de ce concept.* » Une seconde page affichait la suite de la consigne : « *Une fois le texte lu et compris, vous devez répondre à des questions, notamment sur un terme général concernant le concept. Utilisez la souris pour sélectionner votre réponse. Prenez votre temps, vous pouvez relire le texte autant de fois que vous le voulez. Une fois toutes les réponses données, vous pouvez passer à la suite en cliquant sur VALIDER.* » La consigne de la page suivante

concernait la cohérence des textes : « *Certains textes sont incohérents (le concept ne peut logiquement pas être comme décrit dans le texte, ou alors les phrases ne sont pas cohérentes entre elles). Dans ce cas, indiquez-le et passez à la suite.* ». Une dernière page expliquait les modalités pour l'envoi des données (envoi par mail), et encourageait les lecteurs à reprendre le logiciel autant de fois qu'ils le souhaitent. Cette consigne devait être affichée entièrement si c'était la première fois que le lecteur utilisait le logiciel, sinon il pouvait passer automatiquement aux textes.

L'interface de réponse est telle qu'affichée sur la *Figure 17*. Le titre indique le numéro du texte (correspondant au nombre de textes déjà traités par le lecteur + 1). Un sous-titre « thème » indique si le concept porte sur un animal ou une ville. S'en suit un paragraphe généré automatiquement par le générateur.

La partie basse de l'interface permet d'entrer les réponses. La première réponse porte sur la cohérence : le lecteur doit déterminer si le texte est cohérent (« oui / non »). Si la réponse est « non », le bouton « VALIDER » devient cliquable, et le lecteur peut passer au texte suivant. Si la réponse est « oui », le lecteur peut évaluer la valeur du trait sur une échelle (qui rend une valeur entière comprise entre 0 et 10). Enfin, lorsque le lecteur a donné une valeur de trait, il peut juger sur une échelle la difficulté qu'il a éprouvée pour répondre à la question. Une fois la difficulté jugée, il peut passer au texte suivant.

Texte n° 204

Thème : animal

La majorité du temps du caleron est consacrée à ses activités. Cet animal dort environ 7 heures par jour. On ne peut pas vraiment dire que le caleron soit très actif.

Ce texte est-il cohérent ? OUI NON

D'après le texte, cet animal est plutôt (placer le curseur) :

Pas du tout paresseux |-----| Très paresseux

Faire ce choix vous a semblé :

Pas du tout difficile |-----| Très difficile

Figure 17. Visualisation du logiciel de recueil de données pour l'apprentissage.

Le logiciel génère alors un fichier résultat, comprenant les valeurs des caractéristiques attribuées au texte, et les valeurs finales données par les lecteurs (seulement la cohérence si le texte n'est pas cohérent, cohérence, valeur du trait et difficulté dans le cas contraire).

La participation des lecteurs a ainsi permis de recueillir des données de valeurs finales pour 374 textes.

6.3.4 Construction des réseaux bayésiens

Les données récupérées selon la procédure détaillée dans la sous-section précédente n'ont pas toutes été utilisées pour entraîner les réseaux bayésiens. Nous commençons par sélectionner aléatoirement 80% des données pour l'entraînement, les 20% restants seront utilisés par la suite pour évaluer la fiabilité des inférences réalisées par les réseaux bayésiens.

Pour chaque trait (dangerosité, sociabilité,...) de chaque thème (animal, ville), nous construisons 3 réseaux bayésiens. L'un va permettre d'inférer la cohérence, un autre la valeur du trait et le dernier la difficulté.

La construction, ainsi que l'entraînement et les inférences des réseaux bayésiens ont été réalisés à l'aide de la toolbox « Bayes Net Toolbox » pour Matlab (Murphy, 2001).

Dans un premier temps, il faut définir la structure du réseau, c'est-à-dire dessiner le graphe qui contient les nœuds correspondant aux caractéristiques et les nœuds des valeurs finales. Un exemple de structure possible est donné dans la *Figure 18*. Chaque réseau contient un premier niveau de nœuds, qui correspondent aux caractéristiques du concept. Le nombre de caractéristiques dépend du trait correspondant (4 nœuds au premier niveau en moyenne). Pour pouvoir gérer les interactions des caractéristiques 2 à 2, le réseau contient également un niveau de nœuds « cachés » : ces nœuds ne recevront jamais de valeurs observées, ils serviront uniquement au calcul des valeurs finales. Il y a autant de nœuds cachés que de combinaisons possibles de deux caractéristiques distinctes. Chaque paire de nœuds caractéristiques est reliée à un nœud caché. Enfin, tous les nœuds cachés sont reliés au nœud final.

Les nœuds correspondant aux caractéristiques sont des nœuds discrets : ils ont un nombre d'états fini (le nombre de valeurs possibles de la caractéristique qu'ils représentent), auxquels sont associés des probabilités. Les nœuds de la couche cachée, et le nœud final doivent représenter des valeurs continues, leur nombre d'état n'étant pas fini. Ces nœuds sont « gaussiens » : ils ont un continuum de valeurs (situées entre 0 et 1), auxquelles sont associées des valeurs de probabilités, présentées sous forme d'une distribution gaussienne (définie par deux paramètres : la moyenne et l'écart-type de la distribution).

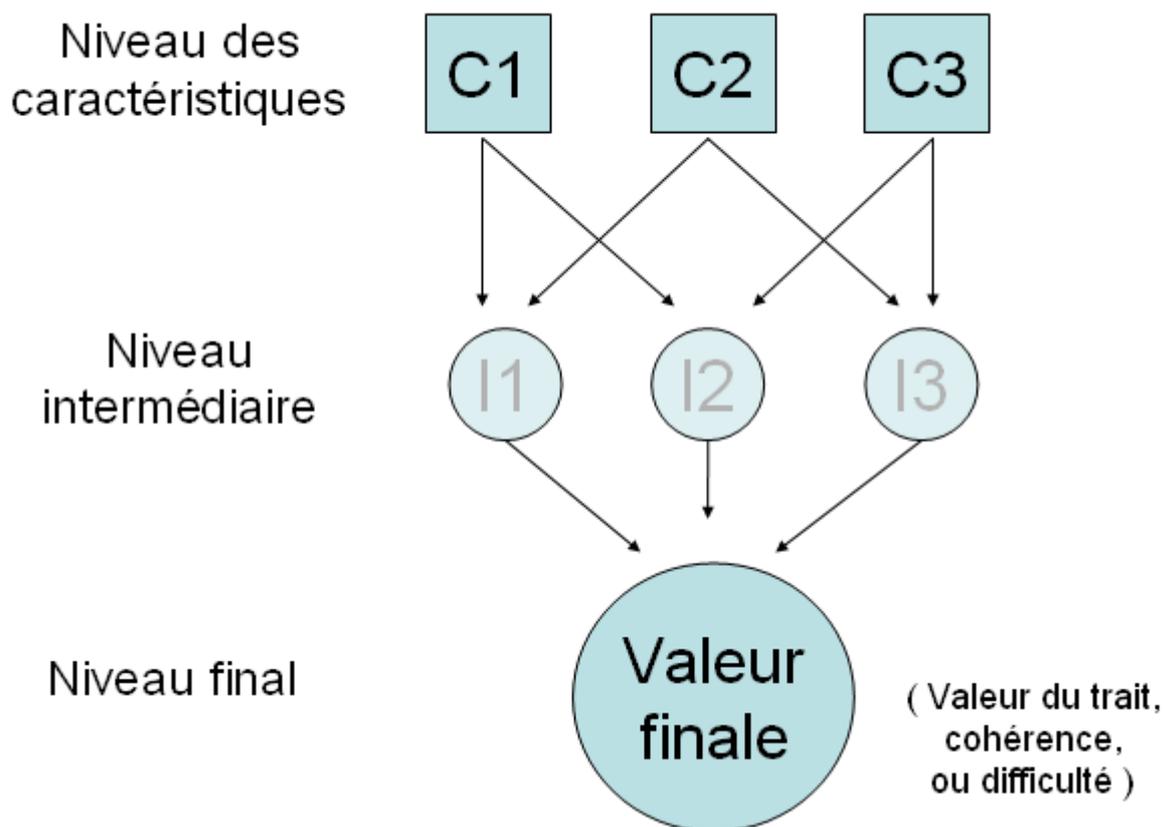


Figure 18. Exemple d'architecture de réseau bayésien appliqué à nos textes. Les nœuds grisés sont les nœuds cachés du réseau. Les nœuds représentés par des carrés sont les nœuds discrets, les nœuds ronds sont continus (contenant une distribution gaussienne).

Les probabilités attribuées aux différents nœuds avant l'apprentissage sont uniformes : c'est-à-dire que le réseau ne contient aucune connaissance a priori sur les probabilités conditionnelles (la probabilité pour un animal d'être très dangereux n'est pas plus grande ni moins grande que de n'être pas du tout dangereux, quelles que soient les valeurs des caractéristiques,...).

6.3.5 Apprentissage des réseaux bayésiens

Les valeurs obtenues lors du recueil de données, et qui correspondent au trait testé sont extraites pour pouvoir procéder à l'apprentissage.

L'apprentissage consiste à attribuer aux nœuds du graphe des distributions de probabilités qui minimisent les écarts avec les données observées. L'algorithme d'apprentissage utilisé est l'algorithme EM (Lauritzen, 1995). Sans entrer dans ses détails techniques, l'algorithme a l'avantage de pouvoir calculer les valeurs des distributions malgré la présence de données manquantes (les nœuds cachés, ou les caractéristiques non

mentionnées). Le moteur utilisé pour l'inférence est l'algorithme de l'arbre de jonction (*junction tree engine*). Pour chaque réseau, l'apprentissage est composé de 50 itérations de l'algorithme EM.

Une fois la phase d'apprentissage achevée, le réseau bayésien est capable d'inférer une distribution de probabilité pour une valeur finale, à partir de n'importe quelle combinaison de caractéristiques (observées ou non). Nous inférons les distributions liées aux trois valeurs finales pour chaque combinaison possible (et donc pour chaque texte potentiellement généré par le générateur de textes).

Pour obtenir une estimation des trois variables finales, nous sélectionnons à chaque fois la valeur moyenne de la distribution inférée. Il s'agit donc de valeurs continues. Pour la valeur de cohérence, il n'y a pas d'intérêt à retenir une valeur continue : un texte est soit cohérent (et pourra être montré au lecteur), soit non cohérent (et ne sera pas affiché). Nous appliquons donc un seuil aux valeurs de cohérence pour obtenir une variable de cohérence binaire : les textes dont la valeur de cohérence est strictement supérieure à 0,8 sont conservés, les autres sont considérés comme incohérents et rejetés.

6.3.6 Vérification de la fiabilité

Pour vérifier la fiabilité des valeurs inférées, nous utilisons la technique de validation croisée : les valeurs inférées par les réseaux bayésiens sont comparées aux 20% des données qui n'ont pas servi à l'apprentissage (corpus de vérification). Pour la cohérence, on compare les choix issus des réseaux bayésiens aux choix des lecteurs (considérer les textes comme cohérents ou non). Pour un total de 74 textes traités par l'algorithme, 63,5% ont été bien classés (textes jugés cohérents par les lecteurs acceptés et textes jugés non cohérents rejetés). On observe 9,5% d'erreurs (textes jugés non cohérents par les lecteurs, mais acceptés selon les réseaux). Enfin, il y a 27,0% de fausses alarmes (textes rejetés alors qu'ils avaient été jugés cohérents). Le taux important de fausses alarmes n'est pas problématique en soi. Il aurait pu être diminué en fixant un seuil d'acceptation moins restrictif que 0,8. Mais cela aurait engendré plus d'erreurs, et nous avons préféré augmenter le taux de fausses alarmes plutôt que ce taux d'erreurs : il nous a semblé plus prudent de rejeter beaucoup de textes plutôt que d'accepter trop de textes incohérents.

Pour les variables continues (valeur du trait et difficulté), les indicateurs permettant d'estimer la fiabilité sont la moyenne et l'écart-type des écarts signés entre les valeurs prévues

par le réseau bayésien et les valeurs données par le lecteur. Pour l'estimation de la valeur du trait, la moyenne des écarts est de -0,05 (écart-type : 0,28). Pour l'estimation de la difficulté, l'écart moyen est de -0,24 (écart-type : 0,35). La valeur de trait estimée par les réseaux bayésiens est donc assez fiable en moyenne. La valeur de difficulté est nettement moins précise.

6.3.7 Sélections des textes

A ce stade, nous disposons donc de paragraphes associés à des traits. Des questions de mémorisation, ainsi que d'inférences peuvent être posées au lecteur (grâce au calcul bayésien de la valeur du trait). Il est possible de ne conserver que les textes cohérents. On connaît approximativement le niveau de difficulté associé à chacun de ces paragraphes. Pour pouvoir utiliser ce matériel dans un test de compréhension, il reste à regrouper ces paragraphes pour former des textes.

Pour pouvoir créer des textes de difficulté faible, moyenne et élevée, nous faisons varier la difficulté inférée par les réseaux bayésiens, ainsi que la longueur des textes. Les paragraphes sont d'abord classés selon leur valeur de difficulté. Les paragraphes « faciles » sont ceux dont la valeur de difficulté inférée par les réseaux est la plus faible, les paragraphes « difficiles » sont ceux dont la valeur de difficulté est la plus élevée.

- Les textes « faciles » sont composés de 2 paragraphes faciles.
- Les textes « moyens » sont composés de 2 paragraphes faciles, et d'un paragraphe difficile.
- Les textes « difficiles » sont composés d'un paragraphe facile, et de 3 paragraphes difficiles.

Nous avons créé 10 corpus de 6 textes, chacun comportant 2 textes faciles, 2 textes moyens et 2 textes difficiles. Chaque paire de textes comprend un texte traitant d'un animal imaginaire et un texte traitant d'une ville imaginaire. Pour chaque texte, 4 questions de mémorisation étaient sélectionnées aléatoirement parmi les questions possibles, et 2 questions d'inférences étaient sélectionnées aléatoirement¹⁶. Malgré l'attribution automatique de la valeur de cohérence, nous avons vérifié manuellement la cohérence de chaque texte.

¹⁶ Le nombre minimum de paragraphes dans un texte étant de 2, le nombre de questions d'inférence à été fixé à deux pour que chaque texte ait le même nombre de questions. Etant donnée la possibilité de répondre correctement au hasard pour les questions de mémorisation, nous avons prévu un plus grand nombre de questions pour cette modalité (4 par texte).

6.4 Procédure

Chacun des 6 textes est affiché indépendamment, l'un après l'autre. A l'affichage, chaque texte est précédé d'un titre : le nom de la ville ou de l'animal imaginaire. Suit le texte, qui peut être affiché de deux manières. La première condition correspond à une mise en page de type paragraphe : chaque paragraphe débute par un retour à la ligne et par un retrait supplémentaire. La seconde condition ne contient pas de mise en forme : tous les paragraphes s'enchaînent sans retour à la ligne (un seul bloc de texte). Ces conditions sont attribuées aléatoirement, de manière à ce qu'un texte sur deux soit affiché sans mise en forme, et de manière répartie (un texte facile est mis en forme, l'autre non, un texte moyen est mis en forme, l'autre non...).

Le texte est affiché sur une seule page, si celle-ci est plus longue que la hauteur de l'écran, il est nécessaire de faire défiler le texte pour pouvoir le lire en entier. En bas de page se situe un bouton pour confirmer la fin de la lecture et passer aux questions.

Au début du test, on explique au sujet que le but est de mesurer sa compréhension. La consigne donnée au sujet avant de d'afficher le premier texte est la suivante : *« Vous allez devoir lire une série de textes. Ces textes décrivent des animaux ou des villes qui n'existent pas. Vous devez lire chaque texte assez attentivement pour pouvoir répondre à des questions de compréhension. Si le texte est plus long que l'écran, vous pouvez le faire défiler¹⁷. Si vous le voulez, vous pouvez relire plusieurs fois le texte, ou revenir sur un paragraphe déjà lu. Une fois que vous avez fini de lire, cliquez sur le bouton qui se situe en bas du texte. Vous devrez alors répondre à 4 questions sur des détails du texte, et à 2 questions sur des caractéristiques générales de l'animal ou de la ville. Il est normal que vous n'arriviez pas à répondre à toutes les questions, dans ce cas répondez au hasard¹⁸. Vous devrez lire 6 textes au total. »*.

Les 4 questions de mémorisation sont présentées les unes après les autres. Une affirmation est présentée au centre de l'écran (exemple : « Il n'y a jamais de spectacles à Clacour. », et deux boutons « Vrai » et « Faux » sont situés respectivement en bas à gauche et en bas à droite de l'écran. Le sujet doit appuyer sur le bouton correspondant à sa réponse.

¹⁷ La modalité pour faire défiler le texte était précisée en fonction du support. Les modalités sont détaillées dans 8.2.4 Interfaces de réponse.

¹⁸ La formulation de cette phrase de la consigne est sans doute maladroite : elle correspond au fait que la tâche est à choix forcée. Une formulation n'incitant pas le sujet à répondre au hasard serait plus judicieuse.

Les 2 questions d'inférences sont aussi présentées l'une après l'autre. En haut de l'écran, un début de consigne commun à chaque question est affiché : « D'après l'interprétation que vous faites de ce texte, [cet animal] / [cette ville] vous semble : ». En dessous, au centre de l'écran, nous affichons une barre avec un curseur. Chaque extrémité de la barre est accompagnée d'un titre : la partie gauche a le titre « Pas du tout » suivi de l'adjectif qui correspond au trait (dangereux, culturelle, ...), et le titre de droite est « Très » et l'adjectif. Le curseur est positionné par défaut au milieu de la barre, le sujet doit le faire glisser à l'endroit qui correspond à son jugement. Une fois que le sujet a positionné le curseur, il doit appuyer sur un bouton « Valider » situé en bas de l'écran.

Une fois que le sujet a répondu à toutes les questions, un écran vide s'affiche. Il peut appuyer quand il le souhaite sur cet écran pour déclencher l'apparition du texte suivant.

6.5 Traitement des données

Les principales variables issues de notre test sont celles liées aux questions de mémorisation et d'inférence. Pour chaque texte lu, nous mesurons le taux d'erreurs aux questions (le nombre de mauvaises réponses divisé par le nombre de questions). Pour les inférences, nous récupérons la position du curseur qui est représentée par une valeur numérique allant de 0 à 1 (0 correspond au curseur placé à l'extrémité gauche et 1 correspond au curseur placé à droite). On calcule l'écart entre cette valeur et la valeur du trait inférée par le réseau bayésien. Les valeurs absolues des deux écarts sont moyennées.

D'autres variables sont mesurées : la durée totale en secondes entre le début de l'affichage du texte et le moment où l'utilisateur clique sur le bouton de fin de lecture et la durée moyenne de réponse aux questions (entre le début de l'affichage de la question et la validation de la réponse).

Toutes ces variables sont rapportées pour chaque texte, afin de pouvoir identifier les deux effets liés aux facteurs difficulté et mise en page.

6.6 Tâche « Remember-Know »

Le test présenté permet de vérifier la mémorisation des éléments du texte. Nous avons également voulu observer quel type de mémoire était mis en œuvre par le lecteur. Pour ce faire, nous avons utilisé la tâche « Remember versus Know » (Tulving, 1985), voir 2.1.3 *La compréhension*.

Chaque texte présenté au sujet contient un pseudo-mot qui est le nom de l'animal ou de la ville imaginaire. Nous testons la mémorisation de ce pseudo-mot. Pour cela, nous citons au lecteur une liste de pseudo-mots. Dans cette liste, un tiers des pseudo-mots sont les noms des concepts présentés dans les textes. Les deux autres tiers sont d'autres pseudo-mots (des distracteurs), construits de la même manière que les noms de concepts (voir 6.2.1 *Concepts manipulés par le générateur*).

Pour chaque pseudo-mot cité, le sujet doit d'abord déterminer s'il s'agit d'un mot qui a été présenté dans les textes précédents. Si le sujet répond « oui », il lui est demandé s'il se souvient du nom (*remember*) ou s'il sait juste qu'il l'a vu (*know*). Les deux questions sont posées l'une après l'autre, car poser toutes les questions en une seule fois (« Est-ce un mot nouveau, ou un mot que vous vous souvenez d'avoir vu, ou que vous savez que vous avez vu ? ») peut entraîner un biais de réponse libéral : et il est montré que dans ce cas, les sujets jugent moins de mots comme étant « nouveaux » (Hicks & Marsh, 1999).

Nous obtenons donc un taux de bonnes détections, un taux de fausses alarmes ainsi qu'un taux de rejets incorrect. Pour dégager des indicateurs sur la tâche RK pour chaque support (chaque set de 6 textes), nous appliquons un des modèles d'analyse proposés par Snodgrass et Corwin (1988) : le modèle non-paramétrique. Ce modèle permet de calculer un indice de discrimination A' , qui reflète la capacité qu'a eu le sujet de bien classer les pseudo-mots selon qu'ils étaient présentés dans le texte ou non. Le modèle permet aussi de calculer un indice de biais B'' qui est lié au critère de décision du sujet, une valeur faible correspondant à un biais conservateur, alors qu'une valeur élevée correspond à un biais libéral.

Les taux de bonnes détections et de fausses alarmes sont d'abord transformés pour ne pas avoir de valeurs égales à 0, qui poseraient problème dans le calcul final.

Le taux de bonnes détections modifié est $H = (\text{nombre_bonne_detect} + 0,5) / (\text{nombre_items_resents} + 1)$.

Le taux de fausses alarmes modifié est $FA = (\text{nombre_fausses_alarmes} + 0,5) / (\text{nombre_items_absents} + 1)$.

Les indicateurs sont calculés de la manière suivante :

- Si $H \geq FA$:

$$A' = 0,5 + \frac{(H - FA) * (1 + H - FA)}{4H * (1 - FA)}$$

$$B'' = \frac{H * (1 - H) - FA * (1 - FA)}{H * (1 - H) + FA * (1 - FA)}$$

- Si $H < FA$:

$$A' = 0,5 - \frac{(FA - H) * (1 + FA - H)}{4FA * (1 - H)}$$

$$B'' = \frac{FA * (1 - FA) - H * (1 - H)}{FA * (1 - FA) + H * (1 - H)}$$

Chacun des deux indices est calculé pour les réponses « Remember » et pour les réponses « Know » ; nous disposons donc de 4 indices en appliquant ce modèle.

6.7 Mesure de la charge attentionnelle

Nous avons vu dans la sous-section 2.1.3 *La compréhension* que le processus de compréhension est étroitement lié aux aspects attentionnels du lecteur. Une mesure concomitante de la charge attentionnelle serait donc nécessaire pour que l'évaluation du processus de compréhension soit complète. Nous procéderons à cette mesure, à partir des résultats observés dans l'expérimentation présentée dans la section 7.2 *L'effet de la lecture électronique sur la posture*. La mise en place et le traitement de cette mesure seront détaillés dans 8.3 *Recueil des données*.

Les textes générés par l'algorithme présenté dans ce chapitre sont donc contrôlés, et permettent de répondre aux pré-requis que nous nous étions fixés : permettre une mesure de la compréhension à différents niveaux (mémorisation, inférences) en nous affranchissant des connaissances a priori du lecteur, et en proposant des textes dont la difficulté est connue.

Chapitre 7 : Aspects posturaux de la lecture électronique

Nous abordons dans ce chapitre le lien entre la posture et la lecture électronique dans deux directions. Deux expériences que nous avons menées y sont rapportées. Dans une première section, nous questionnons l'impact des postures engendrées par les nouveaux formats de dispositifs d'affichage (tablettes) sur les performances de lecture. Dans une seconde section, l'effet de différents types de lecture (associés à des charges attentionnelles plus ou moins élevées) sur la posture du lecteur est observé.

7.1 L'effet de la posture sur la lecture électronique

Cette section de la thèse est une traduction d'un article que nous avons publié sous le titre « *Reading tilted: Does the use of tablets impact performance? An oculometric study* » (Perrin, Paillé, & Baccino, 2014b), pouvant se traduire par « Lire de travers : l'utilisation des tablettes entraîne-t-elle une diminution des performances ? Une étude oculométrique ». L'article original est accessible depuis le site de l'éditeur Elsevier, dans la revue *Computers in Human Behavior*. Cette traduction de l'article est présentée ici uniquement dans un but de partage académique.

Résumé

L'utilisation dans notre vie courante d'appareils électroniques tels que les tablettes induisent de nouveaux comportements posturaux et l'influence de ceux-ci sur les processus cognitifs reste méconnue. Dans cette étude, les aspects posturaux de la lecture sur tablette électronique sont investigués, afin de déterminer si la vitesse de lecture ou la compréhension sont impactées par différentes positions de tête ou de tablette. Le but premier de l'étude est d'évaluer l'effet d'une inclinaison latérale de la tête (roulis) ou de la tablette sur la performance de lecture. Notre étude montre qu'une faible inclinaison n'impacte pas la lecture, les lecteurs étant capables de s'adapter à la situation. Pour chaque variable oculométrique mesurée, nous avons trouvé une forte corrélation entre chaque condition d'inclinaison de la tête et/ou de la tablette (pour chacune, $r > ,73$). Un roulis de la tête, ou encore certains stimuli visuels particuliers peuvent engendrer un mouvement des yeux appelé « cycloverision ». Une seconde expérience a été menée pour vérifier la présence de ce mouvement oculaire, qui pourrait expliquer, dans une certaine mesure, l'adaptation observée dans la première expérience.

7.1.1 Introduction

L'évaluation des performances visuelles et de la lecture sur un affichage électronique est un champ important de la recherche en psychologie cognitive et en ergonomie, du fait de l'importance croissante que prennent les ordinateurs dans le domaine du travail ainsi que dans les loisirs. L'émergence des écrans a marqué radicalement l'histoire de la lecture, et provoqué une rupture avec des millénaires de lecture sur des supports physiques tels que : le parchemin, le manuscrit, les livres, etc. (Baccino, 2004). La plupart des études du domaine ont été consacrées à étudier la différence en termes de performance de lecture entre le papier et les écrans (Dillon, 1992; Gould et al., 1987; Noyes & Garland, 2008). Si la lisibilité a longtemps été meilleure sur papier, cette différence tend aujourd'hui à disparaître. Les améliorations de résolution d'affichage, du contraste et le développement de nouvelles technologies comme l'encre électronique semble conduire à des performances identiques entre les nouveaux écrans et le papier (Siegenthaler et al., 2011).

Les aspects visuels ne sont pas les seuls à avoir évolué : les postures adoptées par les lecteurs changent également. Tandis que l'ordinateur était autrefois uniquement destiné à être installé sur un bureau, il est désormais possible d'utiliser des écrans légers et mobiles tels que les tablettes dans des contextes plus variés. Cela donne au lecteur une plus grande liberté et lui permet d'adopter de nouvelles postures qui pourraient cependant affecter ses performances de lecture. Le but de notre étude est de questionner l'effet de la posture sur le processus de lecture.

Straker et al. (2008) ont comparé les comportements posturaux d'enfants utilisant un ordinateur de bureau, une tablette, et du papier. L'un de leurs résultats a montré que l'utilisation de la tablette pouvait induire des postures moins symétriques que par rapport à un écran traditionnel. Une inclinaison latérale de la tête (roulis) de 11° en moyenne a été mesurée pendant l'utilisation de la tablette, contre un angle de 4° pour l'écran d'ordinateur. Néanmoins, l'effet de cette posture sur la lecture n'a à ce jour pas été étudié. Dans la situation où soit la tête, soit la tablette est penchée, le lecteur ne lit pas un texte horizontal, mais un texte penché relativement à la tête. Cohen, Dehaene, Vinckier, Jobert et Montavont (2008) ont montré que l'identification d'un mot est ralentie lorsque ce mot est présenté non-horizontalement. Des phénomènes similaires ont été traités par Hock et Tromley (1978) au travers du paradigme classique de rotation mentale appliqué à l'identification de lettres. Les sujets devaient déterminer s'ils voyaient une lettre « F » normale ou inversée pour différentes

valeurs de rotation. Les temps de réponses étaient les plus longs lorsque la lettre était fortement tournée. Ils étaient plus courts quand la lettre était droite. Enfin ils augmentaient linéairement avec de petites quantités de rotation, même si la lettre était perçue comme étant droite. Ce type de résultat pourrait-il être observé dans une situation plus écologique, à savoir, pendant la lecture de textes entiers, affichés dans leur intégralité et présentés avec une petite quantité d'inclinaison ? Une hypothèse directe pourrait mener à penser que si le traitement d'une lettre ou d'un mot est impacté par sa rotation, même pour une rotation très faible, alors le traitement d'un texte entier (c'est-à-dire le traitement successif de plusieurs mots) devrait multiplier cet impact par le nombre de mots lus.

L'oculométrie est souvent utilisée pour évaluer la performance de lecture grâce à l'analyse de métriques classiques : le nombre de fixations, la durée de ces fixations, l'amplitude des saccades, etc. (Rayner, 1998; Schmid & Baccino, 2003; Siegenthaler et al., 2012). Cependant au-delà de ces paramètres classiques, d'autres mouvements des yeux peuvent se révéler d'intérêt.

En effet, on sait qu'un roulis de la tête induit un mouvement des yeux appelé la « contre-cycloversion » (Collewijn, Steen, Ferman, & Jansen, 1985). Ce mouvement consiste en une rotation des deux yeux autour leurs axes visuels. La rotation se fait pour chaque œil dans le sens opposé à l'inclinaison de la tête. Ce mécanisme peut ainsi compenser en partie l'inclinaison d'une ligne horizontale sur la rétine¹⁹.

Mais, la cycloversion peut également être induite par une stimulation visuelle²⁰ non horizontale, qu'elle soit fixe (Goodenough, Sigman, Oltman, Rosso, & Mertz, 1979) ou en mouvement de rotation (Ibbotson, Price, Das, Hietanen, & Mustari, 2005). Dans ces situations, la cycloversion suit la direction de la rotation du stimulus. Dans cette étude, nous considérons à la fois l'inclinaison latérale de la tête et de la tablette, les deux composantes devraient donc être impactées. Quand seule la tête est en roulis, la contre-cycloversion devrait

¹⁹ La contre-cycloversion a pour effet de stabiliser en partie l'image rétinienne en contrebalançant le mouvement de la tête. Dans ce cas, le facteur déclenchant la cycloversion est l'information provenant des canaux semi-circulaires de l'oreille interne et on peut parler de la composante « gravitaire » de la cycloversion.

²⁰ Dans ce cas, le facteur déclenchant la cycloversion est la stimulation visuelle et on pourra parler de composante « visuelle » de la cycloversion.

diminuer l'inclinaison des mots sur la rétine, réduisant ainsi l'effet de l'inclinaison sur les performances de lecture. Les mots devraient être moins tournés par rapport à la rétine que par rapport à la tête. De la même manière, si la lecture sur une tablette tournée génère une cycloversion visuelle, elle devrait également compenser l'inclinaison des mots.

Afin de mieux comprendre les phénomènes à l'œuvre lors de la lecture sur support électronique dans ces postures un peu particulières, nous avons mené à bien deux expériences. La première visant à mesurer l'impact de ces dernières sur la performance de lecture et la seconde à approfondir le rôle de la cycloversion lors de cette tâche.

7.1.2 Expérience 1

La première expérience prend la forme d'une approche classique d'oculométrie. Le but est d'évaluer un éventuel effet de la position inclinée de la tête et/ou de la tablette sur la performance de lecture.

7.1.2.1 Méthode

7.1.2.1.1 Participants

28 sujets âgés de 20 à 45 ans (moyenne : 30 ans, écart-type : 7 ans) ont participé à la première expérience. Une limite d'âge de 45 ans a été fixée afin d'éviter d'éventuelles difficultés de lecture liées à la presbytie. L'échantillon était composé de 12 hommes et 16 femmes, habitués à lire sur des supports électroniques sans porter de correction visuelle, hormis des lentilles de contact. Ils devaient être capables de lire et de comprendre des textes en français. Ils ne devaient pas avoir d'antécédents de pathologie visuelle ou musculo-squelettique. Chaque sujet a rempli un formulaire de consentement.

7.1.2.1.2 Appareillage

Afin d'enregistrer les mouvements oculaires, nous avons utilisé un oculomètre Dikablis (Ergoneers GmbH, Manching, Allemagne). Cet appareil est un oculomètre monoculaire (œil gauche), porté sur la tête, avec une fréquence d'acquisition de 25Hz.

Une tablette Latitude ST a été utilisée (Dell, Round Rock, Etats-Unis). Sa luminance a été fixée à 94 cd/m², et le contraste maximal était de 0.010 (Michelson). Ces valeurs ont été mesurées à l'aide d'un LS-110 (Konica Minolta, Tokyo, Japon). Une souris sans fil était combinée à la tablette de manière à ce que les sujets puissent utiliser les deux boutons pour interagir avec la tablette.

Nous avons construit un bureau, une mentonnière avec appui frontal et un support de tablette afin de contrôler la position de la tablette et du corps. Les dimensions de ce montage ont été déterminés à partir des travaux de Young, Trudeau, Odell, Marinelli et Dennerlein (2012), de manière à ce qu'elles coïncident à une posture naturelle. Ces auteurs ont mesuré la posture de 15 sujets qui utilisaient une tablette électronique. Nous avons utilisé leurs données (dans la condition manuelle : « lap-hand ») pour construire le montage (*Figure 19*). Les supports ont été prévus pour assurer un abaissement de tête de 10° et une inclinaison de tablette de 36° . Seule la distance de lecture était différente des données de l'article (55 cm au lieu de 50 cm). Les sujets pouvaient ajuster la hauteur de leur chaise. Les conditions d'éclairages de la salle étaient contrôlées de manière à ce que l'ambiance lumineuse au niveau de la chaise soit à 456 lux, valeur mesurée par un LX100 (Kimo, Montpon Menéstérol, France). Cette valeur d'ambiance lumineuse correspond aux recommandations de l'AFNOR pour le travail de bureau (AFNOR, La Plaine Saint-Denis, France).

L'inclinaison de la tête était contrôlée par une mentonnière ajustable qui permettait une inclinaison de 0° , de 10° dans le sens des aiguilles d'une montre et de 10° dans l'autre sens. La tablette pouvait être tournée de la même manière à l'aide du support ajustable. Le support et la mentonnière ont été réalisés de manière à ce que la distance entre les yeux et les quatre coins de la tablette reste identique lorsque la mentonnière et le support sont tournés du même angle signé.

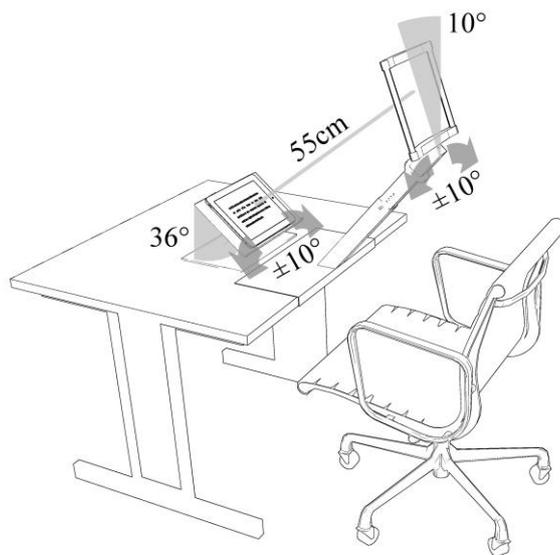


Figure 19: Description schématique du bureau et de la mentonnière utilisés pour contrôler la flexion de la tête et de la tablette

7.1.2.1.3 *Stimuli*

Le script utilisé pour afficher les textes sur la tablette a été écrit sous le langage Python 2.7.4 avec le module Pygame 1.9.1. La résolution de l'écran a été fixée à 1280 * 800. La police utilisée était « Arial » en taille 65 pts, de manière à ce qu'une lettre minuscule « o » tiennent dans un angle visuel de 39 minutes.

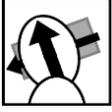
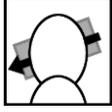
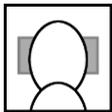
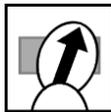
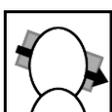
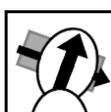
Les textes ont été écrits en français. Ils sont issus de Wikipedia en utilisant la fonction de recherche aléatoire d'articles. Les textes étaient modifiés de manière à être constitués de 30 mots (+/- 3 mots). 5 sujets qui n'ont pas participé à l'étude ont évalué la difficulté de chaque texte sur une échelle de Likert à 5 points. Les textes évalués comme étant « très difficiles » ont été écartés. La fréquence lexicale des mots contenus dans chaque texte a été acquise en se basant sur les données de la base Lexique 3 (New et al., 2001) la moyenne des logarithmes de base 10 de la fréquence lexicale de chaque mot pour 1 million était de 2.593 (écart-type : 1.429). Les textes ont été distribués équitablement dans chaque condition pour que la moyenne des logarithmes des fréquences lexicales et les difficultés moyennes soient similaires pour toutes les conditions. La fréquence lexicale moyenne des conditions était 2.739 (écart-type : 0.028). La difficulté moyenne était 2.01 sur l'échelle à 5 points (écart-type : 0.07). Chaque texte était suivi de deux questions à choix multiple.

7.1.2.1.4 *Plan expérimental et procédure*

La tête et la tablette étaient droites, ou tournées de 10° dans un sens ou dans l'autre. 7 conditions ont été choisies (pour une vue synthétique, voir *Table 5*) : une condition où la tête et la tablette sont droites, deux conditions où les deux sont tournées dans la même direction, deux conditions où seule la tablette est tournée, et deux conditions où seule la tête est tournée. Nous n'avons pas considéré les situations où la tête et la tablette auraient été tournées dans des sens opposés, car ces situations auraient été inconfortables et inhabituelles. L'ordre des conditions a été contrebalancé par un carré latin.

Table 5

Ensemble des conditions expérimentales

		Inclinaison de la tête		
		10° gauche	0°	10° droite
Inclin. de la tablette	10° gauche			X
	0°			
	10° droite	X		

Note. Les conditions où la tête et la tablette sont inclinées dans des directions opposées ne sont pas considérées (marquées d'un "X").

Les sujets étaient familiarisés à la tâche à travers 5 textes. Il leur était demandé de lire le texte afin de pouvoir répondre plus tard à des questions traitant de celui-ci. Quand les sujets avaient fini de lire le texte, ils devaient cliquer sur la souris pour masquer l'affichage. 2 questions leur étaient posées, puis ils pouvaient lire le texte suivant.

Chaque condition comprenait 7 textes. Avant l'apparition de chaque texte, une croix de fixation était affichée à la position de la première lettre du texte. Avant chaque condition, la position de la mentonnière et du support de tablette était réglée. L'inclinaison de la tête était vérifiée à l'aide de lignes qui devaient s'afficher horizontalement sur l'image de la caméra de l'oculomètre. Les sujets étaient encouragés à prendre des pauses dès qu'ils le désiraient.

7.1.2.1.5 Analyse des données

Deux variables étaient directement mesurées durant l'expérience : la durée entre l'apparition du texte et le moment où le sujet clique sur la souris indiquant la fin de la lecture, et le nombre d'erreurs au questionnaire.

Les données oculométriques étaient traitées de la manière suivante :

- Les fixations : les fixations ont été identifiées à partir des données brutes en utilisant l'algorithme de dispersion décrit par Salvucci et Goldberg (2000) : « Identification du Seuil de Dispersion » (I-DT). Dès lors, nous pouvions définir les

autres variables oculométriques. La plupart concernent la première lecture du texte.

- Les fixations de relecture : nous avons fait la distinction entre les fixations et les fixations de relecture. La ligne de texte la plus basse ayant été lue était identifiée (il est rarement arrivé qu'il ne s'agisse pas de la dernière ligne du texte, c'est-à-dire lorsque le sujet n'a pas lu le texte en entier). Les fixations arrivant après la première lecture de cette ligne étaient considérées comme des fixations de relecture.
- Le nombre de fixations en relecture : le nombre de fixations et de régressions était divisé par le nombre total de mots pour supprimer l'effet de longueur du texte.
- La vitesse de lecture : a été calculée en divisant le temps total de fixation par le nombre de mots lus.

Pour chacune de ces variables, nous avons réalisé une analyse de la variance à mesure répétée. Ces ANOVAs ont été ajustées à l'aide de la correction de Greenhouse-Geisser. Nous avons également mesuré la corrélation entre chaque condition. Une analyse de puissance (compromis) a été réalisée avec G * Power 3.1.7 (Faul, Erdfelder, Lang, & Buchner, 2007). En estimant un effet modéré des conditions expérimentales ($f = 0,25$), l'analyse aboutit à une puissance statistique importante ($1 - \beta = 0,972$) avec notre taille d'échantillon.

7.1.2.2 Résultats

La *Table 6* résume les valeurs moyennes pour chaque variable dépendante et les statistiques F associées. Aucun effet de la position de la tête et/ou de la tablette sur la performance de lecture n'a pu être observé. En outre, Edgell (1995) a montré que la meilleure manière de montrer l'absence d'un effet, et d'accepter l'égalité entre les différentes conditions, est de montrer que la probabilité de faire une erreur de type II est suffisamment faible pour être négligée. Nous pouvons donc conclure que les performances ont été les mêmes, quelles que soient les positions de la tête et de la tablette.

De plus, la corrélation moyenne pour la plupart des variable était forte : entre ,731 et ,885 (comme indiqué dans la *Table 6*). Seul le nombre d'erreurs au questionnaire était faiblement corrélé entre les différentes conditions (,222). La faible valeur de ces corrélations peut être expliquée par l'occurrence trop rare d'erreurs aux questionnaires. En effet, les textes

étaient courts et les questions n’adressaient qu’un faible niveau de compréhension : elles portaient sur la mémorisation récente, et non pas sur des processus de plus haut niveau comme l’inférence. Seules quelques erreurs ont été observées, de manière trop rare pour aboutir à une corrélation. Les corrélations des autres variables, qui sont élevées, sont en faveur de l’absence d’effet de l’inclinaison de la tête et de la tablette sur la performance de lecture.

Table 6

Valeurs moyennes (écart-types), ANOVA et corrélations pour les variables oculométriques

	Condition							F(6 ;156)	p (GG)	Corrélation moyenne entre les conditions
Durée avant le clic (s)	13,16 (4,65)	13,10 (3,58)	12,26 (2,94)	13,26 (4,57)	12,37 (3,56)	12,73 (3,89)	12,83 (4,18)	1,69	,177	,823 *
Erreurs par question	0,056 (0,065)	0,048 (0,070)	0,061 (0,067)	0,048 (0,051)	0,056 (0,086)	0,048 (0,067)	0,077 (0,075)	0,78	,543	,222
Durée de fixation (ms)	276,1 (30,5)	273,6 (30,9)	279,3 (33,5)	280,6 (34,8)	279,3 (32,5)	277,1 (31,3)	280,9 (37,3)	1,68	,155	,885 *
Nombre de fixations	1,11 (0,23)	1,11 (0,18)	1,08 (0,17)	1,08 (0,19)	1,06 (0,20)	1,08 (0,23)	1,08 (0,22)	1,17	,330	,790 *
Taille de saccade (°)	3,78 (0,59)	3,79 (0,68)	3,76 (0,64)	3,85 (0,74)	3,80 (0,70)	3,73 (0,71)	3,75 (0,73)	0,56	,704	,853 *
Nombre de régressions	0,17 (0,05)	0,18 (0,05)	0,17 (0,04)	0,17 (0,05)	0,16 (0,05)	0,16 (0,05)	0,16 (0,05)	1,66	,150	,731 *
Vitesse de lecture (mots / seconde)	3,58 (0,89)	3,56 (0,70)	3,56 (0,75)	3,60 (0,88)	3,67 (0,85)	3,65 (0,95)	3,60 (0,87)	0,44	,792	,866 *
Nombre de fixations en relecture	0,37 (0,42)	0,37 (0,29)	0,29 (0,24)	0,40 (0,41)	0,33 (0,29)	0,35 (0,33)	0,34 (0,30)	1,58	,196	,781 *

Note. (*) dans ces cas, toutes les corrélations sont significatives ($p < ,005$)

7.1.2.3 Discussion

Nous avons construit une situation écologique avec une inclinaison assez faible de la tête et de la tablette. Les postures induites n’ont pas impacté la performance de lecture durant l’expérience. La performance est restée stable au vue des données oculométriques, ce qui ne va pas dans le sens des résultats décrits dans les études précédentes portant sur le traitement des mots ou des lettres inclinées (Cohen et al., 2008; Hock & Tromley, 1978). Afficher des textes entiers, comme dans une situation de lecture normale sur tablette, ne semble pas induire

l'effet observé pendant le traitement de mots isolés. Au contraire, les sujets se sont adaptés à l'inclinaison. Ce constat signifie que l'utilisation des tablettes ne devrait pas être impactée par les nouvelles postures induites, même si les lecteurs ne sont pas toujours strictement alignés au texte lu.

L'absence d'effet de l'inclinaison pourrait être due à plusieurs différences en comparaison aux études mentionnées. Dans notre dispositif, les mots ne sont pas présentés isolément. La lecture des mots n'était donc pas segmentée, mais incluse dans le processus global de lecture du texte. Aussi, les différentes inclinaisons mises en œuvre n'étaient pas appliquées aléatoirement : tous les mots d'un même texte étaient inclinés de la même manière. Enfin, l'inclinaison des mots était générée par celle de la tête ou de la tablette et pas au moyen d'un procédé d'affichage.

Ces différences expérimentales peuvent avoir rendu possibles certains mécanismes qui nécessitent la lecture d'un ensemble de mots ou l'inclinaison de la tête ou de la tablette pour être mis en œuvre. Comme mentionné dans l'introduction, notre dispositif pourrait avoir généré la rotation des yeux par le mécanisme de cycloversion. Ce mouvement apparaît quand un sujet incline latéralement sa tête (Collewijn et al., 1985), ou lorsqu'il fait face à une stimulation visuelle inclinée (Goodenough et al., 1979). Si ce mouvement a effectivement été mis en œuvre, il pourrait être une explication possible à l'adaptation que nous avons pu observer dans cette première expérience.

7.1.3 Expérience 2

Comme mentionné précédemment (voir la sous-section *7.1.1 Introduction*), les positions manipulées dans notre étude sont proches de situations connues pour générer un mouvement des yeux appelé cycloversion²¹. Très peu d'études se sont penchées sur la cycloversion pendant la lecture. Dans cette seconde expérience, nous avons étudié la présence de ce mouvement des yeux, qui pourrait être une piste d'explication sur l'adaptation observée à l'aide des mesures oculométriques classiques.

²¹ En effet, le sujet a parfois la tête inclinée ce qui devrait entraîner un mécanisme de cycloversion par sa composante « gravitaire », parfois c'est le texte qui est incliné par rapport au sujet ce qui devrait entraîner une cycloversion par sa composante « visuelle » et parfois la tête et le texte sont inclinés ce qui devrait entraîner une cycloversion par ses deux composantes.

7.1.3.1 Méthode

7.1.3.1.1 *Participants, appareillage et stimulus*

21 sujets qui avaient pris part à la première expérience ont aussi participé à la seconde. L'appareillage est resté le même (voir 7.1.2.1.2 *Appareillage*). Ici, l'oculomètre a uniquement servi à enregistrer l'image de l'œil, et non pas à suivre les déplacements de la pupille. Un seul texte a été utilisé : il avait le même format que les textes de l'expérience 1 (voir 7.1.2.1.3 *Stimuli*), mais a été rédigé de manière à ce qu'une lettre « x » soit située au centre de l'écran.

7.1.3.1.2 *Plan expérimental et procédure*

Au moment de l'équipement de l'oculomètre, nous vérifions que la caméra filmant l'œil ne cache jamais la lettre centrale « x », quelle que soit la position de la tête et de la tablette. La position de l'œil fixant cette lettre était alors enregistrée.

Les conditions utilisées étaient les mêmes que celles de l'expérience 1, dont l'ordre a été déterminé pseudo-aléatoirement. Pour chaque condition, le support de tablette et la mentonnière étaient réglées. Si nécessaire, le sujet devait déplacer sa tête en élévation et latéralement pour que son œil soit à la position enregistrée lorsqu'il fixe la lettre centrale. Le roulis de la tête était contrôlé de la même manière que dans l'expérience précédente : à l'aide de lignes affichées à l'écran. La mesure était répétée 5 fois pour chaque condition.

Chaque mesure était précédée par l'affichage de bruit visuel²² à l'écran. Un point était affiché là où la première lettre allait apparaître (nous n'avons pas utilisé de croix de fixation pour éviter les repères verticaux et horizontaux). Une fois le texte apparu, le sujet devait lire le texte jusqu'à la lettre centrale. Arrivé sur la lettre centrale, il devait ouvrir grand la paupière et fixer le « x » pendant deux secondes. Les images de l'œil étaient enregistrées à ce moment, l'ouverture de paupière servant à acquérir une bonne image de l'iris.

7.1.3.1.3 *Analyse des données*

Comme l'enregistrement était monoculaire, nous n'avons mesuré qu'une « cyclotorsion » : la rotation d'un œil autour de son axe visuel.

²² Une image pour laquelle on attribue à chaque pixel une valeur aléatoire de niveau de gris.

Un script a été construit avec MatLab R2012b et le module Image Processing (MathWorks, Natick, Etats-Unis). Le script était basé sur l'algorithme décrit par Bucher, Heitger, Mast et Bischof (1990). On enregistrait des profils circulaires de l'iris, situés autour du centre de la pupille. Le centre de la pupille était obtenu par une détection de cercle via la transformée de Hough. Les profils contenaient les valeurs de luminosité des cercles de l'iris, c'est-à-dire une représentation numérique des motifs de l'iris. Ces profils étaient enregistrés sous la forme de vecteurs de longueur 360. Un premier groupe de profils était enregistré sur une image de référence issue de la condition où ni la tête ni la tablette ne sont tournées. Pour chaque image enregistrée, les valeurs des vecteurs étaient décalées à différentes distances. Une convolution a été calculée avec les vecteurs de référence pour chaque groupe de vecteurs décalés, et ce afin d'obtenir une valeur d'adéquation. Une fonction polynomiale a été estimée pour décrire les données d'adéquation ainsi obtenues. Comme les vecteurs étaient de taille 360, nous avons considéré que la valeur de décalage qui maximisait la convolution correspondait à la valeur de cyclotorsion en degrés. Des valeurs positives indiquaient une rotation dans le sens des aiguilles d'une montre, du point de vue du sujet.

Pour éviter de traiter des images de l'œil dégradées par des clignements, une paupière pas assez ouverte, ou encore du bruit sur le signal d'enregistrement, nous avons mesuré la différence absolue entre les valeurs de luminosité des profils courants et celles des profils de référence. Nous avons supposé qu'une petite différence correspondait à des profils issus de deux images de bonne qualité, alors qu'une grande différence indiquait une image dégradée. Pour chaque sujet, nous n'avons donc considéré que les profils pour lesquels la différence absolue se situait entre 0 et la moyenne de toutes les différences plus un écart-type.

Une constante a été ajoutée à chaque valeur de cyclotorsion pour chaque sujet de manière à ce que la valeur moyenne de cyclotorsion soit à 0 pour la condition où ni la tablette ni la tête ne sont inclinées.

Un test de Student a été appliqué pour chaque condition (sauf la condition de référence) afin de comparer la valeur des cyclotorsions à 0. Les tailles d'effets ont été calculées avec le d de Cohen.

Afin de distinguer la composante de la cyclotorsion due à l'inclinaison de la tête et celle due à l'inclinaison de la tablette, nous avons vérifié par une ANOVA qu'il n'y avait pas

d'interaction entre l'effet de la tête et celui de la tablette sur la cyclotorsion ; ensuite nous avons construit un modèle de régression multiple de la forme :

$$cyclo_{totale} = b_{tête} * inclin_{tête} + b_{tablette} * inclin_{tablette} + a$$

Avec $cyclo_{total}$ la valeur de cyclotorsion mesurée en degrés ; $inclin_{tête}$ et $inclin_{tablette}$ les valeurs d'inclinaison latérale de la tête et de la tablette en degrés ; $b_{tête}$ et $b_{tablette}$ les facteurs de cyclotorsion dus à la l'inclinaison de la tête et à celle de la tablette ; a l'ordonnée à l'origine en degrés.

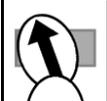
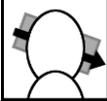
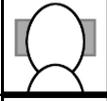
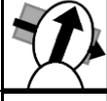
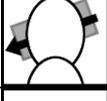
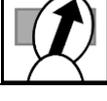
7.1.3.2 Résultats

En raison du rejet des images dégradées, les données de 2 sujets ont été éliminées et pour 4 autres sujets, les valeurs obtenues dans une condition ont du être écartées.

Les tests de Student ont montré que la cyclotorsion était présente dans toutes les conditions où la tête et/ou la tablette sont inclinées (voir *Table 7*).

Table 7

Cyclotorsion moyenne et test de Student

Condition	Cyclotorsion moyenne (sens horaire) [°]	ET	N	IC _{95%}	t	ddl	p	d
	1,99	0,99	18	[1,50 ; 2,48]	8,57	17	< ,001	2,02
	0,98	0,65	19	[0,67 ; 1,29]	6,62	18	< ,001	1,52
	0,76	1,18	18	[0,17 ; 1,34]	2,71	17	,015	0,64
	0	0	19	-	-	-	-	-
	-0,56	1,15	18	[-1,13 ; 0,01]	-2,06	17	,056	-0,48
	-1,18	0,85	18	[-1,60 ; -0,76]	-5,89	17	< ,001	-1,39
	-1,78	0,95	19	[-2,24 ; -1,32]	-8,13	18	< ,001	-1,86

Les cyclotorsions se font dans le sens attendu : dans la direction de l'inclinaison de la tablette et à l'opposé de la direction de l'inclinaison de la tête. Les tailles d'effet sont les plus importantes pour les conditions où les cyclotorsions dues aux inclinaisons de la tête et de la tablette sont dans la même direction. Les effets sont les plus faibles quand les deux modalités sont opposées : c'est-à-dire quand l'observation de la tablette génère une cyclotorsion « visuelle » dans le sens opposé à celui de la cyclotorsion « gravitaire » générée par la tête. Dans ces situations, la cyclotorsion est plus faible, mais elle reste principalement dans la direction de la cyclotorsion « gravitaire » liée à l'inclinaison de la tête (voir *Figure 20*).

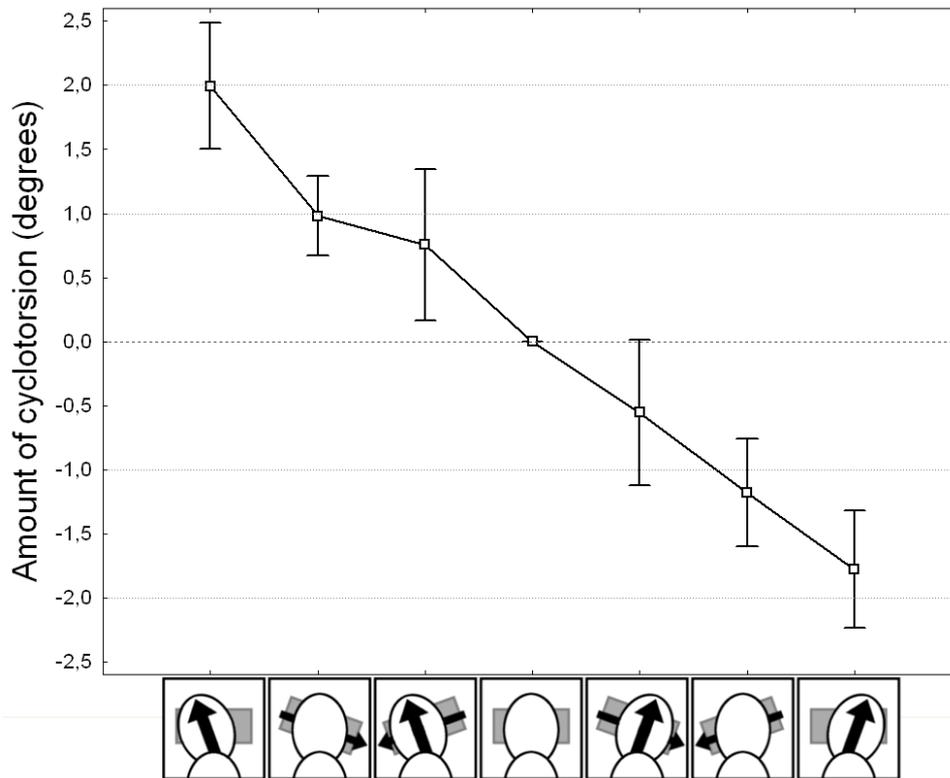


Figure 20. Cyclotorsion moyenne pour chaque condition. Les barres d'erreurs représentent les intervalles de confiance à 95%.

L'ANOVA n'a pas montré d'effet d'interaction entre l'inclinaison de la tête et la celle de la tablette sur la cyclotorsion : $F(2, 122) = 0,225, p = ,799$. Nous avons donc pu procéder à une analyse en régression multiple. Le coefficient de détermination montre que 65% de la variance peut être expliquée par l'équation linéaire (R^2 ajusté = ,649 ; $F(2, 126) = 119,15, p < ,001$). Les coefficients de régression standardisés sont $\beta = -0,355$ pour la tête et $\beta = 0,570$ pour la tablette. Le modèle a été calculé sur les données brutes comme suit :

$$cyclo_{totale} = -0,07 * inclin_{tête} + 0,11 * inclin_{tablette} - 0,03$$

Si nous acceptons ce modèle comme étant une description correcte de la quantité de cyclotorsion générée par la position de la tête et de la tablette, cela implique que les deux modalités sont cumulatives.

Dans ce modèle, les coefficients attribués aux facteurs sont différents de 0 : pour l'inclinaison de tête, $t(126) = 5,854, p < ,0001$ et pour la tablette, $t(126) = 9,395, p < ,001$. Seule l'ordonnée à l'origine ne peut pas être considérée comme étant différente de 0 : $t(126) =$

0,373, $p = ,710$. Cela implique que le modèle est proche d'une combinaison linéaire et que deux conditions symétriques génèrent une quantité opposée de cyclotorsion.

7.1.3.3 Discussion

La deuxième expérience montre que la cyclotorsion peut survenir pendant la lecture sous certaines conditions. Ce constat est intéressant du fait que la littérature traitant de la lecture et des mouvements de torsion de l'œil est très peu fournie. Lam, Chung, Kho et Wong (Lam, Chung, Kho, & Wong, 2000) ont étudié l'effet de ce mouvement sur l'identification de lettres majuscules ; ils ont mesuré la torsion de l'œil due à une inclinaison de 90° de la tête à droite ou à gauche. Comme cette mesure n'était pas effectuée durant la tâche d'identification, on ne peut pas être sûr que la torsion était la même lorsque les sujets étaient simplement assis ou couchés et lorsqu'ils étaient effectivement en train de lire. Notre étude montre que la cyclotorsion a effectivement lieu durant la lecture sur un écran incliné, même si la tête ne l'est pas. C'est un progrès pour la connaissance des mouvements des yeux en lecture, la plupart des études du domaine ne s'intéressant qu'aux fixations et aux saccades (Rayner, 1998).

En proposant un modèle qui explique ce mouvement à travers un double mécanisme, nous sommes en accord avec les recherches plus anciennes. Ce mouvement des yeux peut être induit par la non-verticalité de la tête (Collewijn et al., 1985), c'est la contre-cycloversion « gravitaire ».

La cycloversion peut également être induite par une stimulation visuelle inclinée latéralement (Goodenough et al., 1979). Dans notre expérience, la présence de ces deux mécanismes a été montrée : l'œil se met en rotation dans le sens opposé à l'inclinaison de la tête, et dans le sens d'inclinaison de la tablette.

En plus de montrer les deux types de réponses en cyclotorsion, nous avons avancé le fait que ces mécanismes sont cumulatifs. La quantité totale de cyclotorsion est la somme des deux réponses (posturale et visuelle), que ces deux mécanismes engendrent des torsions qui vont dans le même sens, ou que le sens des torsions soit opposé. En effet, la torsion est à son maximum lorsque la tête est inclinée mais pas la tablette : dans ce cas la contre-cycloversion génère une rotation qui est accrue par la torsion visuelle qui permet de diminuer la non-horizontalité du texte affiché sur la tablette. Inversement, la torsion est la moins importante quand la tête et la tablette sont inclinées : c'est-à-dire quand la rotation due à la contre-cyclotorsion est opposée au sens d'inclinaison de la tablette.

Dans une expérience de Chandrakumar et al. (2010), la cyclotorsion a été mesurée pour plusieurs angles d'inclinaison latérale de la tête. Les sujets devaient observer une grille qui était soit toujours à l'horizontale, soit inclinée de la même manière que la tête. Tout comme dans notre expérience, ils ont distingué la condition où l'inclinaison de la tête est accompagnée de celle de la stimulation visuelle de la condition où seule la tête est inclinée. Le même phénomène a été observé : la cyclotorsion était la plus importante lorsque la tête était inclinée et que la stimulation visuelle était inclinée (par rapport à la tête). Quand la stimulation visuelle n'était pas inclinée par rapport à la tête, la torsion était plus faible. Les valeurs rapportées dans leur étude sont proches de celles issues de notre expérience, pour les conditions où la tête et/ou la stimulation visuelle sont inclinées de 10°.

Les principales différences entre notre expérience et celle de Chandrakumar et al. (2010) sont la stimulation visuelle ainsi que la tâche. Ces auteurs demandaient à leurs participants de fixer une croix centrée sur une grille qui couvrait une grande partie de leur champ visuel. Notre stimulation - le texte affiché sur la tablette - était plus fragmentée et plus bruitée que leur grille qui contenait des lignes bien structurées dans seulement deux directions (la verticale et l'horizontale). Les recherches portant sur la cycloverision visuelle utilisent traditionnellement ce type de stimulation très structurée : des carrés (Goodenough et al., 1979) ou des hélices stylisées (Ibbotson et al., 2005). Néanmoins, dans notre expérience, la lecture du texte sur la tablette était suffisante pour générer la cyclotorsion.

7.1.4 Discussion générale et conclusion

Dans notre première expérience, l'inclinaison latérale de la tête ou de la tablette n'ont eu aucun effet sur la performance de lecture. Les nouvelles postures liées à l'utilisation des tablettes n'ont donc pas les conséquences négatives que nous avons envisagées en nous basant sur les études du traitement de mots isolés (Cohen et al., 2008; Hock & Tromley, 1978). Avec la seconde expérience, nous avons essayé de proposer une explication possible à l'adaptation que nous avons observée. La présence de réponses en cyclotorsion à l'inclinaison de la tête et/ou de la tablette a été montrée. Puisque l'œil tourne en direction de l'inclinaison de la tablette, cela permet une projection plus horizontale des mots sur la rétine. C'est un constat original car la présence de cyclotorsion n'avait pas été montré durant la lecture jusque là.

Les expériences que nous avons conduites font partie de la recherche sur la posture durant les interactions homme machine. Les travaux précédents ont surtout comparé les

différences de lecture dues aux aspects visuels entre plusieurs supports (Benedetto et al., 2013; Menozzi, Lang, Näpflin, Zeller, & Krueger, 2001). Mais les aspects visuels ne sont pas les seuls à évoluer entre les anciens et les nouveaux supports. Le développement des nouvelles technologies sur le plan de la mobilité et des possibilités d'interaction change la posture de l'utilisateur. Certains chercheurs ont étudié ces changements à des fins ergonomiques : ils ont observé l'activité musculaire du cou, des épaules, des avant-bras et des poignets durant l'interaction avec les nouveaux supports (Young et al., 2012; Young et al., 2013). Les préférences et performances des utilisateurs ont également été adressées relativement à une inclinaison frontale (Albin & McLoone, 2014). L'inclinaison frontale n'est pas équivalente à l'inclinaison latérale que nous avons étudiée, car le texte reste horizontal sur la rétine, quel que soit l'angle d'inclinaison. Cependant les auteurs n'ont pas observé d'effet de l'inclinaison sur les performances de lecture. Cela va dans le sens de notre affirmation d'une adaptation aux nouvelles postures engendrées par les tablettes. D'un point de vue plus général, le comportement postural lié à l'utilisation des écrans n'est pas limité au port de tête. Par exemple, Le Bigot et Grosjean (2012) ont étudié l'effet de la position des mains relativement à l'écran sur la sensibilité visuelle. Ils ont rapporté une amélioration des performances visuelles lorsque les mains sont proches de l'écran. Ce type d'effet renforce le besoin de prendre en compte les aspects posturaux durant l'évaluation de la lecture sur support électronique.

Nous avons montré une adaptation globale des lecteurs dans notre étude, on pourrait tout de même soulever l'hypothèse que cette capacité d'adaptation dépend du style perceptif du lecteur. Notre dispositif expérimental était en effet proche de ceux utilisés pour la recherche sur les styles perceptifs : le test classique de « Rod and Frame » (Witkin & Asch, 1948) utilise un cadre large qui pourrait être comparé à notre tablette. Dans ce test, le sujet doit ajuster une baguette inclinée à la verticale gravitaire. Le cadre est incliné latéralement, de manière à ce que le sujet, en fonction de son style perceptuel, parvienne à aligner avec plus ou moins d'erreur la baguette à la verticale. Les sujets ayant plus recours à l'information visuelle (dépendants au champ visuel) ont tendance à faire des erreurs plus grandes, alors que ceux qui exploitent l'information gravitationnelle (indépendants au champ visuel) font des erreurs bien plus faibles (Isableu, Ohlmann, Crémieux, & Amblard, 1997; Witkin, 1949). On pourrait penser que les sujets qui sont moins affectés par l'information visuelle, et qui font des erreurs plus faibles, ont une capacité d'adaptation plus importante à une stimulation visuelle inclinée. En faisant le parallèle à notre étude, cela pourrait impliquer que les sujets indépendants au

champ visuel seraient les plus à même de s'adapter à l'inclinaison de la tablette. Nous avons fait passer à nos sujets un test de Rod and Frame avant l'expérience, mais nous n'avons trouvé que très peu de variabilité interindividuelle dans les erreurs (l'erreur moyenne due à l'inclinaison du cadre s'élevait à $0,96^\circ$, avec un écart-type de 0,85). Les sujets ont donc été capables de placer la baguette à la verticale avec une erreur très faible, nous pouvons donc les considérer comme étant indépendants au champ visuel. Cela est probablement dû au fait que les sujets étaient assez jeunes, et les sujets jeunes sont en général moins susceptibles de faire des erreurs à ce test en comparaison aux personnes plus âgées (Panek, Barrett, Sterns, & Alexander, 1978). Il y a également peu de variabilité interindividuelle dans cette partie de la population (Poulain, Giraudet, & Dobrescu, 2004). Notre étude ne peut donc pas répondre à la question du lien entre style perceptif et capacité d'adaptation à l'inclinaison de la tablette. Il serait intéressant de faire participer des sujets plus dépendants au champ visuel, et de vérifier s'ils pourraient s'adapter de la même manière que les sujets ayant participé à l'étude.

Nous avons montré la présence de cyclotorsion durant la lecture d'un texte incliné latéralement. Mais cette cyclotorsion reste relativement modeste par rapport à l'inclinaison de la tête et/ou de la tablette. Pour 10° d'inclinaison, la réponse en cyclotorsion est de l'ordre de 1 à 2° . Ce mouvement de l'œil compense seulement en partie l'inclinaison de la tête et de la tablette. Si la cyclotorsion peut expliquer en partie l'adaptation, d'autres mécanismes de compensation doivent être à l'œuvre. Une compensation cognitive est sans doute présente aussi.

Notre étude a montré une adaptation du lecteur à une situation d'inclinaison de tête et/ou de tablette, dans un contexte proche des conditions habituelles de lecture. Les positions induites par l'utilisation de nouveaux supports n'impactent pas la performance de lecture. Cette adaptation peut être partiellement expliquée par un mécanisme physiologique : la cycloversion. Les nouveaux supports font évoluer la manière dont nous devons étudier la lecture électronique, de par l'introduction de nouvelles dimensions en comparaison à celles étudiées classiquement. Dans les recherches à venir, les aspects posturaux devraient être pris en compte au même titre que les aspects cognitifs dans la recherche sur les interactions homme-machine.

7.2 L'effet de la lecture électronique sur la posture

7.2.1 Introduction

Les dispositifs électroniques de petite taille sont de plus en plus présents dans notre environnement. Les tablettes, liseuses et smartphones nous accompagnent où que nous allions, grâce à leur format. Le temps où l'outil informatique était uniquement utilisable depuis un bureau est révolu.

Ces nouveaux dispositifs ont rapidement fait l'objet d'investigation afin de mesurer l'impact qu'ils pouvaient avoir sur les performances de lecture (Benedetto et al., 2013; Siegenthaler et al., 2012). Mais les effets possibles de ces technologies ne doivent pas être uniquement considérés sous les aspects visuels et cognitifs. Ces nouveaux formats peuvent engendrer de nouvelles réponses posturales, comme l'attestent les études ergonomiques qui visent à déterminer les risques éventuels qui y sont associés (Straker et al., 2008; Young et al., 2012). Un des paramètres posturaux les plus impactés par l'utilisation des écrans de petite taille est la distance de lecture. Plus l'écran est petit, plus les sujets s'en rapprochent (Maniwa et al., 2013). Cette distance peut également être modulée par la qualité visuelle de l'affichage (Ko, Mohapatra, Bailey, Sheedy, & Rempel, 2014) ou par la durée de la lecture (Maniwa et al., 2013). La distance de lecture est un paramètre important à prendre en compte dans l'étude de la lecture électronique dans le sens où celle-ci contraint l'accommodation ainsi que la convergence, et peut affecter la vision, et/ou générer de la fatigue visuelle (Jaschinski, 2002) ; il existe une importante variabilité interindividuelle dans la distance à l'écran, qui pourrait expliquer le fait que la fatigue visuelle n'affecte pas de la même manière tous les lecteurs (Rempel et al., 2007).

Dans notre étude, ce ne sont pas ces aspects d'ergonomie physique ou de vision que nous allons présenter, mais nous nous proposons de questionner le lien entre la distance de lecture et une capacité cognitive : l'attention. Il est utile de tenir compte des aspects cognitifs pour prendre en compte la distance d'utilisation. En effet les aspects visuels et ergonomiques ne sont pas les seuls à avoir une influence sur les paramètres posturaux : la distance à l'écran peut par exemple varier en fonction de la tâche demandée au sujet (Bababekova, Rosenfield, Hue, & Huang, 2011). Dans cette étude, les auteurs ont montré que la distance de lecture est plus courte pour une consultation de site web que pour la lecture d'un message. Une piste d'explication peut être explorée par la prise en compte du niveau attentionnel. Des auteurs ont pu montrer que la difficulté d'une tâche ou les ressources attentionnelles impliquées peuvent

faire diminuer la distance à l'écran (Balaban et al., 2004; Qiu & Helbig, 2012). Dans une tâche de simulation de défense aérienne, Balaban et al. (2004) ont montré que la distance est plus courte lorsque le nombre de problèmes à résoudre augmente. C'est donc un lien entre la complexité d'une tâche et la posture qui a été mis en lumière par ces travaux. Qiu et Helbig (2012) ont réalisé le même type d'expérience, en demandant à leurs sujets de passer des tests psychologiques plus ou moins complexes. En plus de mesurer la distance à l'écran, les auteurs ont effectué des mesures subjectives de la charge attentionnelle des sujets. Les tâches associées à des valeurs élevées de charge attentionnelle sont celles où les sujets sont au plus proche de l'écran. Si l'on postule une relation de causalité entre la charge attentionnelle et la distance à l'écran, l'effet de la complexité sur la distance peut être remis en question. En effet, la charge attentionnelle n'est pas strictement liée à la complexité de la tâche : bien que la charge augmente avec la complexité jusqu'à un certain niveau, une complexité trop importante peut entraîner un désengagement des ressources attentionnelles du sujet ; une complexité trop importante peut être accompagnée d'une faible mobilisation des ressources attentionnelles (Geissler, Zinkhan, & Watson, 2006).

Les études de Balaban et al. (2004) et de Qiu et Helbig (2012) ne portent pas sur des tâches de lecture électronique. Dans cette étude, nous chercherons à voir si le lien entre le niveau d'attention et la distance à l'écran est confirmé pour une tâche de lecture. Cette confirmation n'est pas triviale, dans la mesure où la lecture électronique est un phénomène complexe, impliquant des phénomènes de bas niveau tels que la détection de contrastes et des phénomènes de haut niveau tels que la mémorisation et la compréhension (Baccino, 2004). En outre, la difficulté d'un texte est moins aisément contrôlée que celle d'une tâche artificielle. L'efficacité de la lecture peut varier selon les connaissances préalables du lecteur (Kendeou & van den Broek, 2007) : un lecteur ayant de solides connaissances sur un domaine traité par un texte le trouvera moins complexe qu'un lecteur ne disposant pas de ces connaissances. Il a en outre été montré que l'emploi de textes faciles et difficiles dans une tâche de lecture électronique peut affecter le niveau d'attention (Schultheis & Jameson, 2004). Les études montrant un lien entre complexité et distance impliquaient l'utilisation de moyens d'interaction (clavier, souris,...). Ces interactions mobilisent la posture, et pourraient également être l'une des causes du rapprochement à l'écran. La réduction de la distance à l'écran est peut-être un moyen de mieux préparer les réponses motrices liées à ces moyens d'interaction. Notre étude cherchera donc aussi à vérifier si une tâche qui n'implique pas d'interactions physiques, une lecture simple, peut suffire à générer le rapprochement à l'écran.

Notre principale hypothèse de recherche qui guidera cette étude est la suivante :

- (1) Une augmentation de la charge attentionnelle pendant la lecture devrait diminuer la distance entre les yeux et l'écran.

Cette hypothèse est accompagnée de deux hypothèses secondaires :

- (2) L'utilisation de textes faciles et difficiles devrait faire varier le niveau d'attention, mais pas de manière systématique puisque le niveau attentionnel n'est pas strictement lié à la complexité.
- (3) Une variation interindividuelle dans les distances à l'écran devrait être observée.

Si l'hypothèse (1) est vérifiée, la mesure de la distance pendant la lecture pourrait être utilisée comme indicateur de charge attentionnelle. La mesure de la distance de lecture peut être réalisée assez facilement par exemple avec un télémètre ou avec une caméra attachée à l'écran et un logiciel de traitement d'image (*face tracking*). Les chercheurs étudiant la charge attentionnelle pourraient donc facilement utiliser ce nouvel indicateur.

Dans une approche plus théorique, la vérification de cette hypothèse apporterait également des données nouvelles sur les mécanismes liés à la variation du niveau attentionnel.

Afin de vérifier ces hypothèses, nous avons réalisé une expérimentation où les sujets devaient lire des textes faciles et des textes difficiles, tandis que nous mesurons leur charge attentionnelle ainsi que leur posture.

7.2.2 Méthode

7.2.2.1 Participants

Au total, 23 participants ont pris part à l'expérimentation. Chacun d'entre eux avait pour langue maternelle le français. Ce groupe était composé de 13 femmes et 10 hommes ayant entre 18 et 45 ans. Leur âge moyen était de 30,0 ans (écart-type : 6,7). La limite d'âge à 45 ans a été choisie pour éviter de faire participer des sujets presbytes : la difficulté d'accommodation aurait pu contraindre leur posture. Ces sujets devaient ne pas avoir d'antécédents de problèmes de vision. Leur acuité visuelle binoculaire minimale en vision de près était fixée à 10/10^{ème}, valeur qui était vérifiée à 33 cm à l'aide d'un Optoprox (Essilor

International, Créteil, France). Les sujets pouvaient porter leur correction en lentille si nécessaire.

Chaque participant a signé un formulaire de consentement lui expliquant que l'étude portait sur la posture pendant la lecture sur support électronique. L'hypothèse principale n'était communiquée qu'à la fin de l'expérimentation.

7.2.2.2 Stimuli

Nous avons rédigé 10 textes narratifs de 698 caractères en moyenne (écart-type : 4,74). Ces textes étaient suffisamment courts pour être affichés sur une seule page. Ainsi les lecteurs n'ont pas eu à interagir avec le support pour pouvoir lire l'intégralité du texte.

Afin de pouvoir moduler la charge attentionnelle liée à la lecture, deux conditions ont été mises en place. Dans une condition « facile », les textes étaient présentés normalement. Dans une condition « difficile », nous avons repris les textes, en enlevant tout signe de ponctuation, et en enlevant toutes les majuscules, sauf celle de la première lettre du texte et celles des noms propres. Nous avons fait ce choix en nous basant sur les travaux de Carver (Carver, 1970), qui avait montré que la lecture de textes sans ponctuation était plus difficile que celle de textes normaux. Les textes étaient affichés avec la police « Verdana » à la taille 18 et étaient justifiés (un exemple de textes est donné dans la *Figure 21*).

<p>Depuis quelques heures, il pleuvait à verse. De l'eau s'étendait sur toute la route. Vers l'avant de la voiture, John entendit un bruit sec. Il stoppa net, ouvrit le capot et regarda. Le moteur sifflait. Un air chaud sortait de l'engin, qui commençait à brûler. Une feuille de papier était tout ce qu'il avait trouvé pour se protéger les doigts. Il tenta d'enlever le bouchon de refroidissement, qui glissa. Sur son front, John sentait les gouttes de pluies qui tapaient. Un peu de liquide de refroidissement versé dans le tuyau, et John eu vite fait de retrouver le bouchon sous la voiture et de le remettre en place. Dans la voiture, il s'assit et tenta de redémarrer. Il avait vaincu ! Le moteur fonctionnait à nouveau.</p>	<p>Depuis quelques heures il pleuvait à verse de l'eau s'étendait sur toute la route vers l'avant de la voiture John entendit un bruit sec il stoppa net ouvrit le capot et regarda le moteur sifflait un air chaud sortait de l'engin qui commençait à brûler une feuille de papier était tout ce qu'il avait trouvé pour se protéger les doigts il tenta d'enlever le bouchon de refroidissement qui glissa sur son front John sentait les gouttes de pluies qui tapaient un peu de liquide de refroidissement versé dans le tuyau et John eu vite fait de retrouver le bouchon sous la voiture et de le remettre en place dans la voiture il s'assit et tenta de redémarrer il avait vaincu le moteur fonctionnait à nouveau</p>
---	---

Figure 21. Exemples de textes affichés. Le texte de gauche correspond à la condition facile (avec ponctuation), le texte de droite correspond à la condition difficile (sans ponctuation).

7.2.2.3 Appareillage

Les participants étaient assis sur un tabouret, qui ne comportait ni de dossier ni d'accoudoirs. Nous avons fait ce choix pour que les mouvements des sujets soient contraints au minimum. Dans le même but, nous n'avons pas fourni de table, les sujets devaient tenir le support électronique directement à la main.

Le support utilisé était une liseuse électronique Kobo Glo (Kobo, Toronto, Canada). Il s'agit d'une liseuse dont la diagonale de l'écran mesure 6 pouces, et dont le poids est de 185g. La résolution de l'écran est de 768 * 1024 pixels. Chaque pixel peut prendre 16 niveaux de gris différents. Nous avons opté pour un support relativement léger, dans le but encore une fois d'encourager la mobilité des sujets. La liseuse possède un système de rétro-éclairage, mais nous ne l'avons pas utilisé.

Les sujets étaient équipés d'un *eyetracker* portatif SMI Glasses 1 dont la fréquence d'enregistrement est de 30 Hz.

La posture était enregistrée à l'aide d'un système de *motion capture* (Vicon, Oxford, Royaume-Uni). Le système est composé de 8 caméras « Bonita B10 ». Les données des caméras sont traitées par le logiciel Nexus 1.8.5. La fréquence d'enregistrement a été fixée à 30 Hz. Les sujets étaient équipés d'un casque sur lequel étaient attachés 4 marqueurs réfléchissants. La position des centres de rotation des yeux était définie par rapport aux marqueurs du casque à l'aide d'une méthode géométrique (Patent No. EP2134249 B1, 2013). Un bras en plexiglas avec 6 ramifications contenant aux extrémités de chacune un marqueur réfléchissant était fixé à la liseuse. La position des coins de la liseuse était mesurée dans un repère défini à partir des positions des marqueurs du bras.

7.2.2.4 Mesures

7.2.2.4.1 Indicateurs de la charge attentionnelle

La charge attentionnelle a été observée à l'aide de 4 types de mesures : des indicateurs de performance sur une double-tâche, des questionnaires de compréhension, des évaluations subjectives et des données oculométriques.

- Double-tâche : le sujet devait effectuer une seconde tâche en même temps qu'il devait lire les textes. La seconde tâche consistait en une identification de son. Un ordinateur équipé d'enceintes émettait des sons durant la lecture. Toutes les 3 secondes, à partir du début de la lecture, l'ordinateur avait une chance sur deux de prononcer un nom de lettre de l'alphabet. Quand un son était émis, il y avait une chance sur deux pour que la lettre citée soit une voyelle. Lorsqu'une voyelle était prononcée, le sujet devait répondre « OK ». Pour une consonne, le sujet ne devait rien dire. Les réponses verbales du sujet étaient enregistrées à l'aide du microphone de l'*eyetracker*. Le début de chaque prononciation du mot « OK » a été consigné sur la piste sonore. Cette tâche a alors permis de dégager trois indicateurs : le taux de bonnes réponses (nombre de « OK » suivant une voyelle divisé par le nombre de voyelles citées), le nombre de fausses alarmes (nombre de « OK » prononcés après une consonne) et la durée moyenne de réponse en cas de bonne réponse.
- Compréhension : chaque texte est suivi de 3 questions ouvertes visant à évaluer la mémorisation du texte. Le nombre de réponses correctes était mesuré.

- Jugements subjectifs : après chaque texte, le lecteur devait attribuer deux notes sur une échelle de 1 à 10 : l'une correspond à la difficulté du texte, indépendamment de la seconde tâche, et l'autre correspond au niveau de concentration que le participant a dû mettre en œuvre pour pouvoir réussir à réaliser les deux tâches (nous ferons référence à ces deux notes comme étant « la difficulté subjective » et « l'attention subjective »).
- Variables oculométriques : les données liées aux mouvements des yeux ont été calculées à l'aide du logiciel SMI BeGaze 3.2. Pour chaque texte, nous calculons la durée moyenne des fixations, le nombre de fixations ainsi que la taille moyenne des saccades (en degrés d'angle visuel). La taille des saccades sera convertie par la suite en millimètres à l'écran à l'aide des données de distance.

7.2.2.4.2 *Indicateur postural*

Afin de calculer une distance de lecture, nous avons repris les coordonnées dans l'espace des centres de rotation des deux yeux, ainsi que celles des quatre coins de la liseuse. A l'aide d'un calcul géométrique, la distance séparant la moyenne des deux centres de rotation (« œil cyclope ») de la moyenne des quatre coins de la tablette (centre de l'écran) a été calculée. Cette valeur est ainsi obtenue pour chaque échantillon d'enregistrement. Afin d'obtenir des indicateurs correspondant à chaque texte, nous avons synthétisé ces données en calculant la moyenne et l'écart-type des valeurs obtenues pour chaque texte.

7.2.2.5 **Plan expérimental et procédure**

Chaque sujet a lu un total de dix textes, dont cinq avec ponctuation (condition facile) et les cinq autres sans ponctuation (condition difficile). L'ordre des textes affichés, ainsi que l'ordre des conditions étaient déterminés pseudo-aléatoirement.

La phase de lecture des textes était précédée par l'équipement de l'*eyetracker* et du casque. Une série de photographies était réalisée pour déterminer la position des centres de rotation des yeux. L'*eyetracker* était ensuite calibré. Les appareils de mesures étaient synchronisés : un clap de cinéma équipé de marqueurs réfléchissants était actionné, de manière à ce que la caméra de scène de l'*eyetracker* le filme et qu'il soit enregistré par le système de *motion capture* ; un son était joué par l'ordinateur dédié à la double tâche de manière à ce qu'il soit enregistré par le microphone de l'*eyetracker*.

Un premier texte était présenté au sujet de manière à le familiariser avec le format de l'expérimentation. Une phase de familiarisation à la seconde tâche était également effectuée de manière à expliquer et à vérifier la compréhension de la consigne avec le sujet.

Chaque texte était précédé d'une croix de fixation placée à la place de la première lettre du texte. Pour chaque texte, le sujet devait appuyer sur l'écran pour démarrer l'affichage. Une fois le texte lu, le participant devait dire « stop », et appuyer sur l'écran pour masquer le texte.

7.2.3 Analyse des données et résultats

7.2.3.1 Données manquantes

A cause de problèmes d'enregistrements, les données complètes d'un sujet n'ont pas pu être enregistrées. Pour ce sujet, seuls les 4 premiers essais sur les 10 ont été obtenus. Pour les 22 autres sujets, nous avons obtenu les valeurs pour chacun des 10 essais. Ainsi, un total de 224 mesures a été réalisé. Les données manquantes expliquent le degré de liberté non entier donné dans les analyses suivantes. Les données de distances moyennes à l'écran pour chaque texte et pour chaque sujet sont représentées dans la *Figure 22*. La moyenne globale des distances est de 323,5 mm et l'écart-type correspondant est de 76,5 mm. On peut donc constater une importante variabilité dans les distances à l'écran.

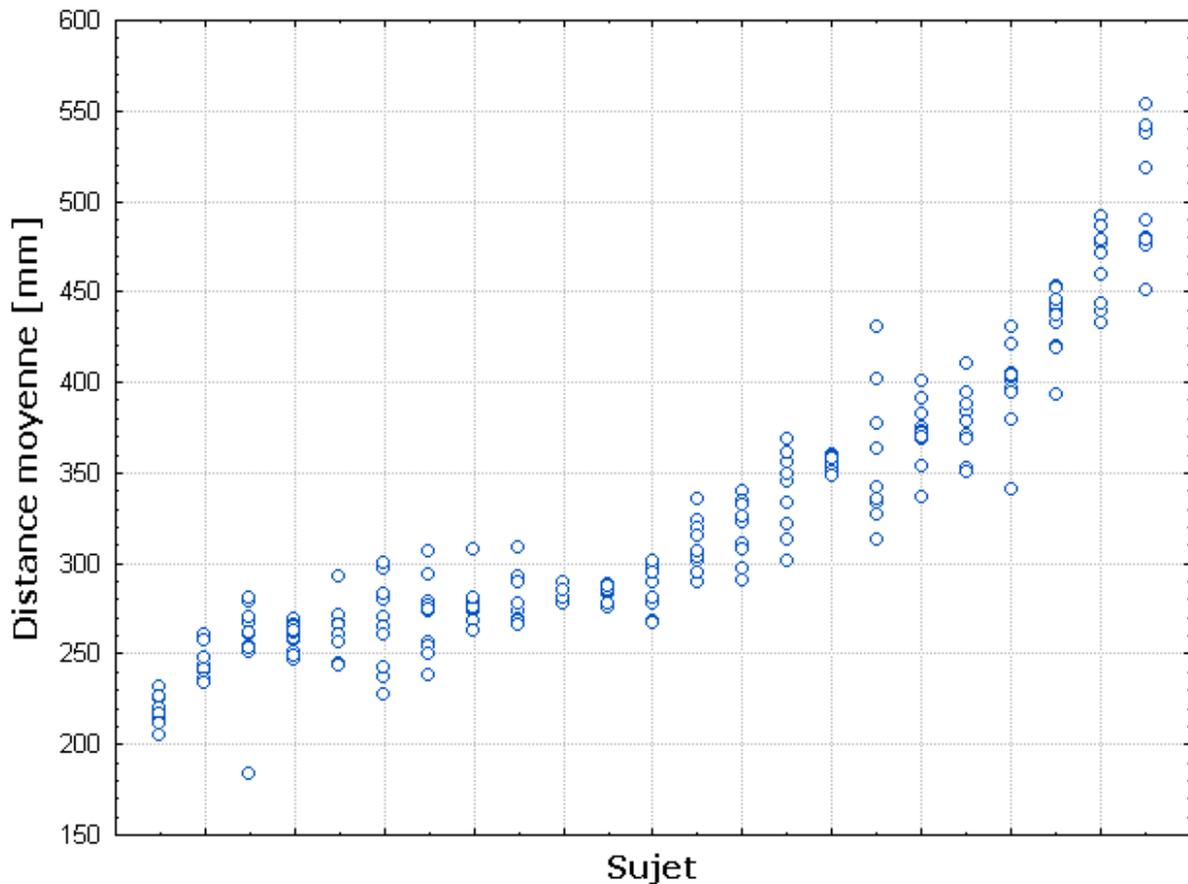


Figure 22. Distances à l'écran moyennes pour chaque sujet et pour chaque texte. Les sujets sont classés dans l'ordre croissant de leur moyenne globale.

7.2.3.2 Effet de la difficulté

Une première analyse permet de montrer l'effet de la difficulté du texte (présence ou absence de ponctuation) sur la distance à l'écran. Nous utilisons un modèle linéaire général pour réaliser une ANOVA à mesures répétées. Ce modèle cherche à expliquer la distance moyenne et l'écart-type de la distance par deux prédicteurs catégoriels : le sujet (facteur aléatoire) et la ponctuation.

L'analyse nous permet de valider la présence d'un effet de la difficulté sur la distance moyenne à l'écran. La distance de lecture est plus faible lorsque les sujets lisent un texte qui ne contient pas de ponctuation (et donc plus difficilement lisible) par rapport à la lecture d'un texte affiché normalement : $F(1 ; 23,846)^{23} = 7,34, p = ,012, \eta^2_p = ,24$. Bien que la différence

²³ Le degré de liberté à virgule est dû à la prise en compte des résultats du sujet pour lequel les données n'ont pu être obtenues que pour 4 textes sur 10.

de distance moyenne soit assez faible, elle est néanmoins significative : les sujets sont en moyenne plus proches de 6,70 mm pour les textes difficiles que pour les textes faciles.

L'analyse montre aussi un effet de la difficulté sur l'écart-type de la distance à l'écran. La variabilité de la distance est plus importante pour un texte difficile que pour un texte facile : $F(1 ; 23,894) = 13,33, p = ,001, \eta^2_p = ,36$. L'écart-type est en moyenne plus grand de 1,60 mm pour un texte difficile.

7.2.3.3 Liens avec la charge attentionnelle

Un second modèle a été construit pour pouvoir décrire les variations de distances moyennes à partir des variables servant à mesurer la charge attentionnelle. On cherche cette fois à expliquer la distance moyenne à l'écran par le facteur sujet (facteur aléatoire) et par l'ensemble des variables associées à la charge attentionnelles (9 facteurs continus, décrits dans 7.2.2.4.1 *Indicateurs de la charge attentionnelle*). Le modèle ainsi généré explique très bien les distances œil/écran : ce modèle aboutit à une valeur de R^2 ajusté = 0,946, $F(31 ; 192) = 126,53, p < ,001$. Le facteur qui explique le plus de variance est le facteur sujet : $F(1 ; 192) = 105,14, p < ,001, \eta^2_p = ,92$. Parmi les 9 variables liées à la charge attentionnelle, toutes ne sont pas significativement liées à la distance à l'écran. Les quatre variables significativement liées à la distance sont les suivantes :

- Nombre de bonnes réponses au questionnaire : $F(1 ; 192) = 14,08, p < ,001, \eta^2_p = ,07$;
- Durée moyenne de fixation : $F(1 ; 192) = 7,10, p = ,008, \eta^2_p = ,04$;
- Attention subjective : $F(1 ; 192) = 6,62, p = ,011, \eta^2_p = ,03$;
- Taux de bonnes réponses à la seconde tâche : $F(1 ; 192) = 5,59, p = ,019, \eta^2_p = ,03$.

Les autres variables utilisées pour mesurer la charge attentionnelle ne sont pas significativement liées à la distance à l'écran : pour chacune d'elles, $p > ,100$.

La distribution des résidus du modèle est représentée dans la *Figure 23*. Cette distribution est proche d'une distribution normale de moyenne 0 et ayant pour écart-type 16,52.

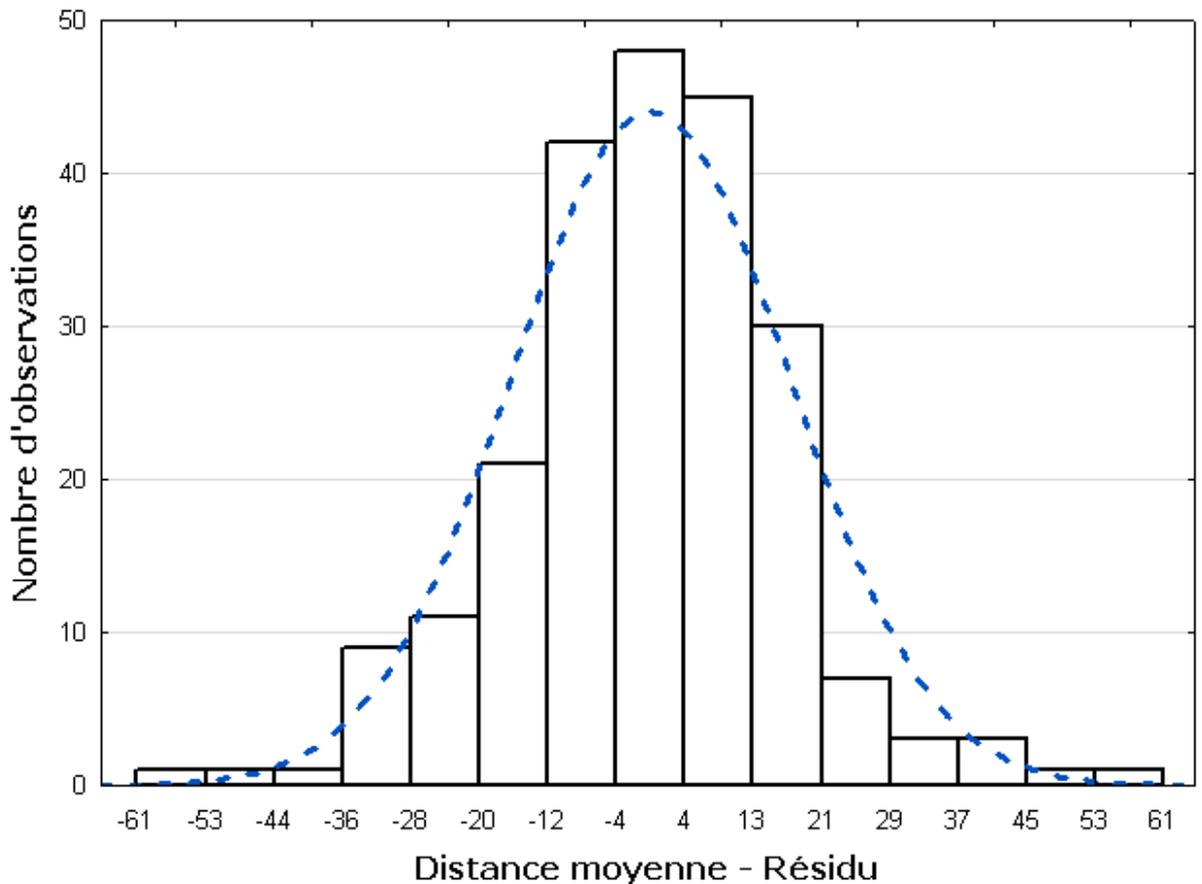


Figure 23. Histogramme des résidus des distances moyennes du modèle (valeurs prévues – valeurs observées). La courbe en pointillés bleus représente la distribution normale qui correspond à la moyenne et à l'écart-type des données obtenues.

Le modèle linéaire peut se représenter par une équation contenant une valeur constante, associant un coefficient à chaque variable continue et une constante à chaque modalité de la variable nominale. Les valeurs brutes des coefficients associés aux variables attentionnelles ne sont pas informatives en soi, car leur impact sur la distance n'a de sens qu'au regard de l'échelle de grandeur de la mesure. Quelques exemples de la variation de distance associée à la variation d'une variable attentionnelle donnée sont présentés pour les quatre variables significatives dans la *Table 8*.

Table 8

Exemples de variations de distance associées à des variations des variables attentionnelles

	Exemple de variation	Variation de distance associée
Nombre de bonnes réponses	+ 1 bonne réponse	- 7,28 mm
Durée moyenne de fixation	+ 10 ms	+ 2,58 mm
Attention subjective	+ 1 point sur l'échelle de notation (sur 10)	- 3,28 mm
Taux de bonnes réponses à la seconde tâche	+ 10 points	+ 3,73 mm

Note. Données estimées à partir du modèle linéaire général.

Nous avons également calculé un modèle pour l'écart-type des distances à l'écran. Ce modèle explique assez peu les différences de variabilité de distances observées : R^2 ajusté = 0,300, $F(31 ; 192) = 4,08$, $p < ,001$.

7.2.4 Discussion

L'analyse des résultats de cette expérimentation a permis de montrer un effet de la difficulté des textes (présence ou absence de ponctuation) sur la distance à l'écran. Lorsque les sujets lisent un texte difficile, ils sont plus proches de l'écran que lorsqu'ils lisent un texte facile. Néanmoins, en moyenne cette différence est faible, de l'ordre de 6 mm seulement. Cette valeur de variation est faible relativement aux variations totales observées pour chaque sujet. Nous acceptons donc l'hypothèse (2) : bien que l'effet de la difficulté soit présent, il n'est pas systématique.

Cette observation justifie le recours aux variables attentionnelles pour mieux appréhender les variations de distance à l'écran. Le modèle mis en place montre que des variables associées à la charge attentionnelle ont un effet sur la distance à l'écran. Les précédents résultats de la littérature montrant une réduction de la distance à l'écran lorsque la charge augmente (Qiu & Helbig, 2012) peuvent donc être transposés à une tâche de lecture. L'hypothèse (1) est confirmée : l'augmentation de la charge attentionnelle en lecture électronique entraîne un rapprochement à l'écran.

Le modèle linéaire général ainsi que l'observation directe des données montre qu'il existe une grande variabilité interindividuelle dans les distances de lecture. Bien que notre

étude n'apporte pas de nouveaux éléments d'explication sur ces différences, elle confirme cette observation déjà faite dans la littérature (Rempel et al., 2007) ainsi que l'hypothèse (3).

Un effet de la difficulté du texte sur la variabilité posturale a été trouvé : l'écart-type de la distance à l'écran est plus grand pour les textes difficiles que pour les textes faciles. Cet effet pourrait s'expliquer par le fait que les sujets se sont rapprochés durant la lecture de textes difficiles, ce mouvement ayant fait augmenter la variabilité de distance à l'écran.

Cette étude montre donc qu'il est possible d'utiliser, en plus des indicateurs traditionnels, la distance à l'écran comme marqueur de la charge attentionnelle. En outre, ce rapprochement est un phénomène dont il faut tenir compte lorsque l'on étudie l'attention. Celui-ci pourrait notamment interagir avec d'autres indicateurs. Par exemple, l'augmentation du diamètre de la pupille est souvent utilisée comme marqueur de l'augmentation de la charge attentionnelle (Jainta & Baccino, 2010). Mais le rapprochement à l'écran a des répercussions physiologiques sur la pupille. Le rapprochement à l'écran entraîne une augmentation de l'accommodation et l'accommodation entraîne une réduction du diamètre pupillaire. L'augmentation de la charge attentionnelle peut donc faire augmenter le diamètre pupillaire, et le faire diminuer en provoquant un rapprochement à l'écran. Les chercheurs utilisant cet indicateur pour mesurer la charge attentionnelle doivent donc tenir compte de ce phénomène, soit en fixant la distance à l'écran (mentonnière), soit en mesurant la distance et en utilisant cette donnée pour éliminer l'effet du rapprochement sur le diamètre pupillaire.

Si notre expérimentation a montré la présence du phénomène de « rapprochement attentionnel », elle n'en explique pas la cause. Plusieurs hypothèses pourraient expliquer la réaction posturale liée à l'attention :

- Le rapprochement à l'écran permettrait d'augmenter la taille des caractères sur la rétine, ce qui pourrait faciliter le traitement visuel des mots et libérer des ressources pour leur traitement sémantique.
- Le rapprochement pourrait également permettre d'augmenter la part du champ de vision alloué au texte, et ainsi diminuer la part allouée à d'éventuels distracteurs (environnement,...). En effet, la programmation des saccades oculaires est liée au déplacement de l'attention spatiale (Deplancke, Madelain, Chauvin, & Coello, 2011). Dans le cas de la lecture, on pourrait envisager que la présence de distracteurs autour du texte peut nécessiter un contrôle plus important de

l'attention visuelle pour que le lecteur puisse réaliser les saccades. En diminuant la part du champ visuel contenant de tels distracteurs, le rapprochement pourrait simplifier le processus de déplacement de l'attention spatiale.

Afin de déterminer si une de ces hypothèses est correcte, il faudrait réaliser de nouvelles manipulations expérimentales, par exemple en maintenant la taille angulaire des caractères fixe en dépit du rapprochement à l'écran, ou en contraignant la position de la tête et en faisant varier la distance de manière contrôlée...²⁴ Le lien entre la posture et le niveau d'attention est donc un phénomène d'intérêt, qu'il serait légitime d'investiguer plus avant.

²⁴ Le lien de causalité entre le rapprochement à l'écran et la modification du niveau attentionnel n'étant pas défini, ce type d'expérimentation permettrait également de savoir si le simple rapprochement aboutirait à une augmentation des ressources attentionnelles pour effectuer la tâche.

Chapitre 8 : Le quotient de lecture électronique

Ce chapitre correspond à l'aboutissement du travail de cette thèse. Nous avons pu montrer dans les chapitres précédents l'aspect multidimensionnel de la lecture électronique. Dimension par dimension, nous avons proposé des tests permettant d'évaluer la performance durant cette activité. Un des postulats de départ des travaux était le suivant : l'évaluation de la lecture électronique ne peut être complète que si elle porte sur l'ensemble des dimensions impliquées dans cette activité complexe.

Nous proposons donc ici une mise en application simultanée de ces tests. Le chapitre décrit la mise en œuvre de la batterie de tests, afin d'évaluer trois dispositifs d'affichage électroniques : un écran d'ordinateur fixe, un vidéoprojecteur et une tablette. Les données de sujets ayant pris part aux tests ont été recueillies et analysées pour répondre à deux objectifs :

- Vérifier que les tests permettent de rendre compte des différences entre supports²⁵.
- Construire un modèle de traitement des données permettant d'obtenir des indicateurs globaux sur la performance de lecture électronique.

L'indicateur général qui décrit la performance pour un support électronique donné sera appelé le « Quotient de Lecture Electronique » (QLE). La suite de ce chapitre montrera comment nous avons mis en application les tests afin de le calculer.

8.1 Les tests de la batterie du Quotient de Lecture Electronique

Le QLE sera obtenu à l'aide des tests présentés dans les chapitres précédents. La batterie est composée à la fois de tests psychophysiques ou psychologiques, mais aussi de tests instrumentés qui n'impliquent pas le lecteur. La variété de ces tests vise à quantifier toutes les dimensions de la posture électronique. Le nom de chaque test et mesure est donné dans la liste ci-après :

²⁵ Ici, et dans la suite de ce chapitre, nous parlons de « différences entre les supports ». Il s'agit d'un abus de langage, il faut entendre que nous évaluons et comparons les performances de lecture associées à l'utilisation des supports dans les conditions d'expérimentations que nous avons mises en place. Les différences ne peuvent pas être expliquées uniquement par le support, mais doivent être considérées dans leur contexte.

- Visibilité :

Test psychophysique : test de caractérisation psychométrique présenté dans la section *4.2 Technique de caractérisation psychométrique*.

Mesure instrumentée de l'écran : mesure des profils de luminance des écrans (une seule fois par écran).

- Lisibilité :

Tests psychologiques : tests d'identification de lettres de décision lexicale (*5.1 Test d'identification de lettres et test de décision lexicale*), test de vitesse de lecture maximale (*5.2 La mesure de la vitesse de lecture maximale*).

Mesure complémentaire : mesure du temps de réponse de base pour les tests d'identification de lettres et de décision lexicale (*5.1.3 Mesure du temps de réponse de base*).

Mesure instrumentée de l'écran : mesure de l'écart entre le temps d'affichage demandé et temps d'affichage réel pour le test de vitesse de lecture (une seule fois par écran).

- Compréhension et posture :

Tests psychologiques : test de compréhension basé sur les textes générés automatiquement (*6.2 Un générateur automatique de textes pour l'évaluation de la compréhension*), tâche « Remember-Know » (*6.6 Tâche « Remember-Know »*) et mesure de la distance de lecture afin de mesurer le niveau d'attention (*7.2 L'effet de la lecture électronique sur la posture*).

8.2 Matériel et sujets

8.2.1 Dispositifs d'affichage électroniques

Nous avons utilisé 3 dispositifs d'affichage électroniques sur lesquels ont été réalisées les mesures :

- Un écran fixe (écran d'ordinateur de bureau). Nous avons utilisé un écran AL1717 (Acer, Taipei, Taïwan). Il s'agit d'un écran TFT-LCD de 17'' de diagonale et de dimension 33,8 cm (horizontale) par 27,0 cm (verticale), soit un ratio de 5:4. Nous avons fixé la définition d'affichage à 1280 * 1024 pixels.

- Un vidéoprojecteur de type EP7150 (Optoma, Milpitas, Etats-Unis). Ce projecteur utilise la technologie DLP. Le vidéoprojecteur a été placé et réglé de manière à afficher sur une surface de 118,7 cm par 89,0 cm. La définition a été fixée à 1024 * 768 pixels.

Pour l'écran fixe et le projecteur, l'affichage a été généré par un PC portable de type VAIO Duo 11 (Sony, Tokyo, Japon), le système d'exploitation associé est Windows 8.1 professionnel.

- Une tablette tactile de type Google Nexus 7 édition 2013 (Asus, Taipei, Taïwan). L'écran utilise la technologie LED IPS LCD. L'écran présente une diagonale de 7'' et de dimension 15,1 cm par 9,5 cm, une définition de 1920 * 1200 pixels, soit un ratio de 16:10, sous Android 4.3.

Les tests ont été codés sous Python 2.7, en utilisant notamment le module Kivy 1.8. Nous avons choisi ce module car il permet de programmer des interfaces graphiques qui peuvent être utilisées à la fois sous Windows et sous Android.

8.2.2 Salle d'expérimentation

La salle de test était une salle sans fenêtres et avec éclairage artificiel, de manière à pouvoir contrôler l'ambiance lumineuse qui était fixée à 220 lux au niveau du bureau.

Les sujets étaient installés sur un fauteuil avec dossier et accoudoirs. Lorsqu'ils effectuaient les tests sur l'écran fixe ou le projecteur, ils pouvaient prendre appui sur un bureau, sur lequel était posé l'écran fixe, ainsi que le clavier et la souris.

Le vidéoprojecteur était situé entre le sujet et un mur blanc qui servait à la projection.

8.2.3 Taille des stimuli

La taille des stimuli a été définie de manière à être constante d'un support à l'autre. Nous nous sommes basés sur des distances théoriques aux écrans :

- Pour l'écran fixe, nous avons considéré la distance de 63 cm en nous basant sur les travaux de Jaschinski (2002), qui avait mesuré la distance moyenne à l'écran lorsque l'on laisse des sujets se positionner librement face à un écran.

- Pour le projecteur, la distance considérée était de 380 cm, cette mesure ayant été réalisée directement à partir de la disposition de la salle.

- Pour la tablette, nous nous sommes basés sur les résultats de l'expérience présentée dans la section 7.2 *L'effet de la lecture électronique sur la posture* ; en effet le support employé dans cette expérimentation (liseuse de 6'') avait un gabarit relativement similaire à celui de la tablette employée pour la mesure du QLE.

A partir de ces distances théoriques, nous avons pu définir la taille à l'écran qui permettait de maintenir une taille angulaire constante entre tous les supports. Pour les tests où des caractères sont affichés, la taille de ces caractères a été paramétrée de manière à ce qu'un caractère « o » minuscule en police « Arial » ait une hauteur correspondant à 13,39 minutes d'angle visuel. Pour le test de caractérisation psychométrique où les stimuli sont des disques, leur diamètre a été fixé à 1° d'angle visuel.

8.2.4 Interfaces de réponse

Pour les tests passés sur l'écran fixe et le vidéoprojecteur, les réponses étaient données via un clavier et une souris sans fil, que le sujet pouvait déplacer librement sur le bureau. Pour la tablette, les réponses étaient entrées à l'aide de l'écran tactile.

Les modalités de réponse sont données dans la section 8.3 *Recueil des données*.

8.2.5 Matériel de mesure

Pour la caractérisation du profil de luminance des écrans, nous avons utilisé un luminancemètre LS-110 (Konica Minolta, Tokyo, Japon).

Pour mesurer l'écart entre les durées d'affichage demandées pour le test de vitesse de lecture et les durées réelles, nous avons utilisé une caméra à haute fréquence d'acquisition Playstation Eye (Sony, Tokyo, Japon) paramétrée à 150Hz.

Afin de mesurer la distance de lecture durant le test de compréhension, nous avons utilisé un télémètre à ultrasons SRF10 (Devantech, Attleborough, Royaume-Uni), composé d'un émetteur et d'un récepteur. Nous avons détaché l'émetteur du circuit imprimé pour l'attacher à un casque (relié par un fil souple) afin d'obtenir la distance entre l'émetteur et le récepteur, attaché lui-même au dispositif d'affichage. Ce télémètre était relié à un convertisseur I2C vers USB : USB-ISS (Devantech). Ainsi, le système a pu être relié au port USB d'un ordinateur qui assurait le traitement des données. Pour la tablette, le récepteur était fixé par velcros au milieu de la partie haute de l'appareil (voir la *Figure 24*). Pour les tests sur écran fixe et vidéoprojecteur, le récepteur était fixé au milieu de la bordure supérieure de

l'écran. Dans le cas de la vidéo-projection, le récepteur se situait donc légèrement en dessous de la zone d'affichage du point de vue du lecteur ; la distance entre le mur et le récepteur était ajoutée pour obtenir une estimation de la distance de lecture.

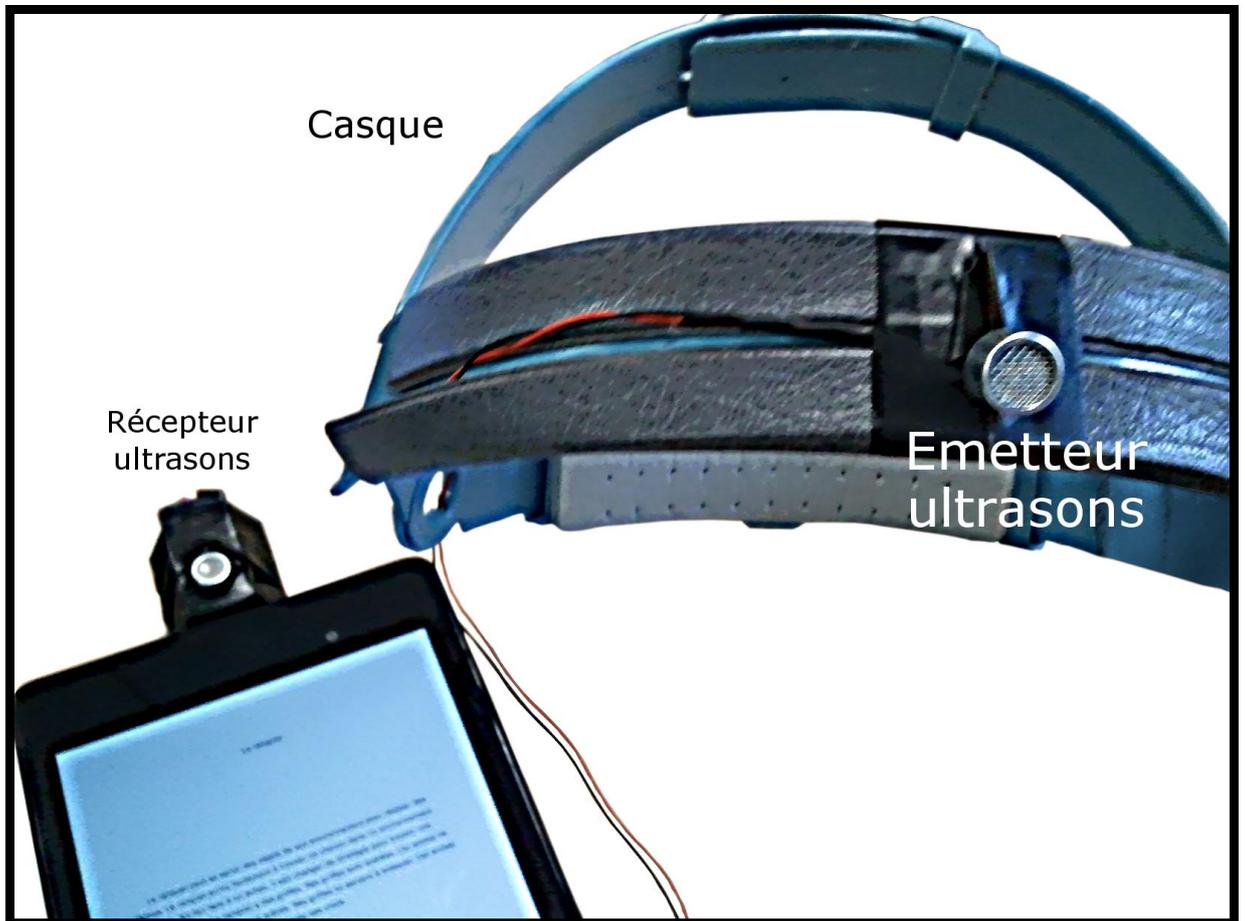


Figure 24. Photo du montage servant à mesurer la distance à l'écran.

8.2.6 Participants

Au total, 36 sujets ont pris part à l'expérimentation. Il s'agissait de 15 femmes et 21 hommes. Leurs âges étaient compris entre 22 et 57 ans, avec une moyenne de 39,0 ans (écart-type : 9,8). Ils étaient tous de langue maternelle française. Si un sujet déclarait avoir eu ou avoir des problèmes liés à sa vision, il n'était pas retenu. Les sujets étaient encouragés à porter leur correction en lentilles de contact ou en lunettes de vue s'ils en avaient l'habitude durant l'utilisation de supports électroniques. L'acuité binoculaire était vérifiée au loin et au près, et devait être au minimum de 10/10 pour que le sujet soit inclus dans l'étude.

Chaque participant a signé un formulaire de consentement expliquant que l'étude portait sur l'évaluation d'une batterie de test visant à mesurer la qualité de lecture sur support électronique.

8.3 Recueil des données

8.3.1 Procédure générale

Pour chaque sujet, le recueil de données prenait la forme de deux passations, durant lesquelles il réalise tous les tests du QLE sur chaque support.

La première passation comprenait la lecture et la signature du formulaire de consentement, la vérification des critères d'inclusion (dont mesure de l'acuité en vision de près et en vision de loin). Ensuite, les tests du QLE étaient expliqués et passés pour un premier dispositif d'affichage électronique.

La seconde passation correspondait à l'application des tests sur les deux autres dispositifs électroniques.

L'ordre de présentation des dispositifs était contrebalancé. Il s'agit d'un contre-balancement complet (les 6 ordres de présentations des supports ont chacun été mis en œuvre 6 fois au total).

Dans le cas particulier de la tablette, l'écran était nettoyé à l'aide d'un tissu microfibre avant chaque passation, et si nécessaire entre deux tests. En effet, la présence de salissures (traces de doigts) à l'écran peut altérer la visibilité de l'écran (Hirota, Uozato, Arai, & Shibata, 2013). Pour les tests passés sur tablette, les sujets doivent répondre par des appuis sur l'écran tactile. Afin d'éviter les salissures à l'écran, les parties où les sujets devaient appuyer se situaient toujours à des positions de l'écran différentes des endroits où les stimuli sont affichés (sauf dans le cas des textes où ceux-ci peuvent éventuellement remplir l'écran).

8.3.2 Procédures associées aux tests

L'ordre de passage des tests est schématisé dans la *Figure 25*, la procédure associée à chaque test, ainsi qu'aux mesures complémentaires est donnée dans les points suivants.

8.3.2.1 Test de visibilité

Caractérisation psychométrique

Le but du test de caractérisation psychométrique est d'obtenir une information sur la visibilité du dispositif d'affichage électronique à partir des différences de niveaux RGB nécessaires pour que le sujet puisse percevoir deux niveaux de gris différents. Le détail de la procédure est donné dans *4.2 Technique de caractérisation psychométrique*.

Le sujet doit déterminer à plusieurs reprises la position d'un disque gris parmi 4 positions possibles. Pour la tablette, le sujet doit donner sa réponse à l'aide d'un pavé directionnel virtuel situé en bas à droite de l'écran. Ce pavé est visible pour les autres supports, mais pas actionnable : le sujet doit alors donner ses réponses à l'aide du pavé directionnel du clavier.

Il est demandé au sujet de donner la position du disque. Il n'y a pas de limite de durée pour donner la réponse, mais le sujet est invité à rapidement donner une réponse au hasard s'il ne le voit pas.

Les données ainsi obtenues permettent de déterminer la courbe de caractérisation subjective, et d'obtenir un indicateur de visibilité : l'aire sous la courbe (plus cet aire est faible, meilleure est la capacité de discrimination des disques). Le plan d'expérience associé à ce test est donc $\underline{S}_{36} * \text{Support}_3$.

Mesure instrumentée de l'écran : caractérisation du profil de luminance de l'écran

Pour chaque dispositif d'affichage électronique, nous avons mesuré à l'aide d'un luminancemètre la luminance associée à chaque niveau de gris affichable par l'écran (256 niveaux : du noir [0, 0, 0] au blanc [255, 255, 255]).

8.3.2.2 Tests de lisibilité

Test d'identification de lettres et de décision lexicale

La procédure associée à ces deux tests est regroupée car ces derniers sont très proches dans leurs formes. Le but du test d'identification de lettres est de mesurer la capacité du sujet à identifier rapidement une lettre parmi une suite de lettres. Pour le test de décision lexicale, c'est la capacité à reconnaître un mot rapidement qui est évaluée. Le détail de la procédure est donné dans *5.1 Test d'identification de lettres et test de décision lexicale*.

Pour le premier test, le sujet doit déterminer le plus rapidement possible et à plusieurs reprises si une lettre cible affichée avant un groupe de lettre fait partie ou non de ce dernier.

Pour le second test, on affiche un mot ou un pseudo-mot, et le sujet doit déterminer le plus rapidement possible s'il s'agit ou non d'un mot du français. Pour la tablette, le sujet doit appuyer sur les boutons « oui » ou « non » qui sont situés en bas à droite de l'écran. Ces boutons sont également affichés pour les autres supports, mais ne sont pas activés : le sujet doit donner sa réponse en appuyant sur les touches « O » ou « N » du clavier.

On demande au sujet de fixer la croix de fixation qui apparaît avant chaque stimulus, puis de regarder le stimulus et de donner sa réponse le plus rapidement possible. Il est expliqué au sujet que des erreurs sont très susceptibles d'être faites, et qu'en cas d'erreur, il doit essayer de ne pas en tenir compte et se concentrer pour la suite des essais.

On mesure la durée de réponse ainsi que le nombre d'erreurs pour chaque essai. Pour le test d'identification de lettres, il y a deux facteurs expérimentaux : présence ou absence de la lettre cible et présence ou absence de lettres visuellement proches de la lettre cible. Le plan expérimental associé est donc $\underline{S}_{36} * \text{Support}_3 * \text{Voisins}_2 * \text{Cible}_2 * R_{15}$. Pour la décision lexicale, nous avons également construit deux facteurs expérimentaux : présentation (mot ou pseudo-mot) et difficulté (mots ou pseudo-mots facilement ou difficilement identifiables). Le plan est donc $\underline{S}_{36} * \text{Support}_3 * \text{Difficulté}_2 * \text{Mot_PMot}_2 * R_{15}$. Pour les deux tests, on calcule la durée de réponse moyenne et le taux d'erreur pour chaque condition.

Test complémentaire : temps de réponse de base

Afin de pouvoir mesurer sur les deux tests précédents l'effet visuel des dispositifs d'affichage, il est nécessaire de mesurer la part de la performance associée à l'interaction avec le support. Nous avons donc mesuré le temps de réponse de base pour chaque sujet et chaque support, le détail de la procédure est donné dans 5.1.3 *Mesure du temps de réponse de base*. L'affichage devient aléatoirement orange ou noir et le sujet doit déterminer le plus rapidement possible cette couleur (à l'aide de boutons « orange » et « noir » pour la tablette, et à l'aide des touches « O » et « N » du clavier pour les autres supports). On obtient ainsi un temps de réponse de base qui sera soustrait aux temps de réponses obtenus par les tâches d'identification de lettres et de décision lexicale.

Test de vitesse de lecture

Le but de ce test est de déterminer la vitesse de lecture maximale associée à la lecture de chaque sujet sur chaque support. Le détail de la procédure est donné dans la section 5.2 *La*

mesure de la vitesse de lecture maximale. Nous avons utilisé ici la modalité orale du test : le sujet doit lire le plus vite possible une phrase affichée pour une durée limitée, et la citer à l'oral une fois que celle-ci est masquée. Lorsque le sujet a donné sa réponse, on affiche à nouveau la phrase pour vérification. Pour l'écran fixe et le vidéoprojecteur, l'expérimentateur vérifie que la phrase a été effectivement lue et entre la réponse au clavier. Pour la tablette, le sujet vérifie sa réponse en appuyant à l'écran, puis appuie sur un bouton « Oui » lorsque la phrase a été correctement lue et « Non » lorsqu'il n'a pas lu la phrase entièrement. Le sujet était tenu de citer la phrase à voix haute, même si ce n'est pas l'expérimentateur qui entrait la réponse, et ce dernier restait à côté du sujet pour vérifier éventuellement ses réponses.

Nous obtenons ainsi une durée minimale d'affichage pour que le sujet puisse lire la phrase, le plan expérimental est donc $S_{36} * Support_3$.

Mesure instrumentée de l'écran : mesure de la durée effective d'affichage des phrases

Le but de cette mesure était de contrôler que les durées d'affichage réelles des phrases sur les différents dispositifs correspondaient bien aux durées d'affichage requises par le logiciel. Nous avons donc fait afficher à chaque écran 621 phrases générées automatiquement. Les durées d'affichage demandées allaient de 40 ms à 640 ms (avec un pas de 1 ms), valeurs qui étaient utilisées dans un ordre pseudo-aléatoire. Les affichages étaient filmés par une caméra ayant une fréquence d'acquisition de 150 Hz. Quand une phrase est affichée, la luminosité de l'écran est réduite. Nous avons écrit un script Matlab pour mesurer les changements de luminosité et ainsi pouvoir mesurer les durées d'affichage des phrases. Les durées d'affichage mesurées par la caméra étaient soustraites aux durées d'affichage demandées. Nous avons calculé la moyenne des différences pour définir l'erreur constante et l'écart-type de ces différences pour définir l'imprécision.

8.3.2.3 Test de compréhension

Lecture et compréhension des textes générés automatiquement

Le but du test est de mesurer la compréhension à différents niveaux (mémorisation et inférence) pour la lecture de textes de difficulté variable. Le processus de génération des textes, ainsi que la procédure de mesure associée sont détaillés dans le *Chapitre 6 : L'évaluation de la compréhension.*

Le lecteur doit lire 6 textes : 2 textes faciles, 2 textes moyens et 2 textes difficiles. Un texte sur deux est mis en forme (présence d'alinéas et de retraits de début de paragraphe), les autres textes sont présentés en bloc. Le plan expérimental est donc $S_{36} * Support_3 * Difficulté_3 * MiseEnForme_2$.

Un bouton est situé en bas du texte, le sujet doit appuyer ou cliquer dessus pour marquer la fin de la lecture. Chaque texte est suivi de 4 questions de mémorisation de type « vrai ou faux », le sujet doit appuyer/cliquer sur le bouton « vrai » ou sur le bouton « faux » pour donner sa réponse. Enfin, chaque texte est suivi de 2 questions d'inférences, où le lecteur doit situer sur un continuum le niveau associé à un trait général sur le concept décrit dans le texte (dangerosité d'un animal, niveau de pollution d'une ville,...). Il déplace un curseur en le faisant glisser avec l'écran tactile ou à l'aide de la souris.

Pour chaque texte, nous mesurons la durée de lecture par mot (durée entre le début d'affichage du texte et moment où le bouton de fin de lecture est actionné, divisé par le nombre de mots composant le texte), le nombre d'erreurs aux questions de mémorisation, l'écart entre les valeurs de traits généraux entrées par le sujet et celles calculées par le générateur de texte, et la durée moyenne de réponse aux questions.

Mesure de la distance à l'écran pendant la lecture

Afin d'obtenir un marqueur de la charge attentionnelle durant la lecture des textes, nous avons mesuré la distance à l'écran. Nous calculons pour chaque texte la moyenne et l'écart-type de toutes les valeurs de distance acquises durant la lecture.

Tâche R-K

La tâche « Remember versus Know » permet d'évaluer la mémorisation des titres de textes, ainsi que de définir quel type de mémoire est mis en œuvre pour la mémorisation. La tâche permet de vérifier la mémorisation des titres des textes présentés. Nous avons calculé les indicateurs globaux de performance à cette tâche à l'aide d'un modèle non-paramétrique (Snodgrass & Corwin, 1988) pour chaque support, le détail du modèle de calcul et de la procédure est donné dans 6.6 Tâche « Remember-Know ». La tâche RK était réalisée entièrement à l'oral (proposition des listes de mots et réponses du sujet), il s'agissait du dernier test pour chaque support.

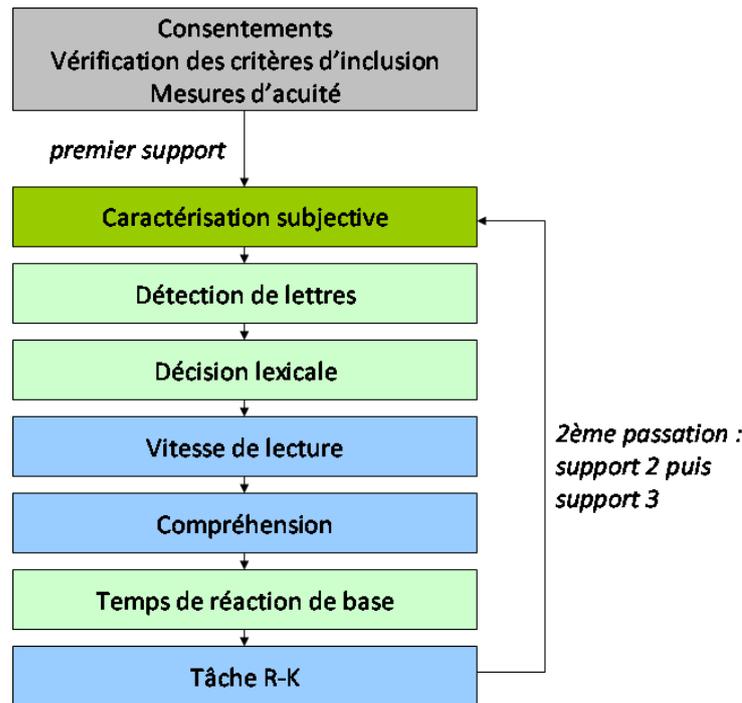


Figure 25. Schéma de déroulement de l'expérimentation.

8.4 Tests d'hypothèses

Une première phase de l'analyse des résultats obtenus grâce au recueil de données consiste à vérifier la présence d'effet des différents facteurs expérimentaux manipulés dans les tests, ainsi que la présence d'effet du dispositif d'affichage électronique sur la performance des sujets.

8.4.1 Visibilité

8.4.1.1 Test de caractérisation subjective

Une ANOVA à mesures répétées montre que l'aire sous la courbe de la fonction de caractérisation subjective est différente entre les supports testés (voir *Figure 26*), $F(2, 70) = 987,56, p < ,001, \eta^2_p = ,99$. Ce résultat est confirmé avec une correction de Huynh-Feldt ($p < ,001$). Une analyse post-hoc réalisée au moyen du test HSD de Tukey montre une différence entre l'aire sous la courbe associée au projecteur et celle de chacun des deux autres supports ($p < ,001$). Le test ne montre pas de différence entre les aires obtenues pour la tablette et celles de l'écran fixe ($p = ,703$).

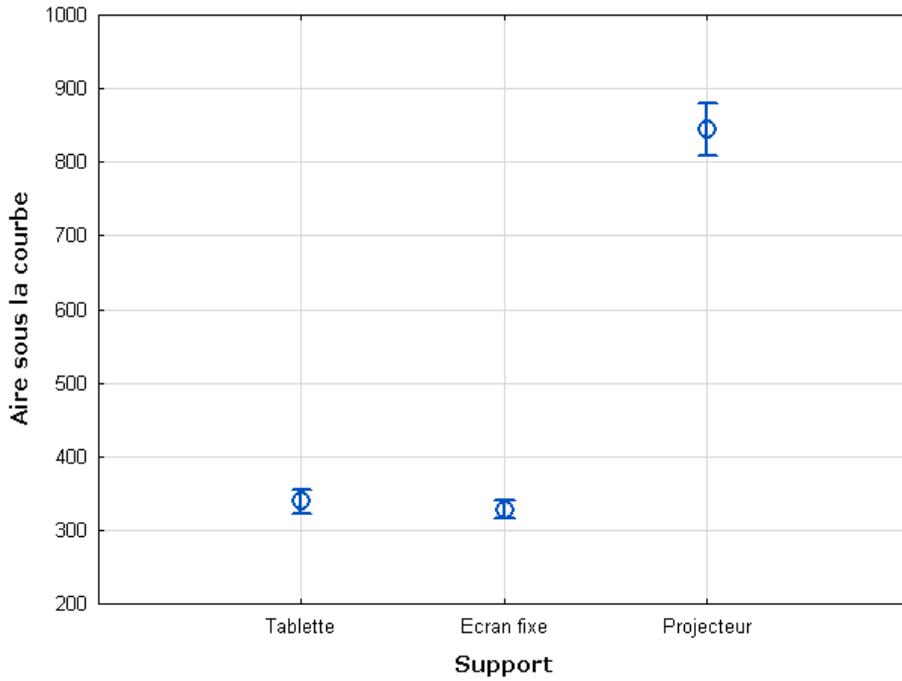


Figure 26. Aire sous la courbe de la fonction de caractérisation psychométrique, en fonction du support. Dans ce graphique, ainsi que dans tous les graphiques suivants, les barres verticales représentent les intervalles de confiance à 95%.

8.4.1.2 Mesure des profils de luminance des écrans

Les profils de luminance des trois dispositifs d'affichage électroniques sont donnés dans la *Figure 27*.

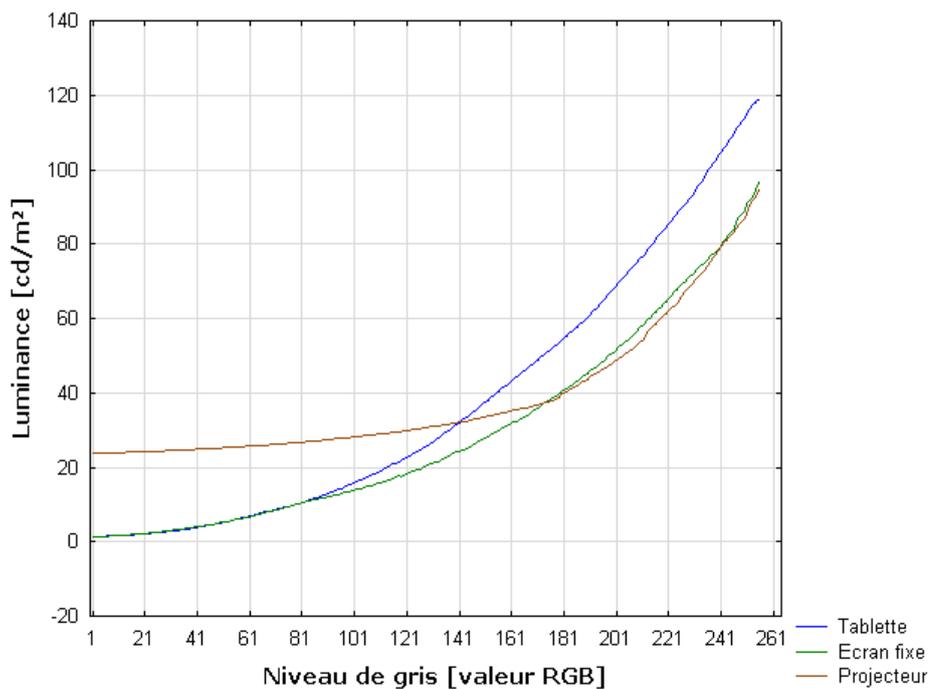


Figure 27. Profils lumineux des trois dispositifs d'affichages testés.

Les valeurs de contraste calculées à partir des données de luminance (voir *Table 9*) vont dans le sens des résultats obtenus par la caractérisation subjective : le contraste est très inférieur pour le vidéoprojecteur que pour les autres dispositifs.

Table 9

Valeurs de contraste des dispositifs d'affichage électroniques testés

	Contrastes		
	Tablette	Ecran fixe	Projecteur
Michelson	0,97750937	0,97469903	0,59938306
Weber	-0,98862679	-0,98718743	-0,74951783

Note. Les contrastes sont calculés à partir de la luminosité minimum (niveau RGB [0, 0, 0]) et de la luminosité maximum (niveau RGB [255, 255, 255]).

8.4.2 Lisibilité

8.4.2.1 Temps de réponse de base

Une ANOVA à mesures répétées montre que le temps de réponse de base (temps de réaction lié à l'interaction) est différent entre les supports testés (voir *Figure 28*), $F(2, 70) = 152,18, p < ,001, \eta^2_p = ,99$. Ce résultat est confirmé avec une correction de Huynh-Feldt ($p < ,001$). Une analyse post-hoc réalisée au moyen du test HSD de Tukey montre que le temps de réponse est plus long sur tablette que sur chacun des autres supports (pour chacun, $p < ,001$). Les sujets répondent en moyenne 0,196 s plus rapidement sur l'écran fixe et 0,165 s plus rapidement sur le vidéoprojecteur que sur la tablette. Le test montre également que le temps de réponse est plus court sur l'écran fixe que sur le vidéoprojecteur ($p = ,037$), en moyenne la réponse est plus rapide de 0,030 s.

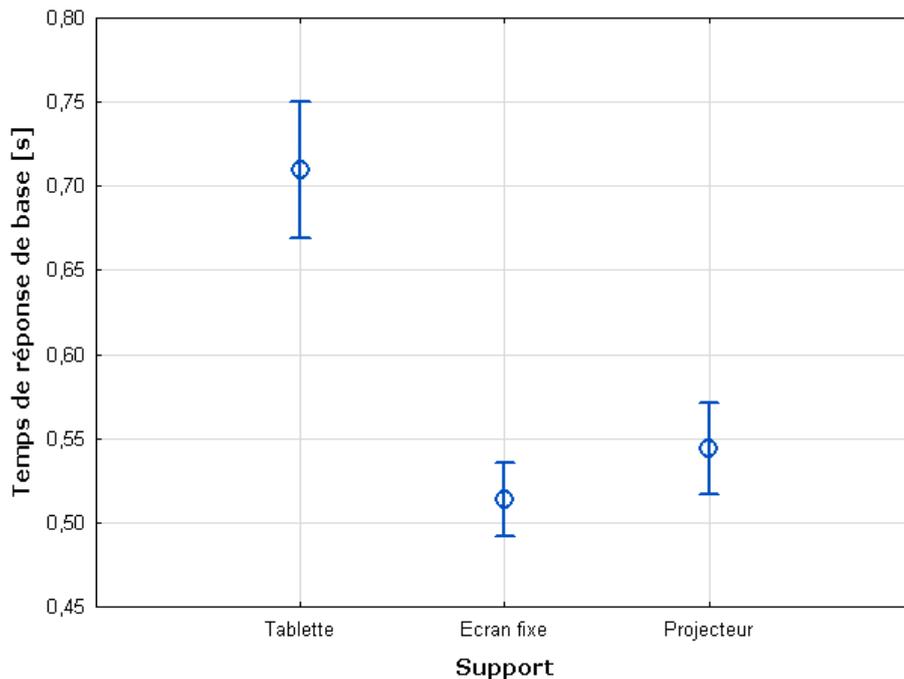


Figure 28. Temps de réponse de base en fonction du support.

8.4.2.2 Test d'identification de lettres

Pour les facteurs expérimentaux, une ANOVA à mesures répétées détecte plusieurs effets sur les temps de réponse corrigés au test de d'identification de lettres (temps de réponse mesuré – temps de réponse de base), une représentation graphique de ces résultats est donnée dans la *Figure 29* :

- Effet de la présence de la lettre cible : la réponse des sujets est donnée plus rapidement quand la lettre cible est présente, $F(1, 35) = 70,04, p < ,001, \eta^2_p = ,67$, en moyenne la réponse est plus rapide de 0,141 s. Ce résultat est confirmé avec une correction H-F ($p < ,001$).

- Effet de voisinage : une lettre est identifiée plus rapidement lorsqu'elle n'est pas accompagnée de lettres visuellement proches (voisines), $F(1, 35) = 406,27, p < ,001, \eta^2_p = ,92$, en moyenne la réponse est plus rapide de 0,247 s. Ce résultat est confirmé avec une correction H-F ($p < ,001$).

- Interaction présence de la cible et voisinage : le ralentissement dû à la présence de lettres voisines est accentué lorsque la lettre cible est absente, $F(1, 35) = 78,49, p < ,001, \eta^2_p = ,69$. Ce résultat est confirmé avec une correction H-F ($p < ,001$). Une analyse post-hoc HSD

de Tukey montre que chacune des 4 conditions est significativement différentes de toutes les autres (pour chaque comparaison, $p < ,001$).

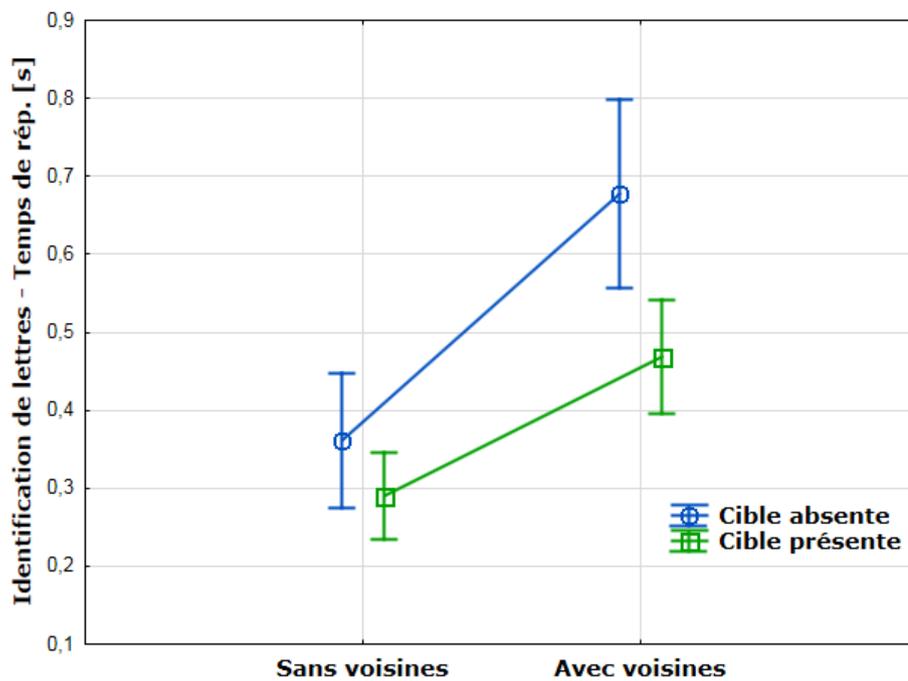


Figure 29. Temps de réponse corrigé en fonction de la présence de lettres voisines (visuellement proches) et en fonction de l'absence ou de la présence de la lettre-cible.

Cette même ANOVA montre une différence de temps de réponse corrigé entre les supports (voir *Figure 30*), $F(2, 70) = 8,80$, $p < ,001$, $\eta^2_p = ,20$. Ce résultat est confirmé avec une correction H-F ($p < ,001$). Le test HSD de Tukey montre que les temps de réponse sont plus longs relativement à la tablette sur écran fixe et sur vidéoprojecteur (respectivement, on observe des réponses plus lentes de 0,092 s et de 0,127s, avec les valeurs $p = ,012$ et $p < ,001$). Ce test ne montre pas de différence entre l'écran fixe et le projecteur ($p = ,503$).

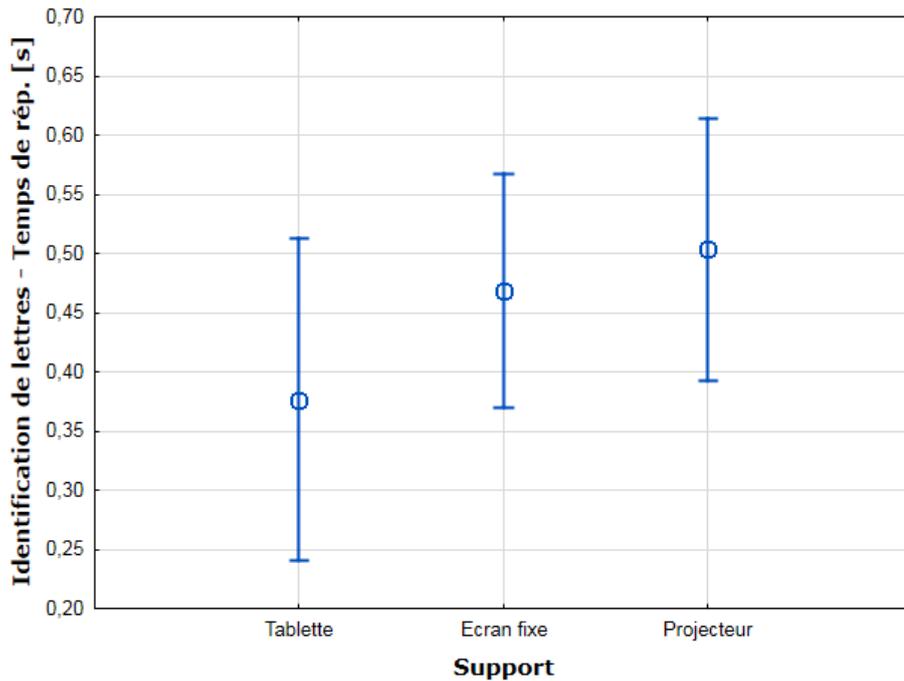


Figure 30. Temps de réponse pour l'identification de lettres en fonction du support.

Enfin, l'ANOVA montre une interaction entre la présence de lettres voisines et le dispositif d'affichage $F(2, 70) = 3,94, p = ,024, \eta^2_p = ,10$. Ce résultat est confirmé avec une correction H-F ($p = ,026$). Un test HSD de Tukey permet de montrer que le ralentissement dû à la présence de lettres voisines est plus faible sur l'écran fixe que sur la tablette, et il est plus faible sur la tablette que sur le projecteur, les ralentissements moyens sont de 0,217 s pour l'écran fixe, de 0,245 s pour la tablette et de 0,279 s pour le projecteur (dans tous les cas, $p < ,001$). L'ANOVA n'a pas détecté d'autres interactions (pour chacune, $p > ,300$).

Une autre ANOVA à mesures répétées a également été conduite pour observer les effets des facteurs expérimentaux sur les taux d'erreurs (lettre considérée comme absente alors qu'elle était affichée ou lettre considérée comme présente alors qu'elle ne l'était pas) :

- Effet de la présence de la lettre cible (*Figure 31*) : on observe en moyenne un taux d'erreur supérieur de 2,40 points lorsque la cible est absente, $F(1, 35) = 33,04, p < ,001, \eta^2_p = ,49$. Cet effet est confirmé par une correction H-F ($p < ,001$).

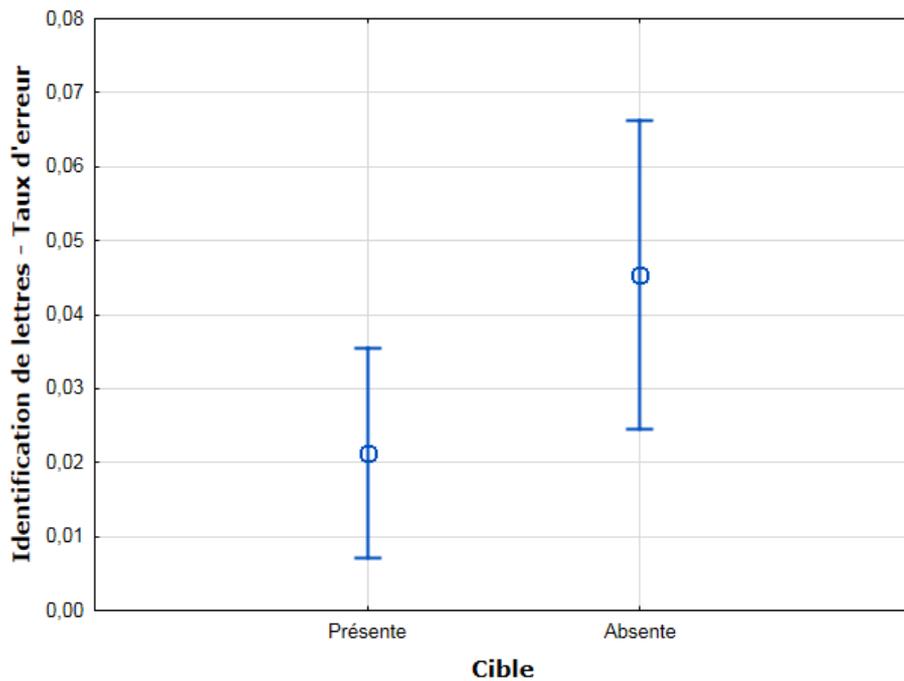


Figure 31. Taux d'erreur à l'identification de lettres en fonction de la présence de la lettre-cible.

- Effet de la présence de lettres voisines (*Figure 32*) : on observe en moyenne un taux d'erreur supérieur de 2,64 points lorsque des lettres visuellement proches sont présentes, $F(1, 35) = 37,73, p < ,001, \eta^2_p = ,52$. Cet effet est confirmé par une correction H-F ($p < ,001$).

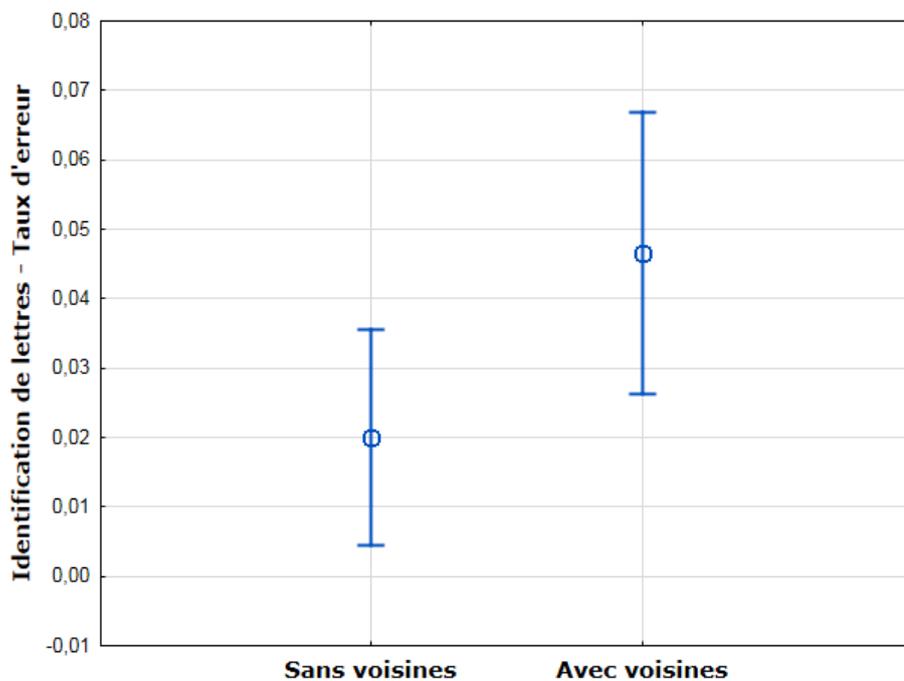


Figure 32. Taux d'erreur à l'identification de lettres en fonction de la présence de lettres voisines.

- Effet du support (*Figure 33*) : on observe une différence sur le taux d'erreur due aux supports, $F(2, 70) = 5,57, p = ,006, \eta^2_p = ,14$. Cet effet est confirmé par une correction H-F ($p = ,006$). Un test HSD de Tukey montre que le taux d'erreur est plus important pour le projecteur que pour la tablette (1,48 points de différence, $p = ,012$) ainsi que pour l'écran fixe (1,43 points de différence, $p = ,016$). Le test ne montre pas de différence entre la tablette et l'écran fixe ($p = ,995$).

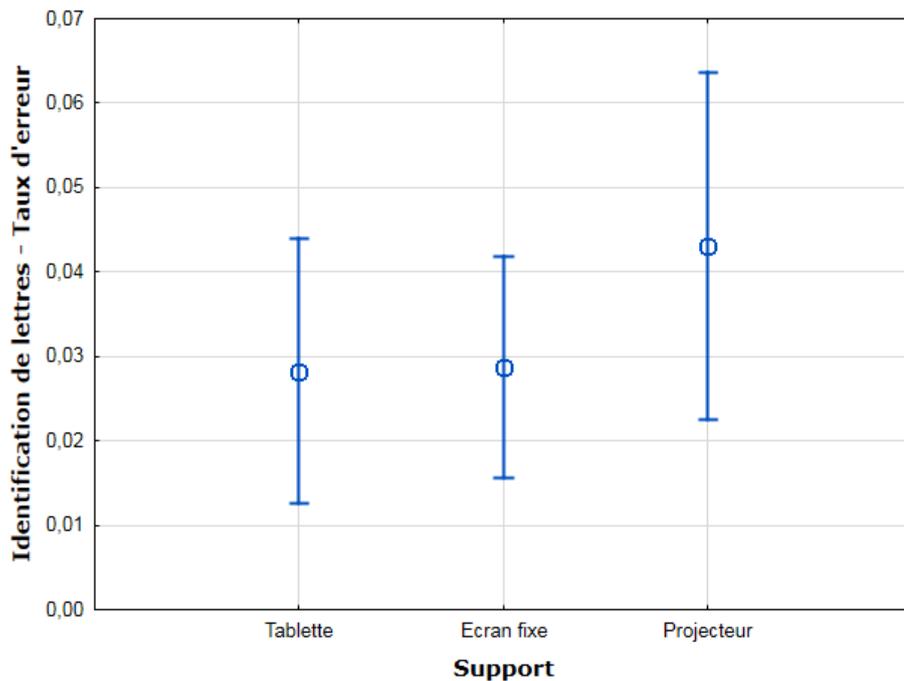


Figure 33. Taux d'erreur à l'identification de lettres en fonction du support.

L'ANOVA ne permet pas de montrer d'effet d'interaction : pour chacune $p > ,050$.

8.4.2.3 Test de décision lexicale

Pour les facteurs expérimentaux, une ANOVA à mesures répétées détecte plusieurs effets sur les temps de réponse corrigés pour le test de décision lexical (temps de réponse mesuré – temps de réponse de base) :

- Effet de la catégorie de stimulus présenté (mot ou pseudo-mot) : la réponse des sujets est donnée plus rapidement lorsque l'on affiche un mot plutôt qu'un pseudo-mot, $F(1, 35) = 41,83, p < ,001, \eta^2_p = ,54$, en moyenne la réponse est plus rapide de 0,181 s (voir *Figure 34*). Ce résultat est confirmé avec une correction H-F ($p < ,001$).

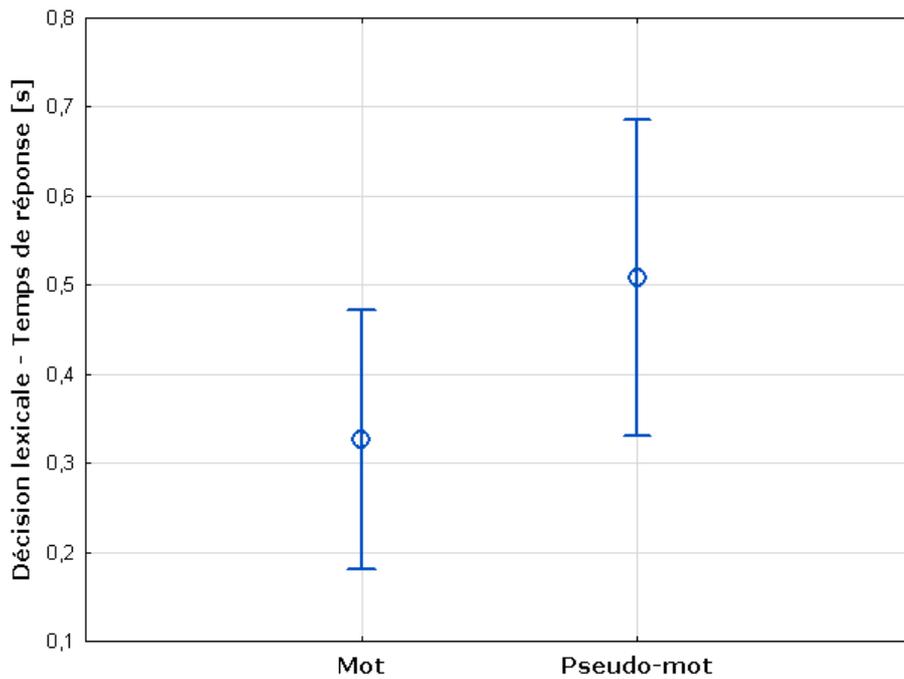


Figure 34. Temps de réponse en fonction du stimulus présenté (mot ou pseudo-mot).

- Effet de difficulté des stimuli (*Figure 35*) : les réponses sont données plus lentement lorsque le stimulus présenté appartient à la catégorie difficile, $F(1, 35) = 169,66, p < ,001, \eta^2_p = ,83$, en moyenne la réponse est plus rapide de 0,163 s. Ce résultat est confirmé avec une correction H-F ($p < ,001$).

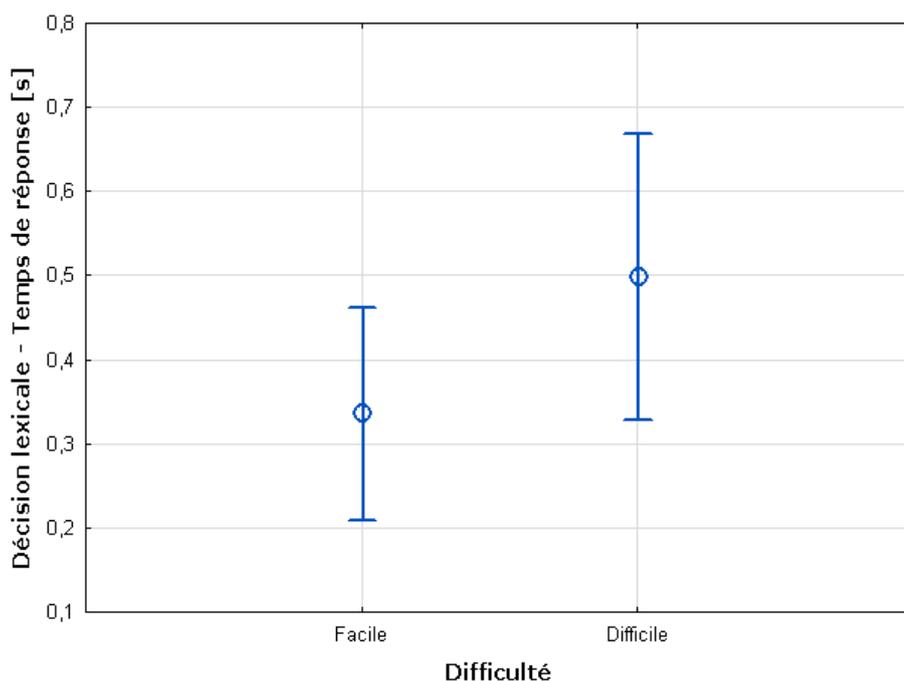


Figure 35. Temps de réponse à la décision lexicale en fonction de la difficulté.

- Effet du support (*Figure 36*) : l'analyse montre un effet du support sur les temps de réponse, $F(2, 70) = 4,37, p = ,016, \eta^2_p = ,11$. Ce résultat est confirmé avec une correction H-F ($p < ,018$). Un test post-hoc de Tukey montre que les réponses sont en moyenne plus rapides de 0,103 s sur la tablette que sur l'écran fixe ($p = ,023$). Le test ne permet pas de statuer sur la différence de 0,090 s observée entre la tablette et le vidéoprojecteur ($p = ,053$). Enfin, le test ne détecte pas de différence entre l'écran fixe et le projecteur ($p = ,937$).

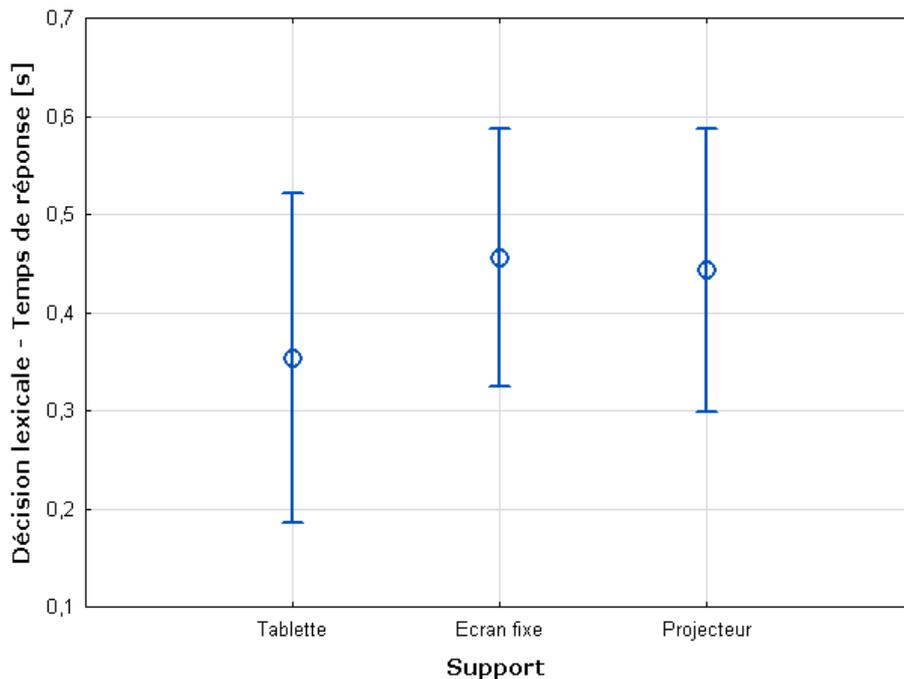


Figure 36. Temps de réponse à la décision lexicale en fonction du support.

L'ANOVA ne détecte aucune interaction : pour chaque interaction testée $p > ,100$.

Une autre ANOVA à mesures répétées a également été conduite pour observer les effets des facteurs expérimentaux sur les taux d'erreurs (mot considéré comme un pseudo-mot ou pseudo-mot considéré comme un mot) :

- Effet de difficulté des stimuli (*Figure 37*) : on observe en moyenne un taux d'erreur supérieur de 3,30 points lorsque le stimulus présenté est difficile, $F(1, 35) = 39,62, p < ,001, \eta^2_p = ,53$. Cet effet est confirmé par une correction H-F ($p < ,001$).

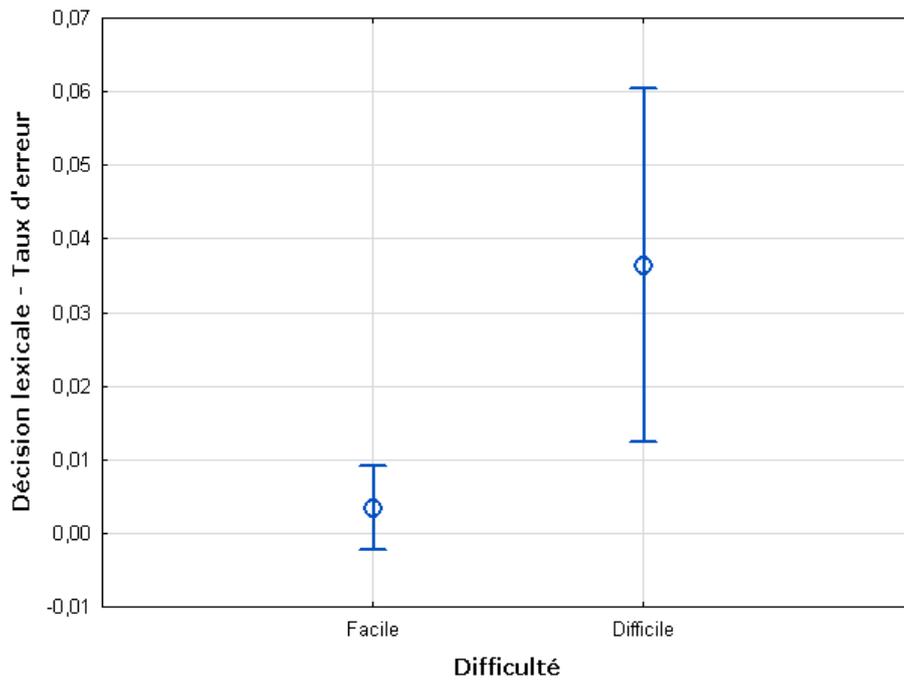


Figure 37. Taux d'erreur à la décision lexicale en fonction de la difficulté.

- Effet du support (*Figure 38*) : l'analyse ne montre pas d'effet du support sur le taux d'erreur, $F(2, 70) = 1,12, p = ,333$.

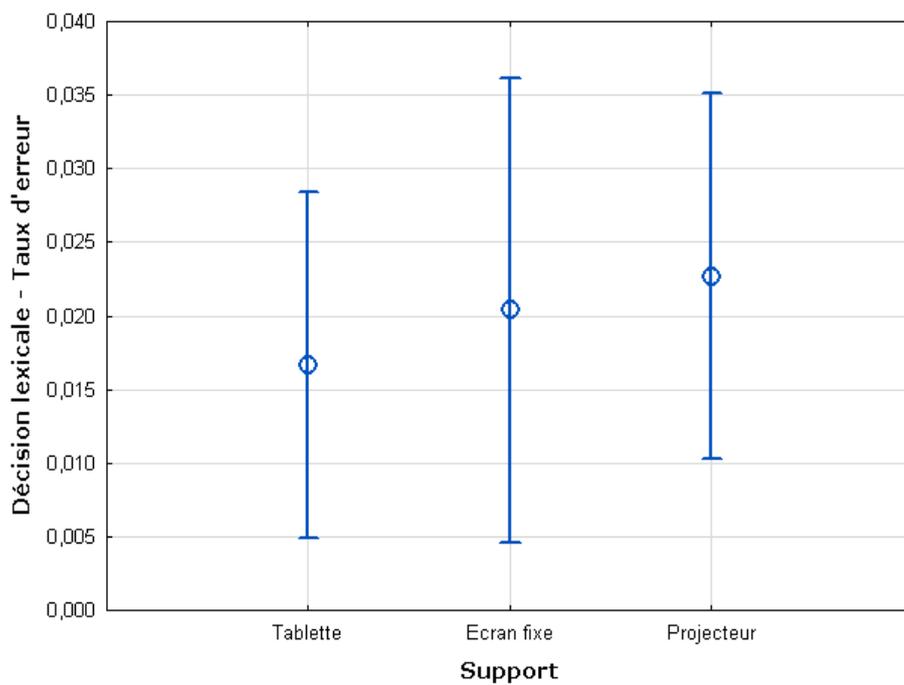


Figure 38. Taux d'erreur à la décision lexicale en fonction du support. L'effet support n'est pas significatif.

L'analyse ne montre pas d'autres différences (pour tous les autres effets et interactions, $p > ,050$).

8.4.2.4 Durée réelle d'affichage des phrases

Nous avons réalisé une ANOVA à un facteur pour comparer les différences entre les durées d'affichages réelles et les durées demandées par l'algorithme pour le test de vitesse de lecture. Cette analyse montre un effet du support (voir *Figure 39*), les supports ne présentent donc pas la même erreur, $F(2, 1554) = 7,13$, $p < ,001$, $\eta^2_p = ,01$. Les erreurs moyennes sont de -6,72 ms (écart-type : 24,93) pour la tablette, -1,98 ms (écart-type : 23,96) pour l'écran fixe et 6,10 ms (écart-type : 88,01) pour le projecteur. Un test HSD de Tukey montre des différences entre le projecteur et la tablette ($p < ,001$) et entre le projecteur et l'écran fixe ($p = ,039$). Le test ne montre pas de différence entre l'erreur de la tablette et celle de l'écran fixe ($p = ,334$). Les erreurs moyennes sont donc très faibles. Ces erreurs sont comptées en différences de durée d'affichage par phrase entière. Or, dans le test de vitesse de lecture, l'unité manipulée est la durée d'affichage par mot. Les phrases comptent 5, 6 ou 7 mots dans ce test. Nous avons donc divisé les erreurs par 6 afin de calculer l'impact de l'erreur sur ce test. Il faut donc modifier les résultats au test de -1,11 ms pour la tablette, de -0,33 ms pour l'écran fixe et de 1,02 ms pour le projecteur. Ces modifications sont donc extrêmement faibles, mais nous les avons tout de même appliquées pour corriger les résultats du test de vitesse de lecture.

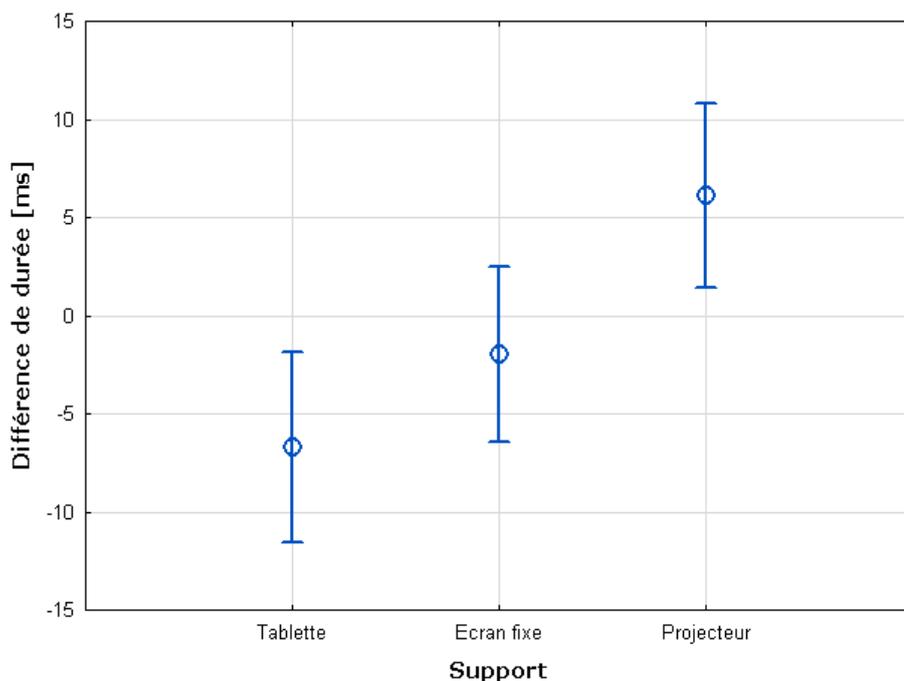


Figure 39. Différence de durée d'affichage entre la durée d'affichage demandée (au niveau du logiciel d'affichage) et la durée réelle.

8.4.2.5 Vitesse de lecture

Une ANOVA à mesures répétées a été conduite : elle ne permet pas de détecter un effet du support sur la durée d'affichage minimale corrigée par mot (et donc sur la vitesse de lecture), $F(2, 70) = 1,68, p = ,193$ (voir *Figure 40*).

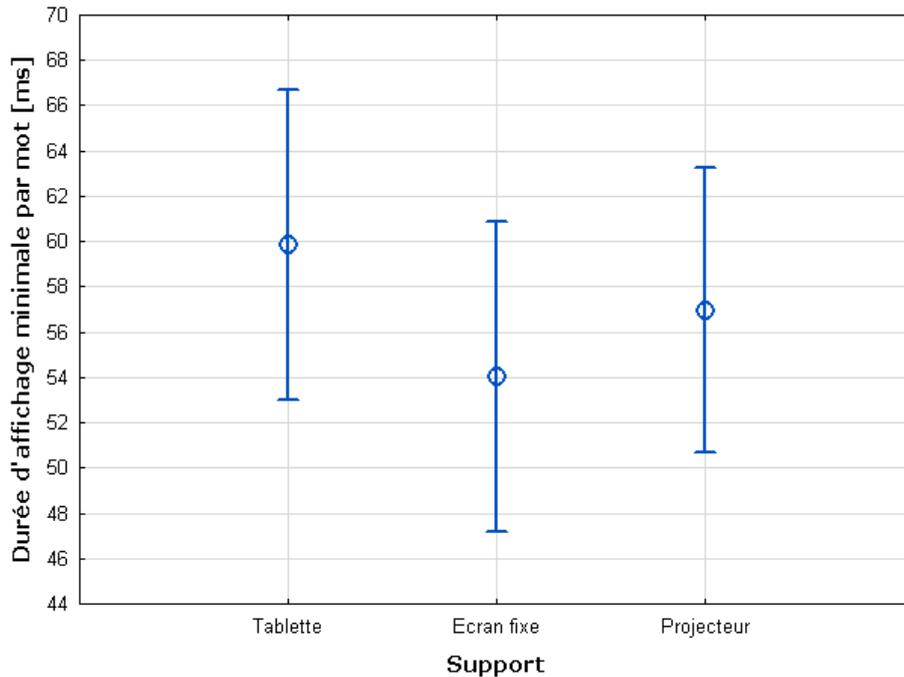


Figure 40. Durée d'affichage minimale par mot pour pouvoir lire une phrase en fonction du support. L'effet support n'est pas significatif.

8.4.3 Compréhension et posture

8.4.3.1 Test de compréhension

8.4.3.1.1 Durées de lecture

Pour les facteurs expérimentaux, une ANOVA à mesures répétées détecte plusieurs effets sur les durées de lecture par mot :

- Effet de la difficulté (voir *Figure 41*) : la difficulté des textes a un effet sur la durée de lecture, $F(2, 70) = 3,69, p = ,030, \eta^2_p = ,01$. Ce résultat est confirmé avec une correction H-F ($p = ,030$). Un test de Tukey montre que la durée de lecture par mot est plus courte pour un texte difficile que pour un texte facile (0,018 s, $p = ,049$). Le test ne permet pas de statuer sur une lecture plus rapide sur les textes moyens en comparaison aux textes faciles (moyenne

de la différence : 0,017 s, $p = ,062$). Le test ne montre pas de différence de durée de lecture entre les textes moyens et les textes difficiles ($p = ,994$).

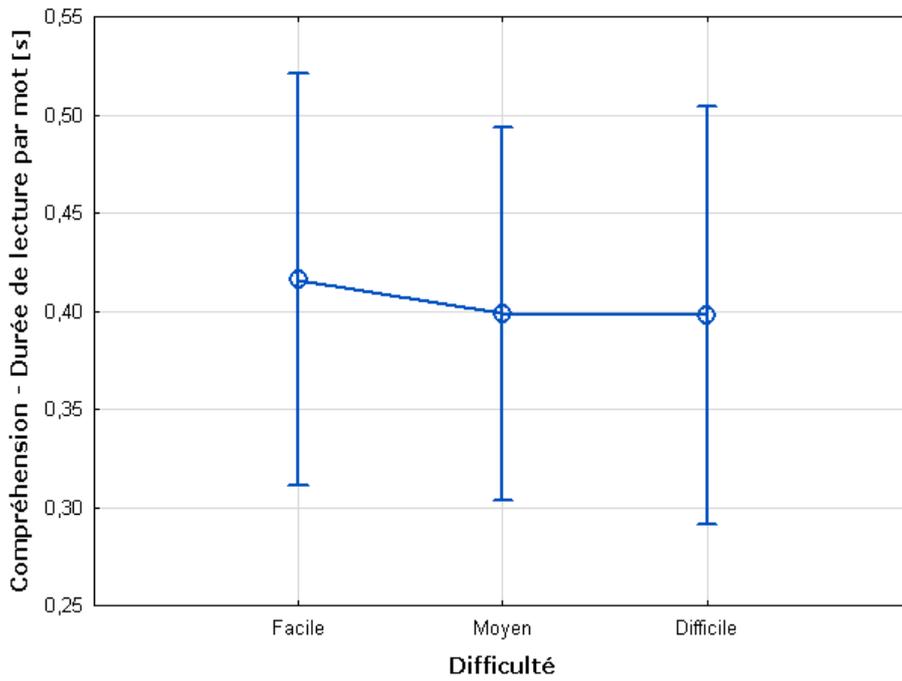


Figure 41. Durée de lecture par mot au test de compréhension en fonction de la difficulté.

- Effet du support (voir Figure 42) : l'ANOVA ne nous permet pas d'observer un effet du support sur la durée de lecture, $F(2, 70) = 1,80, p = ,173$.

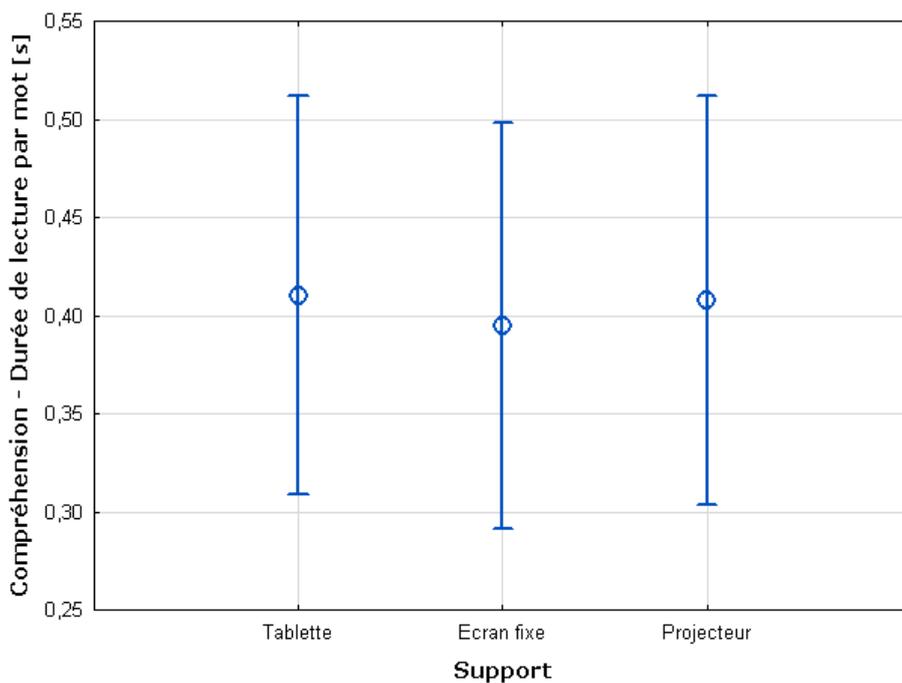


Figure 42. Durée de lecture par mot au test de compréhension en fonction du support. L'effet du support n'est pas significatif.

L'ANOVA ne montre ni d'effet pour la mise en forme, $F(1, 35) = 3,45, p = ,072$, ni d'effets d'interaction : pour chaque interaction testée, $p > ,200$.

8.4.3.1.2 Erreurs de mémorisation

Nous avons également réalisé une ANOVA à mesures répétées sur le taux d'erreurs de mémorisation.

- Effet de la difficulté (voir Figure 43) : la difficulté des textes a un effet sur les erreurs de mémorisation, $F(2, 70) = 18,24, p < ,001, \eta^2_p = ,34$. Ce résultat est confirmé avec une correction H-F ($p < ,001$). Un test de Tukey montre que le nombre d'erreurs de mémorisation augmente avec la difficulté (comparaison facile vs. moyen : $p = ,005$; comparaison facile vs. difficile : $p < ,001$; comparaison moyen vs. difficile : $p = ,018$).

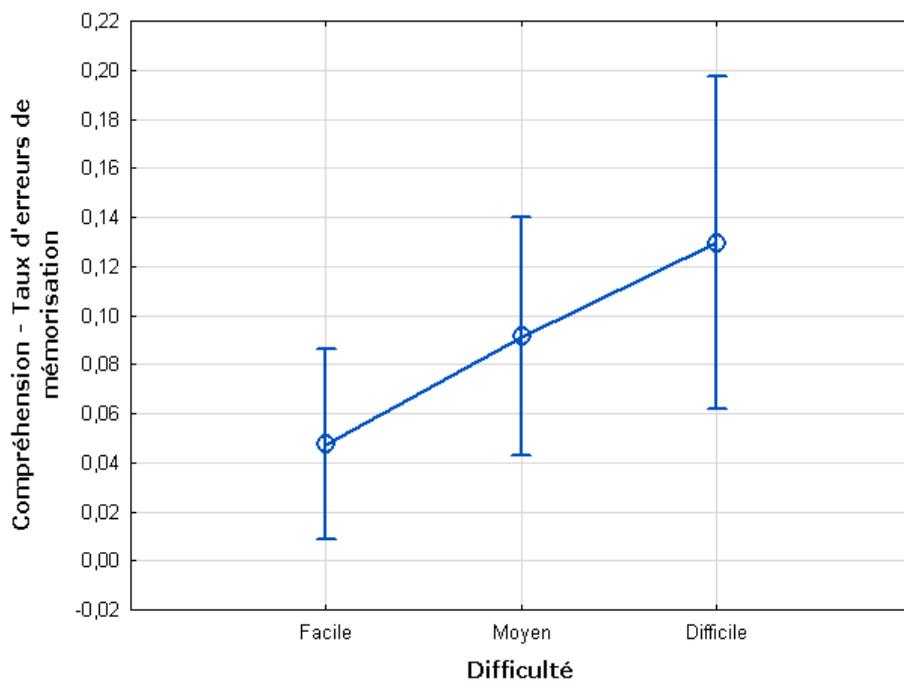


Figure 43. Taux d'erreurs de mémorisation en fonction de la difficulté des textes.

- Effet du support (voir Figure 44) : l'analyse ne révèle pas de différence entre les supports en termes d'erreur de mémorisation, $F(2, 70) > 0,01, p = ,997$.

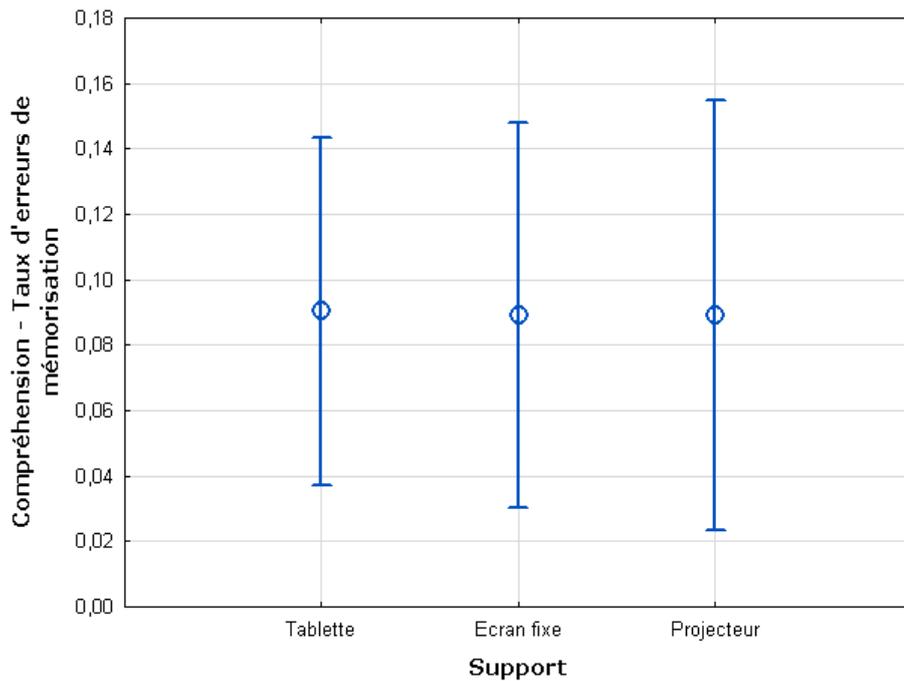


Figure 44. Taux d'erreurs de mémorisation en fonction du support. L'effet du support n'est pas significatif.

L'ANOVA n'a pas pu montrer d'effet de la mise en forme des textes, ni des interactions (dans tous les cas, $p > ,050$).

8.4.3.1.3 Erreurs d'inférences

De la même manière, une ANOVA a été réalisée pour comparer les erreurs d'inférences :

- Effet de la difficulté (voir *Figure 45*) : on observe un effet de la difficulté sur les erreurs d'inférence, $F(2, 70) = 30,55, p < ,001, \eta^2_p = ,47$. Ce résultat se confirme avec une correction H-F ($p < ,001$). Un test de Tukey montre que le nombre d'erreurs d'inférence augmente avec la difficulté (comparaison facile vs. moyen : $p = ,018$; comparaison facile vs. difficile : $p < ,001$; comparaison moyen vs. difficile : $p < ,001$).

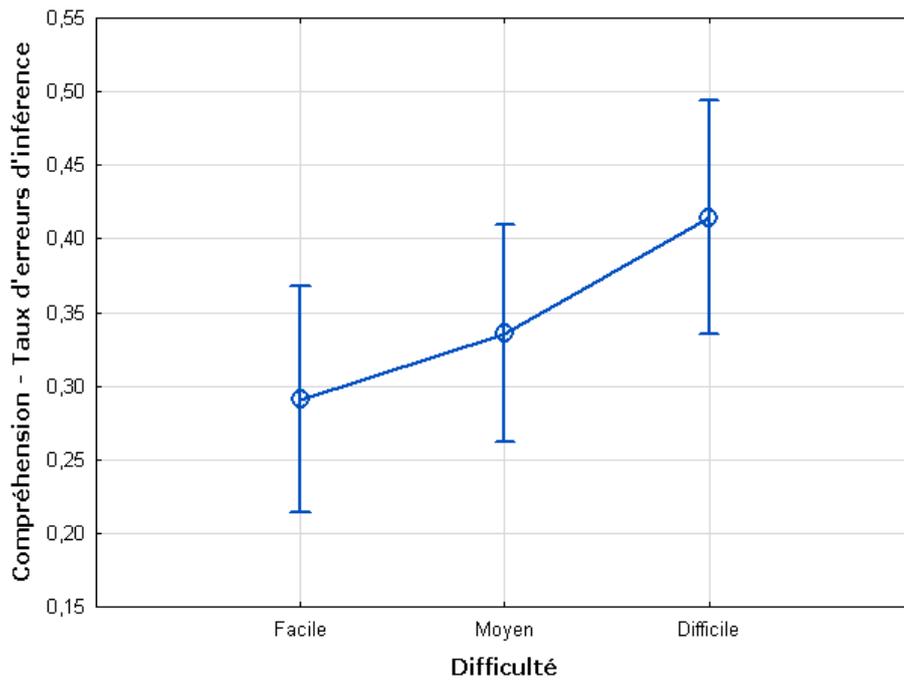


Figure 45. Taux d'erreurs d'inférence en fonction de la difficulté du texte.

- Effet du support (voir *Figure 46*) : l'analyse ne montre pas d'effet du support sur l'erreur d'inférence, $F(2, 70) = 1,09, p = ,343$.

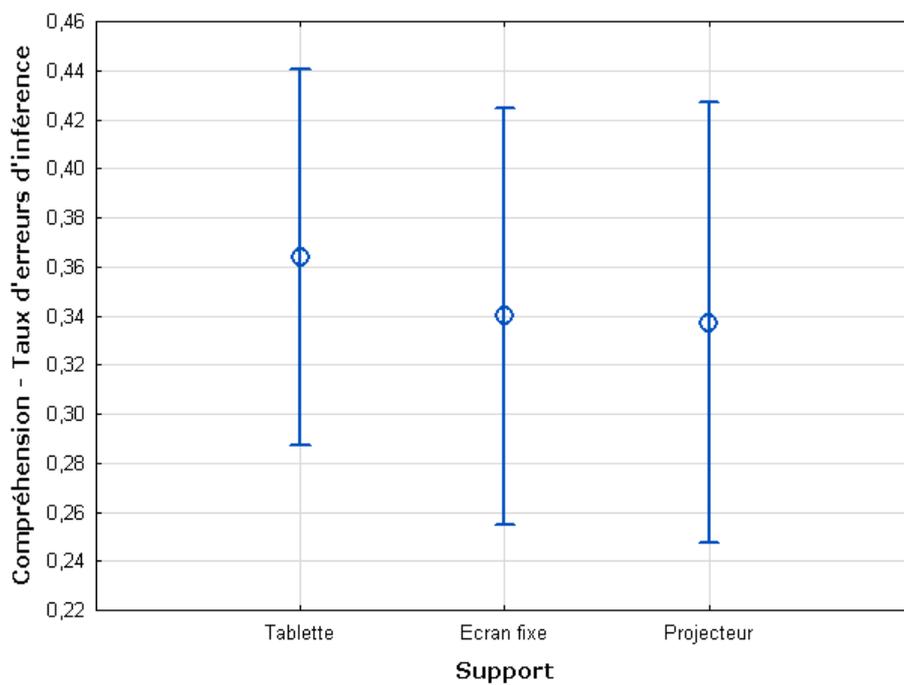


Figure 46. Taux d'erreurs d'inférence en fonction du support. L'effet n'est pas significatif.

L'ANOVA n'a pas pu montrer d'effet de la mise en forme des textes, ni des interactions (dans tous les cas, $p > ,200$).

8.4.3.1.4 Durées de réponse

Une dernière ANOVA à mesures répétées a été construite pour le test de compréhension : celle-ci concerne les durées de réponse :

- Effet de la difficulté (voir *Figure 47*) : l'analyse montre que la difficulté a un effet sur les durées de réponse aux questions, $F(2, 70) = 16,42$, $p < ,001$, $\eta^2_p = ,32$. Ce résultat se confirme avec la correction H-F ($p < ,001$). Un test post-hoc HSD de Tukey montre que la durée de réponse est plus longue pour les textes difficiles par rapport aux textes faciles et aux textes moyens (dans les deux cas, $p < ,001$). Ce test ne détecte pas de différence pour la comparaison entre textes faciles et textes moyens ($p = ,856$).

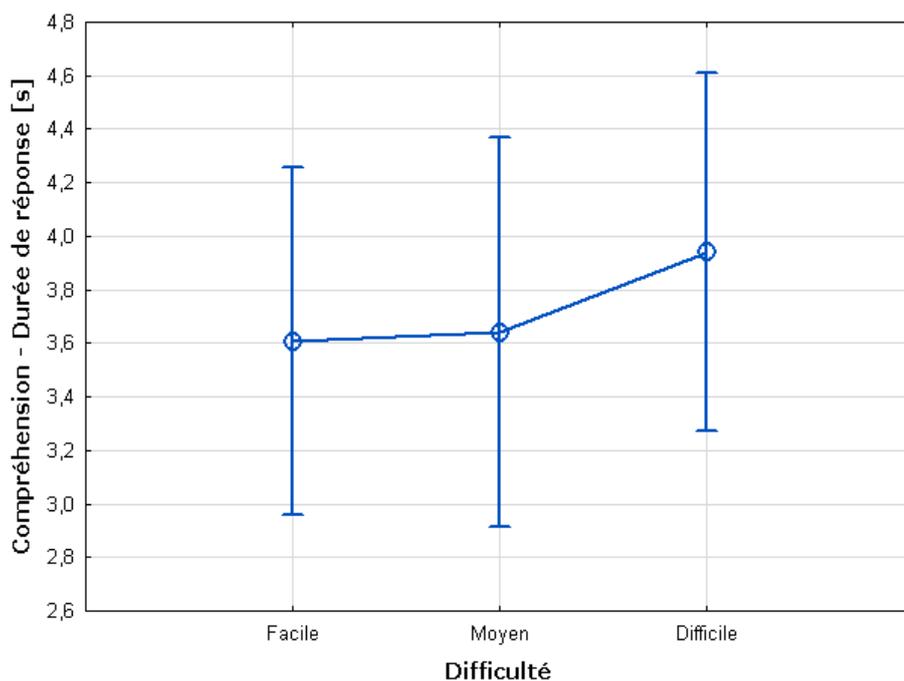


Figure 47. Durée de réponse moyenne aux questions de compréhension en fonction de la difficulté du texte.

- Effet du support (voir *Figure 48*) : l'analyse ne montre pas d'effet du support, $F(2, 70) = 0,51$, $p = 0,601$.

L'analyse ne permet pas non plus de détecter un effet de la mise en forme ni d'effets d'interaction (dans tous les cas, $p > ,050$).

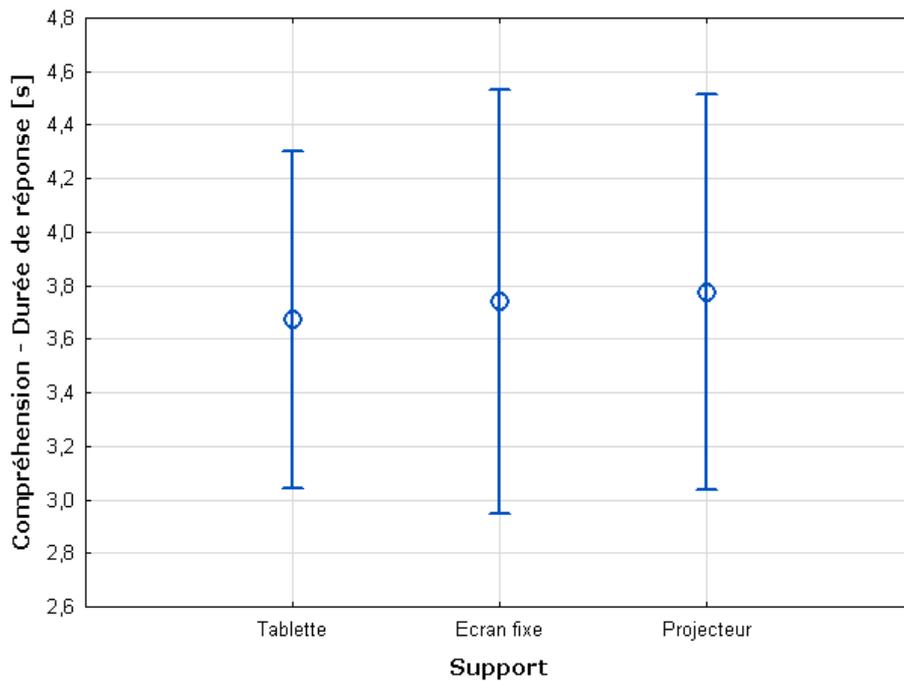


Figure 48. Durée de réponse moyenne aux questions de compréhension en fonction du support. L'effet n'est pas significatif.

8.4.3.2 Posture

8.4.3.2.1 Distance moyenne à l'écran

La distance à l'écran a été enregistrée durant la lecture des textes du test de compréhension. Une première ANOVA à mesures répétées a été réalisée sur la distance moyenne à l'écran pendant la lecture d'un texte :

- Effet de la difficulté (voir *Figure 49*) : l'ANOVA ne permet pas de statuer sur un éventuel effet de la difficulté sur la distance de lecture, $F(2, 70) = 2,73, p = ,072$.

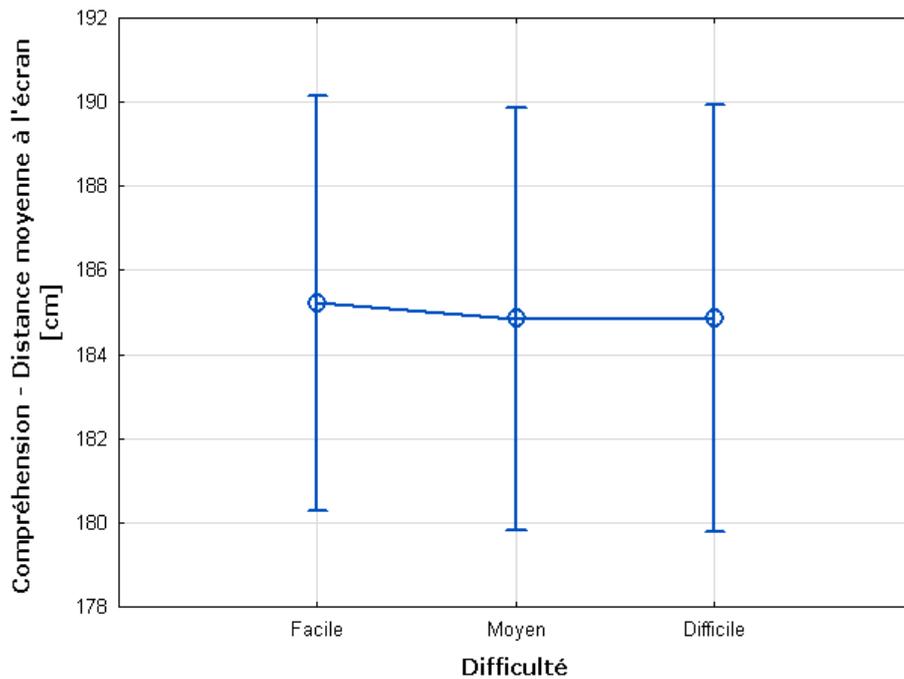


Figure 49. Distance moyenne à l'écran pendant la lecture des textes en fonction de leur difficulté ($p = ,072$).

- Effet du support (voir *Figure 50*) : la distance à l'écran est dépendante du support, $F(2, 70) = 38516,30, p < ,001, \eta^2_p \approx 1$. Ce résultat est validé par la correction H-F : $p < ,001$. Un test de Tukey montre que la distance est la plus courte pour la tablette (en moyenne : 41,81 cm), que la distance à l'écran fixe est plus importante (en moyenne : 62,22 cm), et que la distance est la plus grande pour le vidéoprojecteur (en moyenne : 450,89 cm), pour chaque comparaison, $p < ,001$.

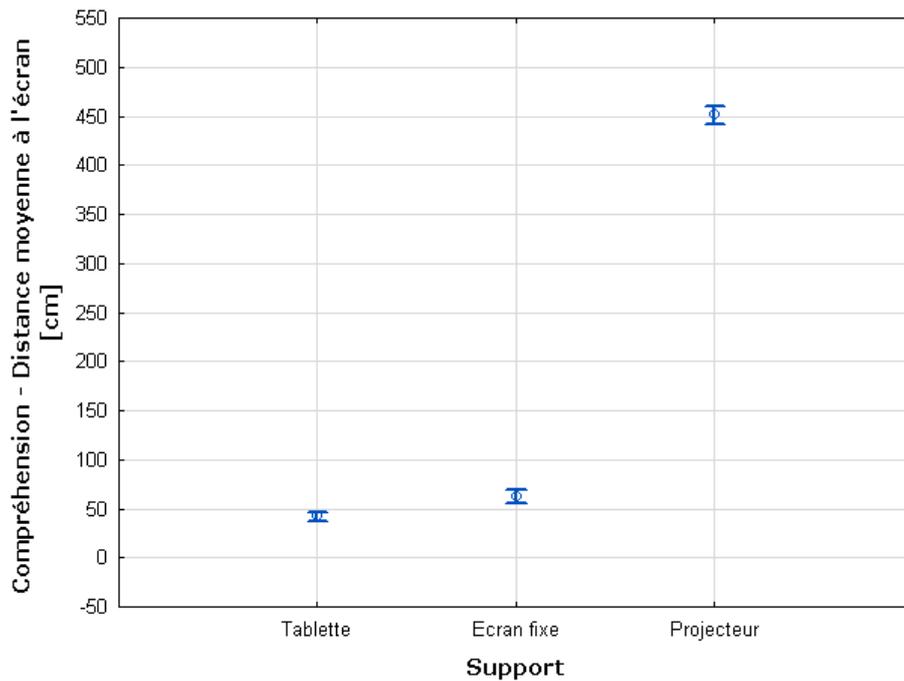


Figure 50. Distance moyenne à l'écran pendant la lecture des textes en fonction du support.

L'analyse ne montre ni effet de la mise en forme, ni effet d'interaction (pour chaque cas, $p > ,100$).

8.4.3.2.2 Ecart-type de la distance à l'écran

Une ANOVA à mesures répétées a été réalisée pour observer l'influence des facteurs expérimentaux sur l'écart-type de la distance à l'écran.

- Effet de la difficulté (voir *Figure 51*) : la difficulté a un effet sur la variabilité de la distance à l'écran, $F(2, 70) = 3,60$, $p = ,032$, $\eta^2_p = ,09$. Un test HSD de Tukey montre que l'écart-type est plus grand pour la condition difficile que pour la condition moyenne (écart de 0,25 cm en moyenne, $p = ,039$). Le test ne détecte pas de différence entre la condition facile et la condition facile ($p = ,097$) et entre les conditions facile et moyenne ($p = ,916$).

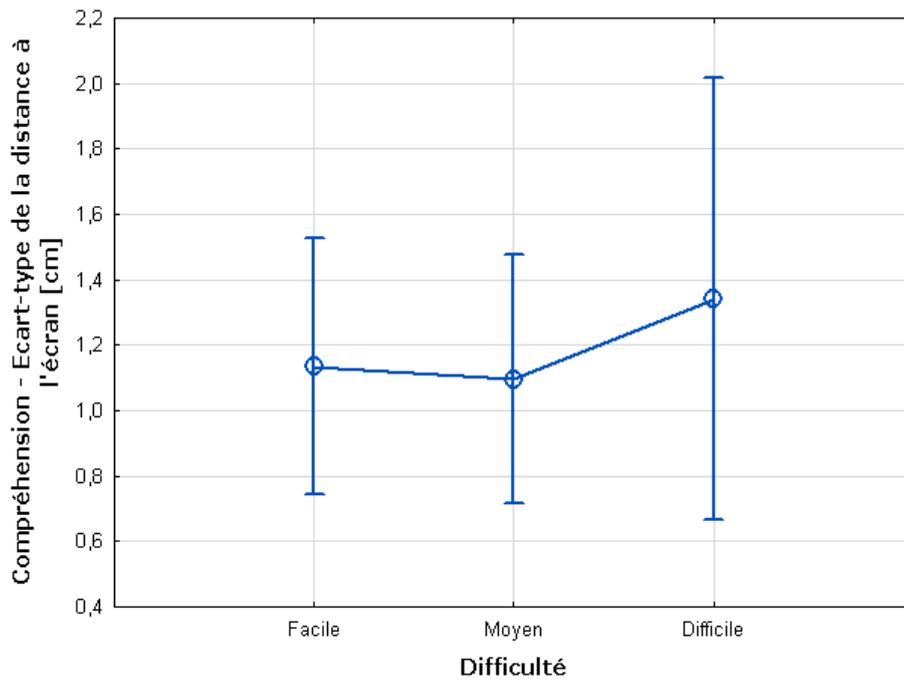


Figure 51. Écart-type de la distance à l'écran pendant la lecture des textes en fonction de la difficulté.

- Effet du support (voir *Figure 52*) : l'analyse ne montre pas d'effet du support sur l'écart-type de la distance à l'écran, $F(2, 70) = 0,69, p = ,506$.

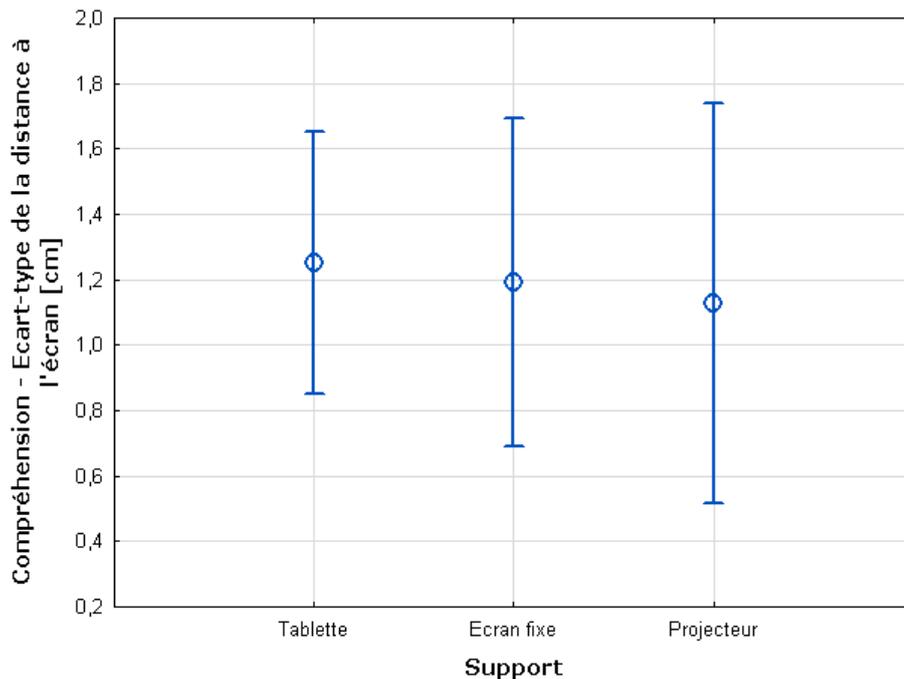


Figure 52. Ecart-type de la distance à l'écran pendant la lecture des textes en fonction de la difficulté.

L'ANOVA ne permet pas de montrer d'effet de la mise en forme ou d'interaction (pour tous, $p > ,050$).

8.4.3.3 Tâche R-K

Nous avons réalisé une ANOVA à mesures répétées pour tester notamment l'effet du support sur les indices de discrimination A' et de biais B'' associés aux variables obtenues suite aux passations de la tâche R-K. L'analyse ne montre pas d'effet du support sur les performances, $F(2, 70) = 0,28, p = ,753$. Les sujets n'ont pas montré une modalité de mémoire plutôt que l'autre : pour la comparaison des résultats « Remember » contre « Know », $F(1, 35) > 0,01, p = ,978$. Les valeurs des indices de discrimination A' sont plus élevées que celles des indices de biais B'' , $F(1, 35) = 3778,20, p < ,001, \eta^2_p = ,99$. Ce résultat est confirmé par une correction H-F ($p < ,001$). La difficulté des textes affecte les réponses : $F(2, 70) = 7,09, p = ,002, \eta^2_p = ,17$, confirmé par une correction H-F ($p = ,002$). Un test post-hoc de Tukey montre que les moins bonnes performances sont observées pour les textes moyens ($p = ,008$ par rapport aux textes faciles et $p = ,003$ par rapport aux textes difficiles). On n'observe pas de différence entre textes faciles et textes difficiles ($p = ,959$). L'analyse ne révèle pas d'effets d'interaction : pour tous, $p > ,150$.

8.5 Modèle de calcul du QLE

Le but de cette section est de décrire la procédure que nous avons mise en place pour dériver des indicateurs généraux de performance de lecture électronique à partir des tests décrits précédemment. Ces indicateurs seront regroupés de manière à obtenir un seul indicateur moyen qui décrit la performance globale de lecture sur un support électronique donné : le « Quotient de Lecture Electronique ».

8.5.1 Calcul des indicateurs pour les tests

Nous calculons tout d'abord un indicateur de performance pour chaque test.

- **Visibilité** : le test de caractérisation psychométrique aboutit à une mesure, l'aire sous la courbe de caractérisation. Plus cette aire est importante, plus la performance du sujet a été faible. Nous calculons la moyenne ainsi que l'écart-type des valeurs d'aires sous la courbe pour les 3 (supports) * 36 (sujets) mesures. A partir de cette moyenne et de cet écart-type, nous calculons la valeur centrée-réduite de chaque mesure (i.e, normalisation des données : on soustrait la mesure brute à la moyenne, et on divise cette valeur par l'écart-type). Enfin, on rapporte la valeur opposée de cette valeur centrée-réduite, afin que les valeurs les plus élevées soient associées aux meilleures performances.
- **Lisibilité** : la lisibilité est évaluée à partir de trois tests (lettres, mots, phrases). Nous construisons donc trois indicateurs. Le test d'identification de lettres aboutit à deux mesures de performance : la durée de réponse et le taux d'erreurs. Plus ces mesures sont faibles, plus la performance du sujet est élevée. Les moyennes et écart-types sont calculés pour les deux mesures. On calcule alors la valeur-centrée réduite des mesures. Pour obtenir l'indicateur, on prend la valeur opposée de la moyenne des deux valeurs centrées-réduites. Le même procédé est appliqué pour obtenir les indicateurs du test de décision lexicale. Le test de vitesse de lecture aboutit à une mesure : la durée minimale d'affichage par mot pour que la phrase puisse être lue. On prend l'opposé de la valeur centrée-réduite de cette mesure comme indicateur.
- **Compréhension**: pour obtenir un indicateur de performance de compréhension, nous utilisons les mesures de 3 variables : l'erreur de mémorisation, l'erreur d'inférence et la durée de réponse. Pour ces trois variables, plus la valeur est

faible, plus la performance a été importante. Chacune de ces variables a été centrée-réduite sur l'ensemble des 36 (sujets) * 3 (supports) * 6 (textes) = 648 mesures. Pour obtenir un indicateur par texte lu, on prend l'opposé de la moyenne des 3 valeurs centrées-réduites. Pour obtenir l'indicateur général qui correspond à un support donné, on calcule la moyenne des indicateurs des 6 textes lus sur ce support.

- Posture / attention : nous avons également mesuré un aspect postural pendant la lecture des textes : la distance à l'écran. Cette mesure est réalisée afin d'avoir un estimateur de la charge attentionnelle (voir 7.2 *L'effet de la lecture électronique sur la posture*) : nous avons montré qu'une augmentation de la charge attentionnelle est liée à un rapprochement à l'écran. Il ne serait pas judicieux de centrer-réduire directement les données de distance, car une très large partie de la dispersion de ces données est due au support (la distance est de l'ordre de 40 cm pour la tablette, alors qu'elle est supérieure à 4 m pour le projecteur). Nous avons donc commencé par calculer les distances moyennes pour chaque combinaison sujet / support.

$$\bar{d}_{ij} = \frac{1}{6} \sum_{k=1}^6 d_{ijk}$$

Avec d_{ijk} la distance moyenne à l'écran pour le texte k , sur le support j par le sujet i . Ces valeurs renseignent une distance de base étant donné le comportement postural d'un sujet par rapport à un support donné. Afin de déterminer un indice de variabilité qui n'est pas associé à cette distance de base, on calcule d'abord pour chaque valeur d_{ijk} son écart à la distance de base. L'écart-type de ces distances corrigées est alors calculé. Pour obtenir un indicateur de charge attentionnelle issu de la posture pour chaque texte lu, on calcule l'opposé de la valeur centrée-réduite des valeurs de distance corrigées. Nous avons donc un indicateur d'attention par texte. Afin de construire un indicateur de posture pour un support et un sujet donnés, nous cherchons à voir dans quelle mesure le sujet a pu diminuer sa distance à l'écran (augmenter sa charge attentionnelle) avec la difficulté des textes. Nous représentons ainsi les 6 indicateurs de charge attentionnelle en fonction des 3 valeurs de difficulté. On attribue respectivement aux textes faciles, moyens et difficiles, les valeurs 0, 1 et 2.

A l'aide de la méthode des moindres carrés, on cherche la fonction linéaire qui décrit au mieux les données (deux exemples sont donnés dans la Figure 53).

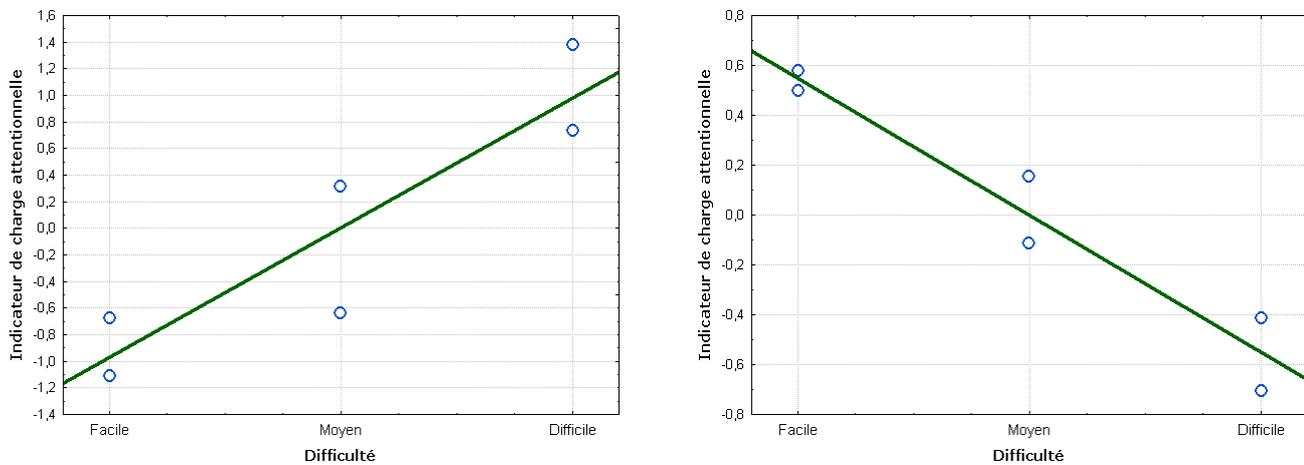


Figure 53. Deux graphes représentant les indicateurs de charge attentionnelle en fonction de la difficulté des textes. Pour le graphe de gauche (sujet 35, projecteur), la charge augmente avec la difficulté, le coefficient directeur de la droite de régression est positif. On observe l'inverse dans le graphe de droite (sujet 34, projecteur), la charge semble diminuer avec la difficulté (désengagement), le coefficient directeur est négatif.

L'indicateur retenu pour décrire l'évolution de la charge attentionnelle est le coefficient directeur de la droite de régression : lorsqu'il est positif, le sujet a pu augmenter son niveau d'attention avec la difficulté, alors qu'une valeur négative serait plutôt liée à un désengagement attentionnel.

Le procédé permet donc d'obtenir 6 indicateurs qui quantifient la performance à différents niveaux de la lecture électronique. Les moyennes de ces 6 indicateurs sont représentées dans un graphique (voir la *Figure 54*).

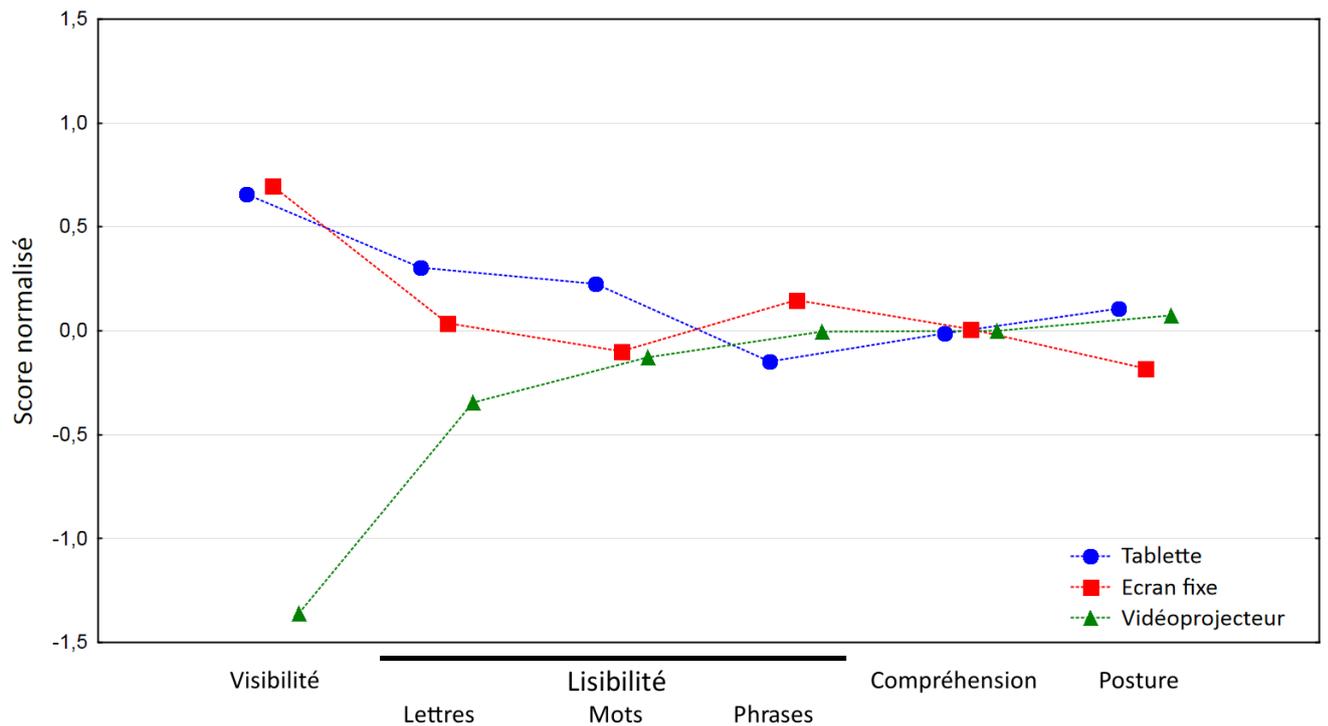


Figure 54. Représentation graphique des valeurs obtenues pour chaque test et pour chaque support afin de calculer le score de Quotient de Lecture Electronique.

Afin de définir un indicateur unique de performance par support et par sujet, nous calculons la moyenne pondérée des 6 indicateurs, de manière à ce que chaque dimension classique de la lecture électronique ait le même poids (compréhension et posture sont regroupées car elles portent sur le même test). Ces valeurs sont ensuite multipliées par 100, et on leur ajoute 100 :

$$QLE_{ij} = 100 * \frac{(V_{ij} + \frac{1}{3} * (L_{lettres_{ij}} + L_{mots_{ij}} + L_{phrases_{ij}}) + \frac{1}{2} * (C_{ij} + P_{ij}))}{3} + 100$$

Avec i le sujet, j le support, V l'indicateur de performance pour la visibilité, $L_{lettres}$ pour l'identification de lettres, L_{mots} pour la décision lexicale, $L_{phrases}$ pour la vitesse de lecture, C pour la performance de compréhension et P pour la posture associée.

Le Quotient de Lecture Electronique associé à chaque support est calculé en prenant la moyenne des indicateurs uniques de performance de chaque sujet obtenus avec ce support. Les valeurs de QLE obtenues pour chaque support sont donc (voir *Figure 55*) :

- Tablette, QLE = 127,81 (écart-type : 30,66) ;
- Ecran fixe, QLE = 121,44 (écart-type : 28,61) ;
- Vidéoprojecteur, QLE = 50,75 (écart-type : 29,25).

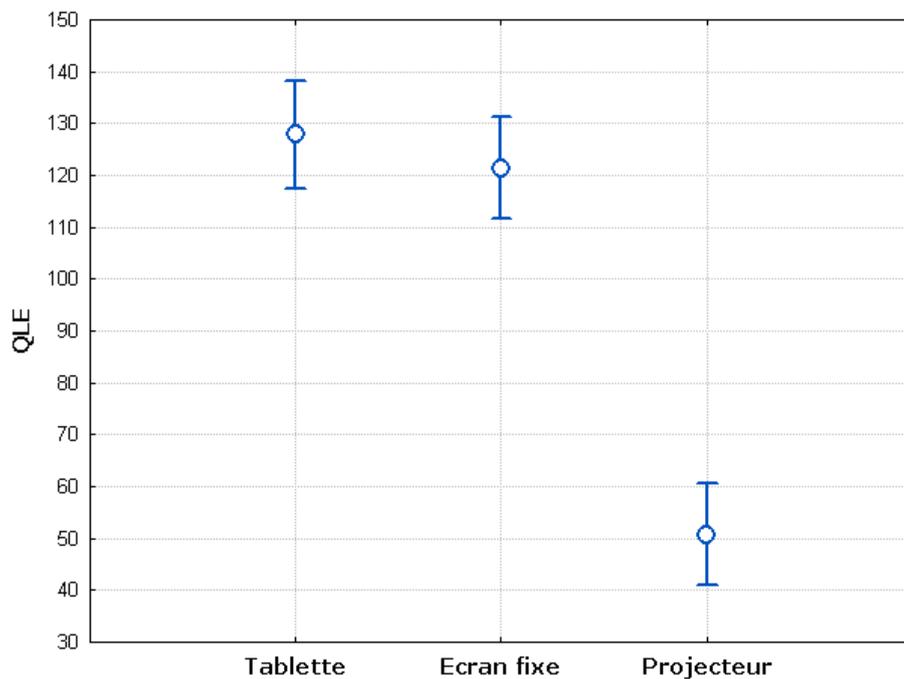


Figure 55. Valeurs finales du QLE pour chaque support.

8.6 Résultats et implications

8.6.1 Visibilité

L'analyse a montré que les niveaux de gris sont bien moins distingués sur le projecteur que sur les deux autres supports, alors que la tablette et l'écran fixe sont relativement similaires sur ce point. Ce résultat est confirmé par les mesures de luminance effectuées : le contraste est beaucoup moins élevé sur le projecteur que sur les deux autres supports, qui ont des valeurs de luminance proches. Cette observation peut s'expliquer du fait du contexte d'affichage : le projecteur affichait directement sur un mur, et non pas sur un écran, et la pièce était éclairée relativement uniformément, ce qui a impacté l'affichage des niveaux de gris

foncés. Le test permet donc d'identifier les différences entre dispositifs d'affichage électroniques.

8.6.2 Lisibilité

8.6.2.1 Temps de réponse de base

La mesure du temps de réponse de base n'est pas un test à part entière de la batterie du QLE, car elle n'est pas liée à une dimension de la lecture électronique, mais est utilisée comme correctif pour deux tests. La mesure porte plutôt sur l'interaction, que nous avons décidé de ne pas évaluer dans le contexte de ces travaux. Néanmoins, nous observons des différences qui peuvent se montrer d'intérêt. La réponse donnée sur l'écran tactile de la tablette est plus longue que sur les autres supports. Mais le dispositif d'interaction en lui-même ne peut pas expliquer entièrement des différences : à dispositif égal, les réponses sont légèrement plus rapides sur l'écran fixe que sur le vidéoprojecteur. Le positionnement dans l'espace du dispositif d'affichage pourrait donc également affecter l'interaction.

8.6.2.2 Identification de lettres

Les facteurs expérimentaux propres au test montrent des effets attendus : les sujets ont une performance moindre lorsqu'ils doivent identifier une lettre dans un groupe de lettres visuellement proches ; c'est aussi le cas lorsque la lettre-cible n'est pas présente dans le groupe de lettres (le sujet doit s'assurer pour chaque lettre affichée qu'il ne s'agit pas de la cible pour pouvoir donner sa réponse).

Ce test montre une performance détériorée pour le projecteur : cette différence peut s'expliquer du fait que ce support était celui qui présentait la résolution et le contraste les plus faibles. C'est sur la tablette que les durées de réponses corrigées sont les plus courtes : il s'agit du support le mieux contrasté et qui présente la meilleure résolution. Le ralentissement de la réponse du à la présence de lettres voisines n'est pas le même sur chaque support : il est le plus important pour le vidéoprojecteur. La faible résolution de ce support pourrait expliquer la difficulté à distinguer une lettre cible de lettres qui lui sont ressemblantes. Mais cet effet ne peut pas s'expliquer uniquement par la résolution puisqu'il est légèrement plus important sur la tablette que sur l'écran fixe. On pourrait alors évoquer des facteurs oculomoteurs tels que la nécessité d'augmenter la convergence et l'accommodation pour voir nettement les caractères affichés sur la tablette (qui est plus proche que l'écran fixe).

8.6.2.3 Décision lexicale

Pour ce test, nous observons qu'un mot est catégorisé plus rapidement qu'un pseudo-mot. Cet effet ne porte que sur les temps de réponse, et pas sur le taux d'erreurs. Habituellement, afin de moduler la difficulté des stimuli présentés, les chercheurs utilisent des mots avec des valeurs de fréquence lexicale distinctes. Néanmoins, cette technique ne permet pas de contrôler la difficulté des pseudo-mots, car, par définition, leur fréquence lexicale est nulle. Nous nous sommes donc basés sur des valeurs obtenues empiriquement (Ferrand et al., 2010) pour définir les conditions faciles et difficiles à la fois pour les mots et les pseudo-mots. Le facteur difficulté impacte à la fois les temps de réponse et les taux d'erreurs des sujets.

Les performances en termes de temps de réponse sont les meilleures pour la tablette, mais nous n'observons pas d'effet du support sur le taux d'erreur. Les réponses plus rapides sur tablette pourraient encore une fois s'expliquer par sa meilleure résolution.

8.6.2.4 Vitesse de lecture

Ce test n'a pas permis de montrer de différences liées au support sur les durées d'affichage nécessaire pour lire des phrases (et donc sur les vitesses de lecture). Deux pistes d'explication peuvent être soulevées pour expliquer ce résultat :

- le test ne serait pas suffisamment discriminant pour détecter une différence,
- les supports testés n'auraient pas eu d'effet sur la vitesse de lecture. Cela signifierait que pour une tâche de lecture de phrases entières, les contrastes et résolutions qui nous avons utilisés n'affectent pas la performance. Des travaux antérieurs ont montré que l'effet du contraste sur la vitesse de lecture n'est pas linéaire (Legge et al., 1987) : la vitesse de lecture reste stable pour les contrastes élevés, et diminue à partir d'un contraste critique (dépendant de la taille des caractères et des caractéristiques du sujet). Les différences de contraste que nous avons observé sur les supports testés ne seraient donc pas suffisantes pour impacter la vitesse de lecture. Le même type d'hypothèse pourrait être avancé pour expliquer l'absence d'effet de la résolution.

8.6.3 Compréhension et posture

8.6.3.1 Compréhension

8.6.3.1.1 Facteurs expérimentaux propres au test

Le but du générateur automatique de textes présentés dans le *Chapitre 6 : L'évaluation de la compréhension* était de proposer des textes dont la difficulté serait contrôlée, afin d'évaluer la compréhension pour plusieurs niveaux de difficulté. Globalement, le facteur difficulté que nous avons mis en œuvre montre effectivement un impact sur la compréhension. La durée de lecture par mot est différente en fonction de la difficulté du texte (textes faciles lus légèrement moins rapidement), mais cet effet est très faible. Les erreurs de mémorisation et d'inférences augmentent avec la difficulté du texte. Enfin, la durée de réponse aux questions est plus longue pour les textes difficiles. Ces résultats valident donc l'attribution des niveaux de difficulté aux textes, et leur effet sur les performances de compréhension.

La mise en page des textes formait un second facteur expérimental : les idées principales du texte étaient soit présentées dans des paragraphes distincts, soit affichées en un seul bloc. Aucun effet n'est ressorti à l'analyse, quelle que soit la variable mesurée, pour ce facteur. Ce résultat pourrait être dû à la structure assez systématique des textes : les idées principales sont toujours présentées les unes après les autres dans le texte ; les lecteurs n'ont peut-être pas eu besoin d'indices visuels tels que la mise en forme pour structurer les différents éléments.

8.6.3.1.2 Effets des supports

Les analyses ne permettent pas de montrer d'effets du support sur la performance de compréhension. Ce résultat n'est pas étonnant dans la mesure où les textes ont été présentés exactement de la même manière sur chaque support. Les textes ont été affichés dans une interface minimaliste : seul le titre, le texte et un bouton pour passer aux questions étaient présents. Nous avons expliqué dans *2.1.3 La compréhension* que c'est l'interface qui peut perturber la compréhension du contenu textuel, à cause des liens hypertextes (Baccino, 2004; Miall & Dobson, 2006), ou alors l'améliorer, par exemple grâce à la présence d'images adaptées au texte (Carney & Levin, 2002). Les résultats de notre expérimentation semblent confirmer le rôle de l'interface, en montrant, qu'à interface égale, le support n'engendre pas d'effet sur la compréhension.

Les résultats de l'analyse montrent donc que nous sommes parvenus à générer des textes pour lesquels la difficulté est contrôlée, comme l'attestent les baisses de performances associées au niveau de difficulté. Néanmoins, il resterait à vérifier la capacité de ce test à faire ressortir des différences dues aux changements d'interfaces (complexité, présence de liens, présence d'illustrations,...).

8.6.3.2 Posture

8.6.3.2.1 Facteurs expérimentaux propres au test

L'analyse ne permet pas de conclure à un effet de la difficulté des textes sur la distance à l'écran. Néanmoins, nous ne pouvons pas non plus totalement exclure un éventuel effet ($p = ,07$). Tout comme dans l'expérimentation présentée dans 7.2 *L'effet de la lecture électronique sur la posture*, il est possible que la hausse de la difficulté n'ait pas systématiquement entraîné une hausse de la charge attentionnelle, ce qui pourrait expliquer que la posture ne soit pas directement liée au niveau de difficulté. En outre, dans l'expérience nous avons vu que la hausse de la difficulté entraînait une hausse de la variabilité de la distance à l'écran. C'est également le cas pour le test de compréhension du QLE, ce qui contredit les résultats de la littérature (Qiu & Helbig, 2012).

8.6.3.2.2 Effets des supports

C'est sans surprise que le facteur support entraîne un effet sur la distance à l'écran. La plus grande distance est observée pour le vidéoprojecteur, cela s'explique par la disposition du matériel (projection sur le mur et sujet assis au bureau). La moyenne de distance obtenue pour l'écran fixe est comparable à une valeur obtenue lorsque les sujets sont libres de se placer comme ils le souhaitent face à l'ordinateur (Jaschinski, 2002). Pour la tablette, la distance moyenne est supérieure à nos attentes. Nous nous étions basés sur les résultats de l'expérimentation présentée dans 7.2 *L'effet de la lecture électronique sur la posture*. La distance plus importante pourrait être due au poids plus important de la tablette utilisée pour le test du QLE (290 g contre 185 g) et à la taille légèrement plus grande (7'' contre 6'') ; la taille et/ou le poids des tablette peut entraîner une distance plus grande à l'écran (Maniwa et al., 2013).

Nous n'avons pas relevé d'effet du support sur la variabilité de la distance à l'écran.

8.6.3.3 Tâche R-K

L'analyse des résultats à la tâche R-K ne révèle pas d'effet du support ou d'effets d'interaction. Le seul effet lié aux conditions expérimentales est l'effet de la difficulté. Pourtant, les performances ne sont pas linéairement liées au niveau de difficulté. Les meilleures réponses surviennent pour les textes faciles mais aussi pour les textes difficiles. Or, ces textes étaient ceux présentés en début et en fin de texte. On pourrait donc soupçonner un effet de primauté et un effet de récence, qui ont permis une meilleure rétention des titres des textes en mémoire, indépendamment de leur contenu.

8.6.4 Quotient de lecture

L'indicateur QLE calculé pour chaque support semble plutôt bien représenter les différences entre supports²⁶. Le vidéoprojecteur était le support auquel la résolution et le contraste les plus faibles étaient associés. Ce dispositif affichait directement sur un mur, avec un éclairage ambiant, et non pas sur un écran de projection avec les lumières éteintes, ce qui altère grandement les qualités visuelles de l'affichage. Cet affichage détérioré est associé à une valeur de QLE très faible (50,75). La tablette et l'écran fixe, dont les contrastes et résolutions étaient meilleurs, et relativement proches aboutissent à des valeurs de QLE plus resserrées (respectivement 127,81 et 121,44). Le score de QLE est le plus élevé pour la tablette, qui était le support disposant de la meilleure résolution et du contraste le plus élevé. Le découpage du QLE en 6 sous-indicateurs permet d'observer à quels niveaux les différences sont présentes ; sur la *Figure 54*, on peut observer que les indicateurs des trois supports sont très proches en termes de compréhension, mais différent par exemple au niveau de la visibilité.

²⁶ Néanmoins, l'utilisation des valeurs centrées-réduites est une technique parmi d'autres. L'utilisation d'une Analyse en Composantes Principales aurait permis de tenir compte de l'inter-corrélation entre les résultats obtenus aux différentes dimensions de la lecture électronique.

Chapitre 9 : Discussion générale - Conclusion

9.1 Tester la lecture électronique : dimensions classiques et posture

Ce travail de thèse a consisté dans un premier temps à définir les mécanismes mis en jeu lors de la lecture électronique. Il s'agit d'un processus complexe pour deux raisons. D'une part, la lecture électronique résulte de l'interaction entre le lecteur, le dispositif d'affichage et le contenu affiché (le texte, notamment). Ces trois éléments ont des caractéristiques propres, qui vont influencer le processus global. D'autre part, la lecture au sens général est une activité complexe : le lecteur met en œuvre une série de traitements, qui agissent à différents niveaux : de la perception des caractères, jusqu'à la compréhension.

Il a donc été nécessaire d'adopter un formalisme pour adresser la complexité de la lecture sur support électronique. Nous avons ainsi considéré les trois dimensions classiques de la lecture électronique (Baccino, 2004) : la visibilité, la lisibilité et la compréhension. L'émergence des appareils portatifs, la diversité des périphériques d'interaction et les différents formats d'écrans nous ont également poussé à prendre en compte une dimension supplémentaire : l'aspect postural de la lecture électronique.

Le but de ce travail était de proposer une méthode générale d'évaluation de la performance associée à la lecture sur un support électronique donné. Cet objectif a impliqué la mise au point de tests permettant d'effectuer des mesures de performance de manière répétée. Nous avons pris le parti de nous inspirer de la psychophysique, qui permet de mesurer des seuils de performance en fonction des variations des stimuli. Un élément essentiel de cette approche est de maîtriser ces stimuli. Si ce point est relativement facile à réaliser pour évaluer les dimensions de bas niveaux de la lecture électronique (identification de formes lumineuses, de lettres, voire de mots), la tâche s'avère bien moins évidente pour les dimensions de haut niveau qui impliquent des stimuli plus complexes tels que les phrases et les textes. En outre, la dimension posturale de la lecture électronique est encore mal connue, et il a été nécessaire de réaliser des études afin de choisir des marqueurs posturaux pertinents pour l'évaluer.

Pour évaluer la visibilité, nous avons adapté un test de Gille et al. (2004) afin de réaliser une caractérisation lumineuse de l'écran en se basant sur les réponses des sujets à une tâche de détection de contraste. Pour améliorer la capacité à effectuer des mesures répétées, nous avons amélioré la procédure en utilisant une tâche de choix forcé à la place de la tâche

de sélection, et nous l'avons rendue plus rapide à l'aide d'une procédure adaptative bayésienne (Kontsevich & Tyler, 1999).

Trois tests ont été construits pour évaluer la lisibilité à trois niveaux linguistiques : la lettre, le mot et la phrase. Les tests permettant d'évaluer le traitement des lettres et des mots sont des tests relativement classiques. Le test d'identification de lettres prend la forme d'une tâche où le lecteur doit déterminer si une lettre-cible est présente ou non dans un ensemble de lettres. Ce test permet également de rechercher des effets d'interaction avec la présence de perturbateurs : des lettres visuellement proches de la lettre-cible, cette proximité de lettres étant basée sur les données expérimentales de Courrieu et al. (2004). Le test évaluant le traitement des mots est une tâche de décision lexicale : le lecteur doit déterminer si le mot affiché est un mot de la langue ou non (pseudo-mot). Contrairement au procédé généralement utilisé dans ce type de tâche, nous n'avons pas utilisé la fréquence lexicale comme facteur expérimental car cette variable ne peut être obtenue que pour les mots et non pas pour les pseudo-mots. Nous avons construit nos conditions à l'aide des données expérimentales du *French Lexicon Project* (Ferrand et al., 2010), qui associe à la fois aux mots et aux pseudo-mots, des temps de réponse moyens et des taux d'erreurs. Le dernier test qui a été développé est une procédure de mesure de la vitesse de lecture maximale, largement inspirée des travaux de Crossland et al. (2008). Des phrases sont affichées pendant une durée limitée, la lecture effective est contrôlée, et une procédure adaptative permet de converger vers la durée d'affichage minimale correspondant à la vitesse maximale de lecture. Un algorithme génère automatiquement un grand nombre de phrases équivalentes et attribue à ces phrases des valeurs de vérité qui permettent de contrôler la lecture effective, ce qui permet au test de pouvoir évaluer de manière répétée la vitesse de lecture. Les mesures issues de ce nouveau test ont été comparées à celles obtenues avec des planches classiques (Legge et al., 1989; Sénécal et al., 2006), et ces mesures sont corrélées entre elles. Bien entendu, cette nouvelle méthode de mesure présente un intérêt pour l'évaluation de la lecture électronique. Mais elle peut servir dans un contexte plus large, notamment pour mesurer les performances de patients atteints de déficits de la vision (Rubin, 2013).

Pour évaluer la compréhension, il est nécessaire de disposer de textes. La compréhension étant un processus qui implique l'acquisition et le traitement de l'information à différents niveaux (Kintsch & van Dijk, 1978), nous avons proposé d'évaluer conjointement la mémorisation des éléments de surface du texte et les inférences réalisées par le lecteur. Dans ce but, nous avons construit un générateur automatique des textes. Il produit des textes

décrivant des concepts irréels (villes et animaux), ce qui permet de limiter les effets des connaissances préalables des lecteurs ; celles-ci pouvant modifier la performance de compréhension (Kendeou & van den Broek, 2007). L'information contenue dans les textes générés est contrôlée (détails, ainsi qu'idées générales), tout comme les niveaux de difficulté et de cohérence. Le générateur permet donc d'obtenir plusieurs ensembles équivalents de textes cohérents, pour lesquels la compréhension du lecteur peut être évaluée en fonction de la difficulté textuelle.

Au-delà de ces trois dimensions classiques, ce travail s'est aussi attaché à explorer le lien entre posture et lecture électronique. Les expérimentations intégrant ces deux domaines portent sur l'ergonomie physique (Straker et al., 2008; Young et al., 2012), sur des aspects visuels (Albin & McLoone, 2014; Maniwa et al., 2013) ou cognitifs (Le Bigot & Jaschinski, 2011; Le Bigot & Grosjean, 2012). Nous avons réalisé deux études pour questionner la relation entre posture et lecture électronique. Une première étude a porté sur l'effet des nouvelles postures entraînées par l'utilisation de tablettes. Elle a montré qu'une inclinaison latérale de la tête n'impactait pas les performances de lecture : les lecteurs s'adaptent à ces situations. Une piste d'explication a été proposée en montrant la présence de cyclotorsions dans ces conditions : les yeux tournent de manière à compenser en partie la différence d'angle d'inclinaison entre la tête et la tablette. Une seconde étude a montré l'effet de la difficulté du texte sur la distance à l'écran. Lorsque les lecteurs lisent des textes difficiles, ils ont tendance à se rapprocher de l'écran. Mais ce rapprochement est surtout lié à charge attentionnelle qui est négativement corrélée à la distance à l'écran. Si le lien entre complexité et distance à l'écran était déjà connu pour d'autres tâches (Balaban et al., 2004; Qiu & Helbig, 2012), il s'agit, à notre connaissance, d'un résultat nouveau pour la lecture sur support électronique. L'expérimentation ne permet pas de déterminer les causes de ce phénomène, nous avons proposé plusieurs hypothèses dans la section 7.2 *L'effet de la lecture électronique sur la posture*. Ce résultat peut être pertinent pour les chercheurs qui étudient l'attention en général : ils pourraient utiliser la distance à l'écran comme un indicateur de la charge attentionnelle. Mais c'est aussi un résultat dont il faut tenir compte vis-à-vis d'autres marqueurs de l'attention : par exemple, le diamètre pupillaire augmente avec le niveau d'attention, mais le rapprochement à l'écran peut entraîner une augmentation de l'accommodation associée à une diminution de ce diamètre. Il est donc important de tenir compte de ce phénomène pour les études où les sujets sont libres de déplacer leur tête. Nous avons utilisé ce résultat en

complément du test de compréhension : la distance à l'écran est mesurée durant la lecture des textes afin de disposer d'un marqueur du niveau d'attention.

9.2 Une approche multidimensionnelle : la batterie de tests du Quotient de Lecture Electronique

Une première phase du travail de thèse a donc consisté à mettre au point ces tests pour évaluer la lecture électronique. La principale nouveauté apportée par ce travail tient dans le fait que nous avons ensuite évalué simultanément des dispositifs d'affichage pour toutes les dimensions de la lecture électronique. Les tests ont été regroupés au sein d'une batterie, afin de quantifier les performances de lecture à tous les niveaux. Le grand nombre de mesure amène cependant un inconvénient : les données sont multiples et l'interprétation des résultats peut être complexe. Nous avons donc également construit des indicateurs pour quantifier synthétiquement les performances associées à chaque dimension. Enfin, ces indicateurs ont été regroupés sous un indicateur global, le Quotient de Lecture Electronique.

La batterie de tests a été appliquée sur trois dispositifs d'affichage électronique : un écran d'ordinateur fixe, un vidéoprojecteur et une tablette. La tablette et l'écran fixe étaient assez similaires en termes de contraste et de définition, avec un léger avantage pour la tablette. Ces paramètres étaient dégradés pour le projecteur, qui affichait avec une faible définition dans un environnement très éclairé, qui a fortement impacté le contraste. Le test de visibilité a montré une performance moindre pour le vidéoprojecteur. La tablette était associée aux meilleures performances pour le traitement des lettres et des mots. Pour les tests de plus haut niveau (vitesse de lecture et compréhension), nous n'avons plus observé de différences. Ce résultat pourrait signifier qu'à partir d'une certaine qualité visuelle, l'affichage n'affecte plus la vitesse de lecture et la compréhension. Pour la vitesse de lecture, cela pourrait s'expliquer au regard des travaux de Legge, Pelli, Rubin & Schleske (1985) et de Legge et al. (1987). Ces auteurs ont montré que des paramètres associés à la qualité visuelle des lettres tels que leur taille et leur contraste ne sont pas linéairement associés à la vitesse de lecture. La vitesse augmente avec la taille et le contraste jusqu'à un certain seuil. Au-delà de cette valeur de seuil, la vitesse de lecture reste stable, indépendamment des valeurs de taille/contraste²⁷. Peut-être sommes-nous dans le même cas pour les supports testés : les supports sont différents

²⁷ La vitesse de lecture chute à nouveau lorsque les caractères sont de très grande taille.

en termes de qualité d'affichage, comme l'attestent les tests de bas niveau, mais la qualité d'affichage n'est pas suffisamment dégradée pour entraîner une baisse de la vitesse de lecture. Le même type d'interprétation pourrait être avancé pour expliquer les résultats observés au test de compréhension : les lecteurs étaient capables de percevoir visuellement les textes avec suffisamment d'efficacité, ils ont alors pu comprendre les textes, indépendamment des variations de qualités visuelles des dispositifs d'affichage.

Globalement, les différences observées se répercutent sur les indicateurs de performances, et les scores de QLE reprennent le classement de performance observé (les scores sont pour le projecteur : 50,75, pour l'écran fixe : 121,44 et pour la tablette : 127,81).

9.3 Limites et ouvertures

Nous avons montré que la batterie de tests du QLE permet de différencier les supports pour la visibilité et la lisibilité. Pour les supports testés, nous n'avons pas observé de différence de compréhension. Ce résultat n'est pas surprenant dans la mesure où les études questionnant l'effet de la lecture électronique sur la compréhension adressent majoritairement l'impact des interfaces. Ces études observent l'effet des liens hypertextes (Salmerón & García, 2011), de la mise en page (Schmid & Baccino, 2003), des distracteurs tels que la publicité (Simola et al., 2011),... Dans notre test, l'interface était minimaliste : le texte était présenté en noir sur fond blanc, seul un titre et un bouton l'accompagnaient. Il pourrait être pertinent d'intégrer nos tests dans des interfaces. Les tests seraient facilement transposables dans d'autres contenus : au lieu d'afficher des lettres, des mots, des phrases ou des textes sur un fond blanc, il faudrait les afficher dans un contexte d'interface réelle. Les tests permettraient alors de pouvoir comparer la performance de lecture entre plusieurs interfaces. La comparaison entre supports pourrait être améliorée en les testant avec des interfaces représentatives (par exemple : dans un navigateur web pour un ordinateur fixe, dans une application pour un smartphone, dans un diaporama pour un vidéoprojecteur, etc.).

En étant appliquée à l'évaluation d'une interface, la batterie de tests du QLE pourrait aider à appréhender son « utilisabilité ». Dans le domaine de l'ergonomie des interfaces homme/machine, une interface est dite utilisable notamment si elle permet de réaliser une tâche avec efficacité et confort. L'évaluation des performances de lecture à l'aide du QLE pourrait s'ajouter aux outils existants qui permettent de mesurer l'utilisabilité (Baccino, Bellino, & Colombi, 2005).

La batterie de test a été prévue pour évaluer un support électronique à un instant donné. Les tests ne sont pas faits pour évaluer la performance de lecture à plus long terme. Or, la lecture sur écran pendant plusieurs heures est parfois liée à un inconfort physiologique : la fatigue visuelle (Rempel et al., 2007; Benedetto et al., 2013). Ce phénomène n'était pas l'objet de notre étude, mais il pourrait être intéressant de voir si des mesures répétées sur le même support, à plusieurs moments de la journée pourraient montrer une baisse de la performance pour des utilisateurs souffrant de fatigue visuelle.

Nos travaux ont porté sur la lecture électronique. Mais l'utilisation des outils numériques n'entraîne pas uniquement un changement pour la lecture, elle impacte aussi le processus d'écriture. L'écriture électronique est souvent réalisée à l'aide d'un clavier, qu'il soit physique ou virtuel²⁸. L'écriture sur clavier est extrêmement différente de l'écriture sur papier : si l'écriture manuscrite implique le tracé de courbes et de traits, l'écriture au clavier consiste à appuyer successivement sur des touches. Or, l'identification de lettres (et donc potentiellement la lecture) activerait des régions du cortex pré-moteur, impliquées dans l'écriture manuscrite (Longcamp et al., 2003). Le changement de modalité d'écriture engendré par les dispositifs numériques, c'est-à-dire le passage du crayon au clavier, pourrait donc perturber les capacités de lecture : un effet négatif de l'utilisation du clavier a été montré chez des enfants pour l'apprentissage de mots nouveaux (Velay, Longcamp, & Zerbato-Poudou, 2004). Les tests que nous avons développés durant ce travail de thèse permettent d'évaluer un support par rapport à ses aspects visuels, ainsi qu'aux aspects posturaux liés à l'activité de lecture, mais pas par rapport à ses modalités d'écriture. Les scores de QLE obtenus découlent de la qualité de lecture à un temps donné, néanmoins ils ne tiennent pas compte de l'effet à long terme du support sur la lecture. Au vu des travaux montrant le lien entre lecture et écriture, il serait peut-être plus pertinent dans certains cas d'utiliser un support ayant une résolution et des contrastes moindres, mais qui permette à l'utilisateur de réaliser les gestes d'écriture (à l'aide d'un stylet par exemple), plutôt qu'un support ayant des performances d'affichage élevées, mais qui oblige à écrire avec un clavier. Le QLE est donc une mesure qui permet de comparer et de classer des supports, néanmoins le classement obtenu n'est valable qu'en termes de performances de lecture, et n'intègre pas certaines autres dimensions.

²⁸ L'écriture peut aussi être réalisée à l'aide d'un stylet numérique, ou même à partir d'interfaces utilisant les mouvements des yeux, mais ces méthodes sont moins courantes.

Dans cette thèse, nous avons considéré les supports de lecture électroniques comme des dispositifs d'affichages. Or, l'affichage n'est pas le seul mode d'interaction des supports, ceux-ci intègrent également souvent des systèmes sonores, des caméras, de l'éclairage. Les personnes souffrant de déficits de la vision peuvent exploiter ces autres possibilités offertes par les dispositifs mobiles pour simplifier leur lecture (Crossland, Silva, & Macedo, 2014) : la synthèse vocale peut transcrire le texte, la caméra intégrée au téléphone peut servir de loupe, le flash de cette caméra peut éclairer un texte papier et le rendre plus contrasté... Pour ces personnes, la lecture électronique prend un tout autre sens dans la mesure où elle permet de lire une information qui n'aurait pas été lisible sans l'apport de ces technologies.

Une autre limite des travaux présentés tient dans le fait que les tests du QLE sont destinés à évaluer la lecture chez le sujet adulte. Ces tests ne sont pas prévus pour l'évaluation de la performance chez l'enfant. Pourtant, les enfants sont de plus en plus exposés aux outils numériques, et leurs capacités à y acquérir de l'information est un enjeu important, elles sont notamment évaluées par le programme d'évaluation international PISA (Thomson & De Bortoli, 2012). Le choix des supports numériques utilisés dans le contexte d'enseignement est également questionné (Diarra et al., 2011; Megalakaki, Aparicio, Porion, Pasqualotti, & Baccino, 2015). Il serait donc pertinent d'utiliser le QLE pour évaluer la lecture électronique chez l'enfant, mais les tests nécessiteraient des adaptations. Ces ajustements pourraient se traduire par la prise en compte de l'âge d'acquisition des mots pour le test de décision lexicale ; des bases décrivant l'âge auquel les mots sont appris sont disponibles et pourraient permettre d'adapter cette tâche à l'âge des lecteurs (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Pour le test de vitesse de lecture, la tâche de vérification consiste à déterminer si une phrase décrivant une catégorisation est vraie ou fausse (par exemple : « Tous les chiens sont des animaux »). Les capacités de catégorisation s'acquièrent au cours du développement, et ne peuvent être utilisées avant un certain âge (Fouquet & Megalakaki, 2013), d'autres phrases et une autre tâche de vérification devraient être construites pour permettre aux plus jeunes lecteurs de passer ce test. Si les enfants sont suffisamment âgés pour effectuer une tâche de catégorisation, les catégories devraient être construites et évaluées a priori par ces lecteurs, pour s'assurer que la catégorisation est acceptable. Il pourrait être pertinent de calculer un degré d'appartenance, c'est-à-dire mesurer l'accord qu'ont les sujets avec la catégorisation proposée (Léger, Boumlak, & Tijus, 2008), et de ne sélectionner les catégories pour lesquelles ce degré est important. Pour le test de compréhension, des textes plus simples devraient être générés.

Une dernière limite tient dans le fait qu'un support électronique ne peut être évalué indépendamment d'un contexte d'utilisation. Dans notre expérimentation présentée dans le *Chapitre 8 : Le quotient de lecture électronique*, l'ambiance lumineuse et la disposition des supports étaient contrôlées. Dans d'autres situations, les résultats n'auraient sans doute pas été les mêmes : par exemple la tablette aurait pu être évaluée en extérieur en temps ensoleillé, ou le vidéoprojecteur aurait pu être évalué en projetant sur un écran dans une salle sans éclairage. Les résultats de QLE sont donc à considérer relativement au contexte, et ne peuvent être pris comme des mesures absolues de la qualité d'affichage des dispositifs. Il serait donc nécessaire de réaliser ces mesures dans des contextes représentatifs des conditions d'utilisation « normales », mais comment définir de telles conditions ?

En dépit de ces limites, nos travaux apportent des outils permettant d'évaluer la lecture électronique de manière répétée et la batterie de test du QLE intègre les dimensions de cette activité. L'utilisation de ces nouveaux outils pourrait permettre de rendre les études évaluant les effets de la lecture électronique plus facilement comparables entre elles. Nous espérons donc que ce travail de thèse pourra permettre des mesures plus précises, qui permettront de faire progresser les connaissances dans le domaine de la lecture électronique.

Liste des références bibliographiques

- Adam, J. M. (2005). La notion de typologie de textes en didactique du français: une notion dépassée ? *Recherches: Revue de didactique et de pédagogie du français*, 42, 11-23.
- Ahn, S. J., Legge, G. E., & Luebker, A. (1995). Printed cards for measuring low-vision reading speed. *Vision Research*, 35(13), 1939-1944.
- Albin, T. J., & McLoone, H. E. (2014). The effect of tablet tilt angle on users' preferences, postures, and performance. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 47(2), 207-211.
- Allard, R., Renaud, J., Molinatti, S., & Faubert, J. (2013). Contrast sensitivity, healthy aging and noise. *Vision Research*, 92(0), 47-52.
- Association française de normalisation. (2002). *Ergonomie des postes et lieux de travail: Tome 1, Principes généraux et conception des postes de travail*. Ergonomie des postes et lieux de travail AFNOR.
- Bababekova, Y., Rosenfield, M., Hue, J. E., & Huang, R. R. (2011). Font size and viewing distance of handheld smart phones. *Optometry & Vision Science*, 88(7), 795-797.
- Baccino, T. (2004). *La lecture électronique* Presses universitaires de Grenoble.
- Baccino, T., Bellino, C., & Colombi, T. (2005). *Mesure de l'utilisabilité des interfaces*.
- Balaban, C. D., Cohn, J., Redfern, M. S., Prinkey, J., Stripling, R., & Hoffer, M. (2004). Postural control as a probe for cognitive state: Exploiting human information processing to enhance performance. *International Journal of Human-Computer Interaction*, 17(2), 275-286.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human perception and performance*, 10(3), 340.
- Ban, H., & Yamamoto, H. (2013). A non-device-specific approach to display characterization based on linear, nonlinear, and hybrid search algorithms. *Journal of vision*, 13(6), 20.
- Barten, P. G. (1999). *Contrast sensitivity of the human eye and its effects on image quality* SPIE press.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276.
- Benedetto, S., Draï-Zerbib, V., Pedrotti, M., Tissier, G., & Baccino, T. (2013). E-Readers and Visual Fatigue. *PloS one*, 8(12), e83676.
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour*, 14(3), 199-208.

- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476), 307-310.
- Bonin, P., Méot, A., Aubert, L., Malardier, N., Niedenthal, P., & Capelle-Toczek, M. C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *L'année psychologique*, 655-694.
- Bouma, H. (1971). Visual recognition of isolated lower-case letters. *Vision Research*, 11(5), 459-474.
- Bucher, U., Heitger, F., Mast, F., & Bischof, N. (1990). A novel automatic procedure for measuring ocular counterrolling: A computer analytical method to determine the eye's roll angle while subjects work on perceptual tasks. *Behavior Research Methods, Instruments, & Computers*, 22(5), 433-439.
- Cail, F., & Aptel, M. (2006). Facteurs de risque pour le membre supérieur dans le travail sur écran: Synthèse bibliographique. *Le Travail humain*, 69(3), 229-268.
- Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14(1), 5-26.
- Carver, R. P. (1970). Effect of a "chunked" typography on reading rate and comprehension. *Journal of Applied Psychology*, 54(3), 288.
- Carver, R. P. (1990). *Reading rate: A review of research and theory* Academic Press.
- Chandrakumar, M., Hirji, Z., Goltz, H. C., Mirabella, G., Blakeman, A. W., Colpa, L. et al. (2010). Effects of earth-fixed vs head-fixed targets on static ocular counterroll. *Archives of Ophthalmology*, 128(4), 413-417.
- Chauveau, J. P. (20-2-2013). *Patent No. EP2134249 B1*.
- Chung, S. T., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research*, 38(19), 2949-2962.
- Coello, Y., Bartolo, A., & Weisbecker, A. (2007). Visual processing of words and spatial information for action. *Journal of vision*, 7(9), 514.
- Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *NeuroImage*, 40(1), 353-366.
- Cole, M. J., Gwizdka, J., Liu, C., Bierig, R., Belkin, N. J., & Zhang, X. (2011). Task and user effects on reading patterns in information search. *Interacting with Computers*, 23(4), 346-362.
- Collewijn, H., Steen, J., Ferman, L., & Jansen, T. C. (1985). Human ocular counterroll: assessment of static and dynamic properties from electromagnetic scleral coil recordings. *Exp Brain Res*, 59(1), 185-196.

- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204.
- Courrieu, P., Farioli, F., & Grainger, J. (2004). Inverse discrimination time as a perceptual distance for alphabetic characters. *Visual Cognition*, *11*(7), 901-919.
- Crossland, M., Legge, G., & Dakin, S. (2008). The development of an automated sentence generator for the assessment of reading speed. *Behavioral and Brain Functions*, *4*(1), 14.
- Crossland, M. D., Silva, S., & Macedo, A. F. (2014). Smartphone, tablet computer and e-reader use by people with vision impairment. *Ophthalmic and Physiological Optics*, *34*(5), 552-557.
- Danlos, L., & Roussarie, L. (2000). Génération Automatique de Textes. In J.M. Pierrel (Ed.), *Ingénierie des Langues* Hermès Science.
- Davoli, C. C., Du, F., Montana, J., Garverick, S., & Abrams, R. A. (2010). When meaning matters, look but don't touch: The effects of posture on reading. *Memory & Cognition*, *38*(5), 555-562.
- De Viron, L., Bernhard, D., Moriceau, V., & Tannier, X. (2011). Génération automatique de questions à partir de textes en français. *Traitement Automatique des Langues Naturelles*, 33.
- Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *JASIS*, *41*(6), 391-407.
- Dehaene, S. (2007). *Les Neurones de la lecture : La nouvelle science de la lecture et de son apprentissage* Odile Jacob.
- Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, *9*(7), 335-341.
- Dehais, F., Causse, M., Vachon, F., & Tremblay, S. (2012). Cognitive conflict in human-automation interactions: a psychophysiological study. *Applied ergonomics*, *43*(3), 588-595.
- Deplancke, A., Madelain, L., Chauvin, A., & Coello, Y. (2011). Attention spatiale et contrôle saccadique : données comportementales et neurobiologiques en faveur d'une conception motrice du contrôle attentionnel. *L'année psychologique*, *111*(03), 549-576.
- DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*(3), 1616-1641.
- Diarra, M., Kubryk, G., Megalakaki, O., Pasqualotti, L., Rico-Duarte, L., Binon, C. et al. (2011). The use of Interactive White Boards for the evaluation of reading activities in school. *Proceedings International Conference: The Future of Education* (pp. 161-164).

- Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics*, 35(10), 1297-1326.
- Dyson, M. C. (2004). How physical text layout affects reading from screen. *Behaviour & Information Technology*, 23(6), 377-393.
- Dyson, M. C., & Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, 54(4), 585-612.
- Edgell, S. E. (1995). Commentary on "Accepting the null hypothesis". *Memory & Cognition*, 23(4), 525.
- Egidi, G., & Gerrig, R. J. (2009). How valence affects language processing: Negativity bias and mood congruence in narrative comprehension. *Memory & Cognition*, 37(5), 547-555.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Fechner, G. T. (1860). *Elemente der Psychophysik* (2 Vols). Breitkopf and Hartel. Vol. 1 trans, by HE Adler (1966). *Elements of psychophysics*.
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A. et al. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior research methods*, 42(2), 488-496.
- Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, 8(4), 215-221.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 627-635.
- Fouquet, N., & Megalakaki, O. (2013). Construction et compréhension des catégories taxonomiques des animaux, végétaux et objets fabriqués chez des enfants de 3 à 6 ans. *Enfance*, 2013(02), 117-137.
- Friedman, S. M., Munoz, B., Rubin, G. S., West, S. K., Bandeen-Roche, K., & Fried, L. P. (1999). Characteristics of discrepancies between self-reported visual function and measured reading speed. Salisbury Eye Evaluation Project Team. *Investigative Ophthalmology & Visual Science*, 40(5), 858-864.
- Gage, W. H., Winter, D. A., Frank, J. S., & Adkin, A. L. (2004). Kinematic and kinetic validity of the inverted pendulum model in quiet standing. *Gait & posture*, 19(2), 124-132.
- Gateau, T., Durantin, G., Lancelot, F., Scannella, S., & Dehais, F. (2015). Real-Time State Estimation in a Flight Simulator Using fNIRS. *PloS one*, 10(3).
- Gates, D. M. (2008). *Automatically Generating Reading Comprehension Look-Back Strategy: Questions from Expository Texts* DTIC Document.

- Geissler, G. L., Zinkhan, G. M., & Watson, R. T. (2006). The influence of home page complexity on consumer attention, attitudes, and purchase intent. *Journal of Advertising*, 35(2), 69-80.
- Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. *Perception & Psychophysics*, 14(3), 471-482.
- Gibson, E. J. (1969). Principles of perceptual learning and development.
- Gille, J., Arend, L., & Larimer, J. O. (2004). Display characterization by eye: contrast ratio and discrimination throughout the grayscale. *Electronic Imaging 2004* (pp. 218-233) International Society for Optics and Photonics.
- Gonzalez, R. C., & Woods, R. E. (2002). *Digital image processing* (2 ed.).
- Goodenough, D. R., Sigman, E., Oltman, P. K., Rosso, J., & Mertz, H. (1979). Eye torsion in response to a tilted visual stimulus. *Vision Research*, 19(10), 1177-1179.
- Gopnik, A., & Sobel, D. M. (2000). Detectingblickets: How young children use information about novel causal powers in categorization and induction. *Child development*, 71(5), 1205-1222.
- Gould, J., Alfaro, L., Barnes, V., Finn, R., Grischowsky, N., Minuto, A. et al. (1987). Reading is slower from crt displays than from paper: attempts to isolate a single variable explanation. *Human Factors*, 29(3), 269-299.
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29(2), 228-244.
- Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition: Effects of masked homophone primes. *Journal of Memory and Language*, 33(2), 218-233.
- Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: from pixels to pandemonium. *Trends in Cognitive Sciences*, 12(10), 381-387.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32(6), 717-742.
- Hess, E. H. (1965). Attitude and pupil size. *Scientific american*.
- Hicks, J. L., & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic bulletin & review*, 6(1), 117-122.

- Hirota, M., Uozato, H., Arai, S., & Shibata, Y. (2013). Effect of Unclean Display on Accommodative Response. *Investigative Ophthalmology and Visual Science*, 54(6), 4257.
- Hock, H. S., & Tromley, C. L. (1978). Mental rotation and perceptual uprightness. *Perception & Psychophysics*, 24(6), 529-533.
- Hopkins, W. G. (2004). Bias in Bland-Altman but not regression validity analyses. *Sportscience*, 8(4).
- Huang, L., & Dobkins, K. R. (2005). Attentional effects on contrast discrimination in humans: Evidence for both contrast gain and response gain. *Vision Research*, 45(9), 1201-1212.
- Huang, Y. Y., & Menozzi, M. (2014). Effects of discomfort glare on performance in attending peripheral visual information in displays. *Displays*, 35(5), 240-246.
- Hubel, D. (1994). *L'oeil, le cerveau et la vision: les étapes cérébrales du traitement visuel* Pour la science.
- Ibbotson, M. R., Price, N. S. C., Das, V. E., Hietanen, M. A., & Mustari, M. J. (2005). Torsional eye movements during psychophysical testing with rotating patterns. *Exp Brain Res*, 160(2), 264-267.
- Isableu, B., Ohlmann, T., Crémieux, J., & Amblard, B. (1997). Selection of spatial frame of reference and postural control variability. *Exp Brain Res*, 114(3), 584-589.
- Jainta, S., & Baccino, T. (2010). Analyzing the pupil response due to increased cognitive demand: An independent component analysis study. *International Journal of Psychophysiology*, 77(1), 1-7.
- Jaschinski, W. (2002). The proximity-fixation-disparity curve and the preferred viewing distance at a visual display as an indicator of near vision fatigue. *Optometry & Vision Science*, 79(3), 158-169.
- Johansson, J., Pansell, T., Ygge, J., & Seimyr, G. Ö. (2014). The effect of contrast on monocular versus binocular reading performance. *Journal of vision*, 14(5), 8.
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain Activation Modulated by Sentence Comprehension. *Science*, 274(5284), 114-116.
- Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, 35(7), 1567-1577.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, 95(2), 163.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.

- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1-2), 262-284.
- Ko, P., Mohapatra, A., Bailey, I. L., Sheedy, J., & Rempel, D. M. (2014). Effect of Font Size and Glare on Computer Tasks in Young and Older Adults. *Optometry & Vision Science, 91*(6), 682-689.
- Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39*(16), 2729-2737.
- Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods, 44*(4), 978-990.
- Lam, A. K. C., Chung, E., Kho, J., & Wong, S. (2000). Digital measurement of torsional eye movement due to postural change and its effect on reading performance. *Current Eye Research, 21*(4), 763-766.
- Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100*(1), 150.
- Latham, K., & Whitaker, D. (1996). A comparison of word recognition and reading performance in foveal and peripheral vision. *Vision Research, 36*(17), 2665-2674.
- Lauritzen, S. L. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis, 19*(2), 191-201.
- Le Bigot, N., & Grosjean, M. (2012). Effects of Handedness on Visual Sensitivity in Perihand Space. *PloS one, 7*(8), e43150.
- Le Bigot, N., & Jaschinski, W. (2011). Hand position at computer screens: effect on visual processing. *Proceedings of the 29th Annual European Conference on Cognitive Ergonomics* (pp. 85-88) ACM.
- Lee, C. (1999). Eye and head coordination in reading: roles of head movement and cognitive control. *Vision Research, 39*(22), 3761-3768.
- Lee, D. S., Shieh, K. K., Jeng, S. C., & Shen, I. H. (2008). Effect of character size and lighting on legibility of electronic papers. *Displays, 29*(1), 10-17.
- Léger, L., Boumlak, H., & Tijus, C. (2008). BASETY: Extension et typicalité des exemplaires pour 21 catégories d'objets. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 62*(4), 223.
- Legge, G. E., Ross, J. A., Luebker, A., & LaMay, J. M. (1989). Psychophysics of reading. VIII. The Minnesota Low-Vision Reading Test. *Optom.Vis.Sci., 66*(12), 843-853.
- Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading-I. Normal vision. *Vision Research, 25*(2), 239-252.

- Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading. V. The role of contrast in normal vision. *Vision Research*, 27(7), 1165-1177.
- Legge, G. E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1985). Psychophysics of reading: II. Low vision. *Vision Research*, 25(2), 253-265.
- Lemaire, B., Guérin-Dugué, A., Baccino, T., Chanceaux, M., & Pasqualotti, L. (2011). A cognitive computational model of eye movements investigating visual strategies on textual material. *cogSci 2011 Proceedings*, 1146-1151.
- Leveau, N., Jhean-Larose, S., & Denhière, G. (2011). EMOVAL: Evaluation automatique de la valence et de l'activation émotionnelles des textes à l'aide d'une méta-norme de 5656 mots-racines. *Psychologie française*, 56(4), 209-221.
- Longcamp, M., Anton, J. L., Roth, M., & Velay, J. L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *NeuroImage*, 19(4), 1492-1500.
- Longcamp, M., Anton, J.-L., Roth, M., & Velay, J.-L. (2005). Premotor activations in response to visually presented single letters depend on the hand used to write: a study on left-handers. *Neuropsychologia*, 43(12), 1801-1809.
- Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction*, 21(3), 251-283.
- Maissa, C. A., Guillon, M., Wong, S., & Lebois, C. (2010). Time Controlled Visual Acuity - Repeatability and Its Relation to Conventional Visual Acuity. *ARVO 2010*.
- Maniwa, H., Kotani, K., Suzuki, S., & Asao, T. (2013). Changes in posture of the upper extremity through the use of various sizes of tablets and characters. In *Human Interface and the Management of Information. Information and Interaction Design* (pp. 89-96) Springer.
- Mansfield, J. S., Legge, G. E., & Bane, M. C. (1996). Psychophysics of reading. XV: Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science*, 37(8), 1492-1501.
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17(6), 578-586.
- Medhi, I., Sagar, A., & Toyama, K. (2006). Text-free user interfaces for illiterate and semi-literate users. *Information and Communication Technologies and Development, 2006. ICTD'06. International Conference on* (pp. 72-82) IEEE.
- Megalakaki, O., Aparicio, X., Porion, A., Pasqualotti, L., & Baccino, T. (2015). Assessing visibility, legibility and comprehension for interactive whiteboards (IWBs) vs. computers. *Educational Psychology*, (ahead-of-print), 1-20.
- Menozzi, M., Lang, F., Näpflin, U., Zeller, C., & Krueger, H. (2001). CRT versus LCD: effects of refresh rate, display technology and background luminance in visual performance. *Displays*, 22(3), 79-85.

- Miall, D. S., & Dobson, T. (2006). Reading hypertext and the experience of literature. *Journal of Digital Information*, 2(1).
- Michelson, A. A. (1927). *Studies in optics*. Chicago, Ill.: The University of Chicago Press.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41.
- Mostow, J., Beck, J., Bey, J., Cuneo, A., Sison, J., Tobin, B. et al. (2004). Using automated questions to assess reading comprehension, vocabulary, and effects of tutorial interventions. *Technology Instruction Cognition and Learning*, 2, 97-134.
- Murphy, K. (2001). The bayes net toolbox for matlab. *Computing science and statistics*, 33(2), 1024-1034.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'année psychologique*, 447-462.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, 36(3), 516-524.
- Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352-1375.
- Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity throughout adulthood. *Vision Research*, 23(7), 689-699.
- Panek, P. E., Barrett, G. V., Sterns, H. L., & Alexander, R. A. (1978). Age differences in perceptual style, selective attention, and perceptual-motor reaction time. *Experimental Aging Research*, 4(5), 377-387.
- Pasqualotti, L., & Baccino, T. (2014). Online advertisements: how are visual strategies affected by the distance and the animation of banners? *Frontiers in Psychology*, 5.
- Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America A*, 7(10), 2032-2040.
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46(28), 4646-4674.
- Perrin, J.-L., Paillé, D., & Baccino, T. (2014a). A new sentence generator providing material for maximum reading speed measurement. *Behavior research methods*, 1-10.
- Perrin, J.-L., Paillé, D., & Baccino, T. (2014b). Reading tilted: Does the use of tablets impact performance? An oculometric study. *Computers in Human Behavior*, 39(0), 339-345.
- Podgorny, P., & Garner, W. R. (1979). Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. *Perception & Psychophysics*, 26(1), 37-52.
- Poulain, I., Giraudet, G., & Dobrescu, N. (2004). Age-related changes in perception of verticality with a static or kinetic visual-field disturbance. *Perception*, 33(ECVP Abstract Supplement).

- Qiu, J., & Helbig, R. (2012). Body posture as an indicator of workload in mental work. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(4), 626-635.
- RAND Reading Study Group, & Snow, C. (2002). *Reading for Understanding: Toward an R&D Program in Reading Comprehension* RAND Corporation.
- Rapp, D. N., & van den Broek, P. (2005). Dynamic text comprehension an integrative view of reading. *Current Directions in Psychological Science*, 14(5), 276-279.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10(3), 241-255.
- Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and aging*, 21(3), 448.
- Rayner, K., Slattery, T. J., & Bélanger, N. N. (2010). Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review*, 17(6), 834-839.
- Rempel, D., Willms, K., Anshel, J., Jaschinski, W., & Sheedy, J. (2007). The effects of visual display distance on eye accommodation, head posture, and vision and neck symptoms. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(5), 830-838.
- Risse, S. (2014). Effects of visual span on reading speed and parafoveal processing in eye movements during sentence reading. *Journal of vision*, 14(8), 11.
- Robbins, H., & Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3).
- Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9(5), 487-494.
- Rubin, G. S. (2013). Measuring reading performance. *Vision Research*, 90, 43-51.
- Sadoski, M. (2001). Resolving the effects of concreteness on interest, comprehension, and learning important ideas from text. *Educational Psychology Review*, 13(3), 263-281.
- Salmerón, L., Baccino, T., Cañas, J. J., Madrid, R. I., & Fajardo, I. (2009). Do graphical overviews facilitate or hinder comprehension in hypertext? *Computers & Education*, 53(4), 1308-1319.
- Salmerón, L., & García, V. (2011). Reading skills and children's navigation strategies in hypertext. *Computers in Human Behavior*, 27(3), 1143-1151.
- Salvucci, D., & Goldberg, J. (2000). Identifying Fixations and Saccades in Eye-Tracking Protocols. *ETRA '00: Proceedings of the symposium on Eye tracking research & applications* (pp. 71-78) Palm Beach Gardens, Florida, United States: ACM.

- Schenk, T., & McIntosh, R. D. (2010). Do we have independent visual streams for perception and action? *Cognitive Neuroscience*, *1*(1), 52-62.
- Schleicher, A., Zimmer, K., Evans, J., & Clements, N. (2009). PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science. *OECD Publishing (NJI)*.
- Schmid, S., & Baccino, T. (2003). Perspective Shift and Text Format: An Eye-Tracking Study. *Current psychology letters*, *9*(3).
- Schultheis, H., & Jameson, A. (2004). Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. *Adaptive hypermedia and adaptive web-based systems* (pp. 225-234) Springer.
- Seiple, W., Szlyk, J. P., McMahon, T., Pulido, J., & Fishman, G. A. (2005). Eye-movement training for reading in patients with age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, *46*(8), 2886-2896.
- Senécal, M. J., Gresset, J., & Overbury, O. (2006). Minnesota Low-Vision Reading Test, version française: Échelle d'acuité visuelle MNREAD. Longueuil: Institut Nazareth & Louis-Braille.
- Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489-493.
- Siegenthaler, E., Bochud, Y., Bergamin, P., & Wurtz, P. (2012). Reading on LCD vs e-Ink displays: effects on fatigue and visual strain. *Ophthalmic and Physiological Optics*, *32*(5), 367-374.
- Siegenthaler, E., Wurtz, P., Bergamin, P., & Groner, R. (2011). Comparing reading processes on e-ink displays and print. *Displays*, *32*(5), 268-273.
- Simola, J., Kuisma, J., Öörni, A., Uusitalo, L., & Hyönä, J. (2011). The impact of salient advertisements on reading and attention on web pages. *Journal of Experimental Psychology: Applied*, *17*(2), 174.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34.
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71-86.
- Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652-654.
- Strahm, M., & Baccino, T. (2006). L'impact des schémas non analogiques conceptuels dans la compréhension de textes expositifs: stratégies visuelles selon l'expertise. *Psychologie française*, *51*(1), 25-40.
- Straker, L. M., Coleman, J., Skoss, R., Maslen, B. A., Burgess-Limerick, R., & Pollock, C. M. (2008). A comparison of posture and muscle activity during tablet computer, desktop computer and paper use by young children. *Ergonomics*, *51*(4), 540-555.

- Thomas, J. B. (2009). *Colorimetric characterization of displays and multi-display systems* (PhD, Université de Bourgogne).
- Thomson, S., & De Bortoli, L. (2012). Preparing Australian students for the digital world: Results from the PISA 2009 digital reading literacy assessment.
- Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 9(1), 40-50.
- Trauzettel-Klosinski, S., Dietz, K., & the IReST Study Group. (2012). Standardized Assessment of Reading Performance: The New International Reading Speed Texts IReST. *Investigative Ophthalmology & Visual Science*, 53(9), 5452-5461.
- Traxler, M. J., Johns, C. L., Long, D. L., Zirnstein, M., Tooley, K. M., & Jonathan, E. (2012). Individual differences in eye-movements during reading: Working memory and speed-of-processing effects. *Journal of Eye Movement Research*, 5(1), 5.
- Tressoldi, P. E., Stella, G., & Faggella, M. (2001). The development of reading speed in Italians with dyslexia A longitudinal study. *Journal of learning disabilities*, 34(5), 414-417.
- Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, 35(17), 2503-2522.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, 26(1), 1.
- van den Broek, P., Risden, K., Fletcher, C. R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In I. Lawrence Erlbaum Associates (Ed.), *Models of Understanding Text* (pp. 165-187).
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension* Academic Press New York.
- Velay, J. L., Longcamp, M., & Zerbato-Poudou, M. T. (2004). De la plume au clavier: est-il toujours utile d'enseigner l'écriture manuscrite. In Dunod (Ed.), *Comprendre les apprentissages: Sciences cognitives et éducation* (pp. 69-82).
- Video Electronics Standards Association. (2001). *Flat panel display measurements standard, version 2.0* VESA. doi:10.1002/col.10084.
- Wade, S. E. (1990). Using think alouds to assess comprehension. *The Reading Teacher*, 442-451.
- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics*, 18(3).
- Werlich, E. (1975). *Typologie der texte*.
- Witkin, H. A. (1949). The nature and importance of individual differences in perception. *Journal of Personality*, 18(2), 145-170.

- Witkin, H. A., & Asch, S. E. (1948). Studies in space orientation. IV. Further experiments on perception of the upright with displaced visual fields. *Journal of Experimental Psychology*, 38(6), 762.
- Young, A., & Greig Bowers, P. (1995). Individual difference and text difficulty determinants of reading fluency and expressiveness. *Journal of Experimental Child Psychology*, 60(3), 428-454.
- Young, J. G., Trudeau, M., Odell, D., Marinelli, K., & Dennerlein, J. T. (2012). Touch-screen tablet user configurations and case-supported tilt affect head and neck flexion angles. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41(1), 81-91.
- Young, J. G., Trudeau, M. B., Odell, D., Marinelli, K., & Dennerlein, J. T. (2013). Wrist and shoulder posture and muscle activity during touch-screen tablet use: Effects of usage configuration, tablet type, and interacting hand. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 45(1), 59-71.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338-353.
- Zambarbieri, D., & Carniglia, E. (2012). Eye movement analysis of reading from computer displays, eReaders and printed books. *Ophthalmic and Physiological Optics*, 32(5), 390-396.
- Zhu, E. (1999). Hypermedia Interface Design: The Effects of Number of Links and Granularity of Nodes. *Journal of Educational Multimedia and Hypermedia*, 8(3), 331-358.
- Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, 24(3), 875-887.

Résumé long

Les dispositifs d'affichage électroniques sont de plus en plus présents dans notre environnement, et nous les utilisons régulièrement pour consulter l'information qu'ils exposent. La lecture électronique est donc devenue une activité récurrente. A l'heure actuelle, les supports ont pris des formes multiples, (smartphones, tablettes, vidéoprojecteurs,...), l'intérêt de cette thèse est de proposer des outils et méthodes pour pouvoir comparer les performances de lecture associées à chacun de ces supports.

De l'arrivée de la lumière sur la rétine à la compréhension du texte, de nombreux traitements perceptifs et cognitifs sont réalisés à différentes échelles par le lecteur. La lecture est l'association de ces processus. Cette activité est impactée par les dispositifs d'affichages numériques à différents niveaux. La visibilité, qui est en lien avec la capacité du lecteur à percevoir les stimulations visuelles et la lisibilité, en lien avec sa capacité à identifier les lettres et les mots, sont impactées par la qualité des écrans. La compréhension peut être également altérée par les dispositifs, car l'affichage de l'information peut y prendre des formes multiples. A ces trois dimensions classiques de la lecture électronique, nous ajoutons la dimension posturale. En effet, de plus en plus de supports étant mobiles, ils amènent des réponses posturales nouvelles.

Nous construisons dans cette thèse des outils pour pouvoir évaluer un dispositif d'affichage sur chacune des dimensions caractérisant la lecture électronique. Afin de pouvoir comparer les supports, les tests doivent donc permettre d'effectuer des mesures répétées. Le choix du matériel de test est donc primordial : les stimuli affichés au lecteur doivent être équivalents pour permettre d'obtenir des réponses stables à chaque recueil de données, mais ils doivent également limiter les effets liés à la lecture répétée (apprentissage, changement de stratégie de lecture,...).

Pour la visibilité, nous proposons d'utiliser une méthode de caractérisation d'écran basée sur la détection de contrastes des lecteurs.

Afin d'évaluer la lisibilité, nous proposons d'utiliser un test d'identification de lettres, une tâche de décision lexicale et une méthode d'évaluation de la vitesse maximale de lecture. Pour ce dernier test, nous avons mis en place un algorithme qui génère automatiquement des phrases équivalentes, afin de pouvoir répéter les mesures sans que les lecteurs ne puissent apprendre le contenu de ces phrases.

Un test de compréhension est également construit, de manière à évaluer la mémorisation ainsi que les inférences que les lecteurs produisent en lisant des textes. Pour ce test, nous avons développé un générateur automatique de textes descriptifs que nous contrôlons sur plusieurs aspects (difficulté, cohérence, information contenue).

Le détail technique de chaque test est détaillé dans l'ouvrage. Pour le test de vitesse de lecture, nous présentons également une étude de validation de la mesure.

Les liens qui associent lecture électronique et posture sont encore peu connus. A l'aide de deux études nous cherchons à questionner d'une part l'effet de la posture sur la performance de lecture et d'autre part l'effet de la lecture sur la posture.

La première étude sur la posture adresse l'effet d'une inclinaison latérale de la tête (mouvement pouvant être engendré par l'utilisation de tablette) par rapport à la performance de lecture. L'expérience montre, à l'aide de variables oculométriques, qu'une inclinaison latérale de la tête ne modifie pas la performance mesurée : on observe une adaptation des lecteurs. Nous montrons qu'une explication possible peut être trouvée dans les réponses oculomotrices des lecteurs, dans la mesure où ils effectuent des cyclotorsions en réponse à ce contexte postural.

La seconde étude posturale porte sur l'effet de la complexité d'un texte sur la distance à l'écran. Les résultats montrent que les sujets se rapprochent de l'écran lors de la lecture de textes difficiles. Des mesures concourantes montrent que la variation de la distance à l'écran est inversement corrélée au niveau d'attention des lecteurs.

Nous utilisons ce dernier résultat pour évaluer la posture durant la lecture électronique.

Une fois la batterie de test construite, nous évaluons la lecture sur trois supports : un écran d'ordinateur fixe, un vidéoprojecteur ainsi qu'une tablette. Le recueil de données a été réalisé pour répondre à deux objectifs : d'une part pouvoir comparer ces dispositifs, et d'autre part pouvoir construire des indicateurs de la qualité de lecture sur support électronique. Ces indicateurs sont synthétisés sous la forme d'une valeur numérique représentant la performance associée à un support : le Quotient de Lecture Electronique (QLE).

Les tests ont montré que la visibilité est moins bonne pour le projecteur que pour la tablette et l'écran fixe. La tablette, dont le contraste et la définition étaient les plus élevés, était associée aux meilleures performances pour le traitement des lettres et des mots.

Néanmoins, ces différences disparaissent pour les tests de lecture de phrases et de compréhension.

Ces différences se répercutent dans les indicateurs de performances, et les scores de QLE reprennent le classement de performance observé (les scores sont pour le projecteur : 50,75, pour l'écran fixe : 121,44 et pour la tablette : 127,81).

Les tests permettent donc de différencier les supports pour les dimensions de bas niveaux, associées au traitement visuel. A interfaces équivalentes, les dimensions de lecture de plus hauts niveaux semblent moins impactées par le support en lui-même. L'effet des interfaces serait également à évaluer, en effet, ce sont ces interfaces qui font que la lecture sur support électronique est moins linéaire, parfois plus complexe, mais aussi parfois plus riche que la lecture classique sur papier.

Ces travaux aboutissent donc à une nouvelle méthodologie d'évaluation permettant des mesures répétées et comparatives entre différents dispositifs d'affichage, ce qui pourrait permettre de faire progresser les connaissances actuelles sur la lecture électronique.