



**HAL**  
open science

# Contributions des architectures textuelles à l'accessibilité non-visuelle des documents numériques

Laurent Sorin

► **To cite this version:**

Laurent Sorin. Contributions des architectures textuelles à l'accessibilité non-visuelle des documents numériques. Traitement du texte et du document. Université Toulouse le Mirail - Toulouse II, 2015. Français. NNT : 2015TOU20132 . tel-01578078

**HAL Id: tel-01578078**

**<https://theses.hal.science/tel-01578078v1>**

Submitted on 28 Aug 2017

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

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



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse II - Jean-Jaurès*

Discipline ou spécialité : *Informatique*

---

---

Présentée et soutenue le *15 Décembre 2015* par

**LAURENT SORIN**

Titre : *Contributions des architectures textuelles à l'accessibilité non-visuelle des documents numériques*

---

---

### JURY

Archambault Dominique – *Rapporteur*

Aussenac-Gilles Nathalie – *Co-encadrante*

Chêne Denis – *Membre du Jury*

Dubois Emmanuel – *Président du Jury*

Lemarié Julie – *Co-directrice*

Lenay Charles – *Rapporteur*

Maurel Fabrice – *Membre du Jury*

Mojahid Mustapha – *Directeur de thèse*

Professeur à l'Université Paris VIII

Directrice de Recherche CNRS

Chercheur sénior – Orange Labs.

Professeur à l'Université Toulouse III

Maître de Conférences à l'Université Toulouse II

Professeur à l'Université de Technologie de Compiègne

Maître de Conférences à l'Université de Caen

Maître de Conférences à l'Université Toulouse III

---

**École Doctorale et spécialité** : Mathématiques, Informatique et Télécommunications de Toulouse (MITT) – Image, Information, Hypermédia.

**Unité de Recherche** : Institut de Recherche en Informatique de Toulouse (IRIT) – UMR 5505.

**Directeurs de Thèse** : Mustapha Mojahid et Julie Lemarié.

**Rapporteurs** : Dominique Archambault et Charles Lenay.



*« Les mots sont sans fin, mais la communication est d'ordre verbal autant que non-verbal.  
Entendre un mot est une chose, entendre une absence de mot en est une autre. »*

Jiddu Krishnamurti - Journal.

Je voudrais très sincèrement remercier Julie Lemarié, Nathalie Aussenac et Mustapha Mojahid pour avoir dirigé ma thèse et y avoir largement contribué. Merci de votre confiance, de vos conseils, et de votre bonne volonté, et surtout merci de m'avoir laissé faire des choix qui me correspondaient. Grâce à vous j'ai eu la chance de passer 3 années enrichissantes tant sur le plan scientifique qu'humain.

Un grand merci à Dominique Archambault et Charles Lenay d'avoir accepté de rapporter cette thèse, et ainsi par la même d'y contribuer en y apportant leur longue expérience de la recherche scientifique.

Je remercie spécialement Emmanuel Dubois, qui fut mon tout premier contact à l'IRIT et qui m'a aiguillé sur la thèse concrétisée aujourd'hui par ce manuscrit ; merci également d'avoir accepté d'être membre du jury : la boucle est bouclée !

Merci à Denis Chêne et à Fabrice Maurel pour leur participation au jury de cette thèse, je suis très heureux d'avoir eu des points de vue aussi différents et riches sur mes travaux.

Merci à tous les membres de l'équipe ELIPSE à l'IRIT et plus particulièrement ceux avec qui j'ai partagé de bons moments, et ceux qui ont su me remonter le moral dans les moments difficiles et ceux qui m'ont aidé et conseillé. Merci à Jean-Philippe Fauconnier avec qui j'ai été très heureux de collaborer durant ces quelques semaines sur la formalisation automatique de la mise en forme. Merci à Antoine Dupuis qui a participé au projet à l'occasion de son stage de Master.

Merci à Claude Griet qui a été mon contact privilégié à l'Institut des Jeunes Aveugles de Toulouse, sans qui je n'aurais simplement pas pu mener à bien mes expérimentations. Je la remercie pour sa réactivité et sa gentillesse. Merci aussi à Mathieu Delpy (anciennement) de l'IJA qui a largement participé à ce projet. Merci également à tous les membres de l'IJA et à tous les volontaires qui ont participé à mes tests ; sans eux non plus rien n'aurait été possible.

Merci à toute ma famille, mes parents, mon frère et ma sœur, mes grands-parents. Votre soutien et votre amour me font toujours chaud au cœur. Merci à mon père qui a fait tant de sacrifices pour ses enfants, et qui est toujours là pour nous.

Enfin, je dédie ce travail à celle qui m'a soutenu (et supporté !) pendant ces 3 années, qui m'a aidé à oublier mes doutes et à continuer d'avancer. À toi Justine, mon amour.

## Résumé

---

L'accessibilité numérique joue un rôle décisif dans l'éducation, l'inclusion sociale et l'autonomie des individus souffrant d'une déficience. Dans cette thèse, nous nous sommes intéressés à une composante quasi-universelle des documents numériques : la mise en forme des textes. L'objectif de nos travaux était de proposer aux déficients visuels accédant à des contenus textuels un accès fonctionnel à la sémantique de leur mise en forme. En effet, loin de n'être qu'un ornement à visée esthétique, les marques de mise en forme bâtissent de véritables architectures textuelles.

Une étude exploratoire que nous avons menée auprès d'élèves déficients visuels inclus en classe ordinaire, portant sur leurs difficultés et besoins en matière d'accessibilité aux documents, a notamment montré que ces élèves ont des difficultés pour accéder aux contenus pédagogiques textuels, qui entraînent des situations de surcharge cognitive au quotidien. Une partie de ces difficultés provient de l'absence d'accès aux architectures textuelles véhiculées par la mise en forme, qui fait obstacle à la construction, de façon anticipée, d'une représentation mentale de l'organisation globale du document. Parallèlement à cette étude exploratoire, l'examen de la littérature atteste que les technologies d'assistance utilisées par les déficients visuels pour accéder aux contenus ne restituent qu'une partie de la sémantique de la mise en forme. Or, quantité de travaux montrent le rôle crucial joué par la mise en forme en lecture visuelle, notamment sur la compréhension.

Nos travaux ont donc visé à extraire formellement la sémantique de la mise en forme (les architectures textuelles correspondantes), à concevoir des stratégies de restitution multimodales de cette sémantique selon une approche holistique (prenant en compte les caractéristiques de la mise en forme, des utilisateurs et de la tâche de compréhension de textes), et enfin à évaluer l'impact de ces stratégies de restitution. Autrement formulé, le but était de savoir comment rendre compte des architectures textuelles des documents numériques, de manière non-visuelle, et dans quelle mesure les utilisateurs déficients visuels pouvaient bénéficier de cette démarche.

Afin d'extraire la sémantique de la mise en forme, nous avons développé une méthodologie de formalisation basée sur l'analyse des marques typographiques, dispositionnelles et lexico-syntaxiques de mise en forme. Cette méthodologie, qui s'inscrit dans le cadre théorique du Modèle d'Architectures Textuelles [Virbel 1985; Virbel 1989], a l'ambition de permettre de modéliser universellement la sémantique de toute architecture textuelle. Une première implémentation à l'aide d'un algorithme d'apprentissage automatique a montré des résultats prometteurs quant à la modélisation d'architectures simples.

Afin de restituer les informations métatextuelles extraites, nous avons élaboré différentes stratégies de restitution orales (multimodales), dont nous avons évalué l'efficacité auprès de 72 volontaires dont 60 voyants et 12 non-voyants. Les résultats de ces deux expérimentations révèlent que tout d'abord, la restitution des architectures textuelles permet d'améliorer la compréhension du contenu tout en diminuant la charge cognitive des utilisateurs, mais également que l'utilisation de modalités non-verbales permet aux utilisateurs de se sentir moins désorientés pendant l'accès aux documents. L'ensemble de nos travaux montrent que l'accessibilité des contenus numériques pour les déficients visuels peut encore être améliorée, afin que ces contenus soient non seulement accessibles mais que cet accès soit le plus performant, satisfaisant et efficient possible sur le plan cognitif. En conclusion, l'étude des architectures textuelles véhiculées par la mise en forme, et la conception d'équivalents

fonctionnels, constituent donc une nouvelle piste pertinente pour l'amélioration de l'accessibilité des documents numériques, pour les utilisateurs en situation de handicap visuel ou présentant un handicap incapacitant pour la lecture.

Mots clés : déficience visuelle, accessibilité, Modèle d'Architectures Textuelles, mise en forme, interfaces non-visuelles, compréhension.

# Sommaire

---

Sommaire .....	vi
Préambule et indices de lecture .....	1
Introduction.....	2
<i>Chapitre I - Aspects théoriques de la mise en forme des textes, et bilan empirique de ses rôles pour le lecteur .....</i>	<i>7</i>
1. La mise en forme des textes.....	8
2. Le modèle d'architectures textuelles (MAT) .....	10
3. Conclusion.....	25
<i>Chapitre II - Accessibilité des contenus textuels pour les déficients visuels : l'enjeu de la restitution de la mise en forme.....</i>	<i>26</i>
1. Introduction .....	28
2. Définir l'accessibilité pour mieux cerner ses enjeux .....	31
3. Comment les déficients visuels perçoivent et traitent l'information .....	34
4. Approches actuelles pour l'accessibilité des contenus numériques.....	44
5. Les technologies d'assistance pour l'accès non-visuel aux contenus numériques : approches et limites	54
6. Vers la restitution des architectures textuelles pour l'accessibilité des contenus numériques .....	73
7. Étude en situation : accès et accessibilité de l'information pédagogique pour les élèves déficients visuels .....	78
8. Conclusion et problématique générale.....	89
<i>Chapitre III - Méthodologie de formalisation de la Mise en Forme Matérielle ..</i>	<i>90</i>
1. Introduction .....	92
2. Socle théorique pour la formalisation de la MFM.....	93
3. Segmentation des objets textuels des documents.....	98
4. Définition et identification des conventions partagées de MFM .....	105
5. Modélisation des conventions locales de MFM .....	115
6. Vue d'ensemble du processus d'annotation et opérationnalisation technique .....	119
7. Vers une formalisation automatique de la MFM par apprentissage automatique.....	123
8. Conclusion.....	126
<i>Chapitre IV - Restituer la Mise en Forme Matérielle à l'oral : conception et opérationnalisation.....</i>	<i>128</i>
1. Introduction .....	130
2. Typologie des informations de MFM à restituer.....	131

3. Vers une restitution optimisée de la MFM : étude globale de l'espace de conception .....	133
4. Proposition d'une stratégie de restitution de la MFM .....	151
5. READ : une plateforme expérimentale pour implémenter et évaluer des stratégies d'accès non-visuelles aux documents .....	155
6. Conclusion.....	164
<i>Chapitre V - Évaluations de la restitution orale de la mise en forme matérielle</i> .....	<i>167</i>
1. Introduction .....	168
2. Concepts clés.....	170
3. Évaluation de la mise en saillance des titres à l'oral auprès de sujets non-voyants .....	172
4. Évaluation de 4 stratégies de restitution de la MFM auprès de sujets voyants.....	182
5. Évaluation de 2 stratégies de restitution de la MFM auprès sujets non-voyants .....	224
6. Conclusion sur l'ensemble des résultats.....	250
<i>Conclusion générale et perspectives</i> .....	<i>256</i>
7. Synthèse des contributions .....	256
8. Perspectives : futurs travaux .....	261
<i>Références</i> .....	<i>264</i>
<i>Annexes</i> .....	<i>279</i>
<i>Table des matières du manuscrit</i> .....	<i>310</i>
<i>Glossaire</i> .....	<i>316</i>
<i>Abréviations</i> .....	<i>317</i>
<i>Table des figures</i> .....	<i>317</i>
<i>Liste des tableaux</i> .....	<i>320</i>
<i>Liste des publications</i> .....	<i>321</i>
<i>Résumé de la thèse en Anglais – Abstract</i> .....	<i>322</i>





---

## Préambule et indices de lecture

---

Ce document rapporte les travaux réalisés dans le cadre d'une thèse de doctorat en Informatique entre Décembre 2012 et Décembre 2015, au laboratoire IRIT (Institut de Recherche en Informatique de Toulouse) de l'Université Toulouse III au sein de l'équipe ELIPSE (Étude de L'Interaction Personne-Système). Cette thèse était encadrée par Mustapha Mojahid (IRIT – ELIPSE), Nathalie Aussenac-Gilles (IRIT – MELODI : Méthodes et ingénierie des Langues, des Ontologies et du Discours) et Julie Lemarié (laboratoire CLLE : Cognition Langues Langage et Ergonomie – Université Toulouse II Jean-Jaurès). Ces travaux ont été financés par la Région Midi-Pyrénées et le Pôle de Recherche et d'Enseignement Supérieur (PRES) Toulouse.

Ce document est structuré en 7 chapitres ; chaque chapitre comprend plusieurs parties numérotées à l'aide de chiffres romains, plusieurs sections numérotées à l'aide de chiffres arabes et parfois des sous-sections ordonnées par des lettres capitales. Chaque chapitre débute par un résumé de son contenu suivi de son propre sommaire. Une table des matières complète, une table des figures, une liste des tableaux, un glossaire ainsi que les annexes sont disponibles à la fin du document.

Afin de faciliter la lecture, le titre du chapitre courant est affiché en en-tête sur les pages paires du document, et le titre de la partie courante est affiché en en-tête sur les pages impaires. Chaque page est numérotée ; les numéros sont situés en pied de page.

Pour lecteurs accédant à une version électronique du document, il est possible de naviguer dans le document à partir des différents sommaires et tables des matières, chaque item étant un lien vers la partie correspondante dans le document. Enfin, à noter que les références croisées présentes dans le texte sont des liens fonctionnels.

## Introduction

---

*« J'ai connu l'époque où l'on n'avait pas toutes les possibilités qu'offre l'informatique... ça n'avait rien à voir. »*

C'est ce que rapportait un volontaire, aveugle de naissance, ayant participé aux tests menés dans le cadre de nos travaux. Sans aller jusqu'à présenter l'outil informatique comme une solution absolue, les nouvelles technologies et la recherche ont un rôle important à jouer dans l'autonomie et l'inclusion sociale de toute personne présentant une déficience, quelle qu'elle soit. Ainsi, au-delà de l'intérêt scientifique du développement de méthodes et outils pour les déficients visuels relatés dans ces travaux, l'intérêt sur le plan humain de l'accessibilité des interfaces informatiques et contenus numériques est certain. Cette thèse vise à modestement contribuer à ces différents domaines.

Nos travaux sont basés sur trois constats différents, dont la validité sera argumentée dans les trois paragraphes suivants : la mise en forme des textes a une double portée sémantique et fonctionnelle ; elle joue un rôle important dans la compréhension des textes ; et enfin la mise en forme n'est que partiellement prise en compte par les outils et techniques actuels d'accès non-visuel aux documents, utilisés par les déficients visuels. Ainsi, l'objectif de ces travaux était d'évaluer comment et dans quelle mesure **la restitution de la mise en forme des textes permettrait d'améliorer l'accessibilité des documents numériques**. Le terme restitution englobe ici deux niveaux d'amélioration ; le premier niveau concerne la disponibilité des informations de mise en forme, et vise à permettre l'accès à des informations qui auparavant étaient ignorées ou faiblement restituées. Le second niveau concerne l'optimisation de cet accès sur un plan cognitif et s'inscrit dans une réflexion plus générale sur l'utilisabilité des interfaces et méthodes d'accès non-visuel aux documents numériques.

La mise en forme des textes correspond à des marques typographiques et dispositionnelles (couleurs, polices, graisse, alinéas, disposition...) qui permettent de distinguer des *objets textuels* : titres, énumérations, paragraphes, etc. En effet, elle ne constitue pas simplement un ornement du texte à visée esthétique, mais communique au lecteur des informations relatives au texte lui-même, en l'occurrence des informations sur son organisation et sa structuration, que l'on nomme ici des architectures textuelles [Virbel 1989]. Ces informations jouent un rôle important dans la lecture visuelle et contribuent à la compréhension [Lemarié et al. 2008]. Une première illustration, certes un peu naïve de ce constat, est qu'un texte auquel on enlèverait toute mise en forme, ainsi transformé en un seul bloc de texte compact, subirait non seulement une perte de sa sémantique d'origine mais serait également plus restreint en termes fonctionnels ; il serait par exemple plus difficile d'y retrouver une information.

On attribue le développement de la mise en forme des textes écrits aux XII<sup>ème</sup> et XIII<sup>ème</sup> siècle à un besoin fonctionnel de pouvoir traiter des volumes de texte grandissants, qui ne pouvait être satisfait qu'à travers une lecture silencieuse (et non pas orale) et une structuration visuelle des textes, comme indiqué par [Maurel 2006]. Cette organisation visuelle de l'information pourrait être rapprochée de l'utilisation de la prosodie à l'oral [Nespoulous et al. 2002], qui compenserait nos capacités cognitives limitées en facilitant le traitement de la parole à l'aide d'une structure « intonative ». L'imprimerie aurait contribué à la démocratisation de la lecture silencieuse, et avec elle à la multiplication des phénomènes de mise en forme dans les textes. Quoi qu'il en soit, la psychologie des apprentissages et la psycholinguistique confirment la portée fonctionnelle de la mise

en forme et son importance pour la lecture. En effet, les *signaux visuels* (mise en forme des textes) permettent d'améliorer la mémorisation des éléments signalés [Lorch 1989], et jouent un rôle prépondérant dans la compréhension des textes qui nécessite la construction d'une représentation globale cohérente du contenu [Van Dijk and Kintsch 1983], processus largement appuyé par la structuration visuelle [Lemarié et al. 2012]. D'une manière plus générale, les signaux visuels jouent un rôle de facilitateurs de traitement de l'information écrite [Lemarié et al. 2008]. Ainsi, sur ces bases empiriques, nous avons choisi de nous intéresser à la tâche de compréhension d'un texte expositif (exposant, développant des concepts), et aux bénéfices potentiels de la restitution de la mise en forme dans le cadre de cette tâche.

Le troisième et dernier constat sur lequel se basent nos travaux est que la mise en forme, sa sémantique et les fonctions de lecture visuelles associées ne sont que très partiellement retranscrites par les outils et techniques mis à disposition des déficients visuels. En effet, l'approche actuelle d'accès non-visuel aux documents (appliquée par les technologies d'assistance) est principalement basée sur la sémantique du langage utilisé pour structurer et représenter le contenu des documents. Or, ces langages (comme le HTML) ne permettent de décrire qu'une partie des architectures textuelles accessibles aux lecteurs voyants (voir Chapitre II). Ce propos est cependant à nuancer au vu de l'apparition de nouvelles normes d'accessibilité prenant en compte une partie supplémentaire de la sémantique de la mise en forme des textes [W3C - World Wide Web Consortium 2010]. Nous pensons que le statut de la mise en forme, plus souvent considérée comme un élément esthétique des textes ou un facteur de confort de lecture que partie intégrante de leur sémantique, a pu limiter la prise en compte des architectures textuelles dans les langages de représentation de documents numériques et dans leur accès par les technologies d'assistance.

Afin de mettre à l'épreuve l'hypothèse principale ayant guidé nos travaux, à travers l'évaluation de l'impact de la restitution des architectures textuelles auprès d'utilisateurs déficients visuels, il était tout d'abord nécessaire de pouvoir modéliser ces architectures textuelles et de mettre en place différentes stratégies de restitution pertinentes pour cette évaluation. Les paragraphes suivants développent les différentes questions de recherche correspondantes à cette démarche.

La première question de recherche qui s'est posée a été *comment formaliser la mise en forme*? En d'autres termes, comment analyser et extraire l'information véhiculée par la mise en forme dans le but de la rendre accessible. Le cadre théorique posé par le Modèle d'Architecture Textuelle (MAT) [Virbel 1985; Virbel 1989], modèle logico-linguistique, nous permet de mieux comprendre la genèse de la mise en forme signifiante des documents, ici appelée Mise en Forme Matérielle (MFM). Ce modèle initialement utilisé dans le domaine de la génération automatique de texte [Pascual 1996; Luc 2000], nous a permis de développer une méthodologie de formalisation automatisable de la MFM, avec des premiers résultats encourageants quant à son application automatisée à l'aide d'algorithmes d'apprentissage [Fauconnier et al. 2014]. Cependant, formaliser les architectures textuelles d'un document relève de l'interprétation parfois *subjective* de ses marques de MFM ; en effet un phénomène de MFM peut être polysémique, et donc donner lieu à plusieurs interprétations. De plus, même si certains principes régissent l'utilisation de la MFM dans les textes, les auteurs de documents peuvent ne pas les respecter en faisant des erreurs, par exemple. Ainsi nous ne prétendons pas pouvoir formaliser tout document existant à l'aide de cette méthodologie, mais plutôt poser les bases de l'opérationnalisation du MAT à l'annotation de la MFM des documents.

La deuxième question de recherche ayant guidé ces travaux a été *comment restituer les informations et fonctions de MFM à l'oral* ? Le choix du mode audio plutôt que le mode tactile (ou une combinaison de ces modes), et donc de la synthèse vocale comme socle pour la restitution, est argumenté au début du Chapitre IV. La question initiale se divise en deux questions ; la première étant : *faut-il restituer l'ensemble des informations formalisées* ? On peut d'ores et déjà répondre par la négative ; la séquentialité imposée par l'oral et ses contraintes biomécaniques ne sont pas compatibles avec la densité et la quantité d'informations visuelles d'un document. Sans un filtrage préalable, restituer une telle quantité d'informations (potentiellement non-pertinentes eu égard aux objectifs de l'utilisateur) pourrait avoir un impact négatif sur la charge cognitive [Giraud 2014], et engendrerait un allongement du temps de parcours du document et donc des frustrations [Lazar et al. 2007]. Cette question sera abordée sous l'angle de l'ergonomie cognitive par l'étude de la tâche réalisée à l'aide d'un document ; on filtrera donc les informations de mise en forme et on modulera leur restitution en fonction de leur pertinence pour la réalisation des objectifs de la tâche de compréhension, mais aussi plus généralement en fonction du contexte d'utilisation. Ce qui nous amène à la seconde sous-question : *dans quelle(s) modalité(s) restituer ces informations (verbale ou non-verbale)* ? L'oral offre en effet un panel de modalités différentes comme les icônes audio, la prosodie ou encore la spatialisation du son qui présentent chacune des caractéristiques différentes, et dont l'utilisation pourra être plus ou moins pertinente en fonction du contexte de la restitution.

Enfin, la dernière question qui s'est posée, une fois des pistes de réponse données aux deux premières, a été *comment évaluer la restitution de la mise en forme* ? En nous inspirant de travaux existants [Lemarié et al. 2006; Giraud 2014] et toujours dans une approche d'ergonomie cognitive, nous avons évalué la restitution à la fois sur les performances de compréhension du texte dont on restitue les architectures textuelles, mais aussi par des évaluations subjectives de la charge cognitive, c'est-à-dire les ressources mentales engagées pour arriver à ces performances.

Ainsi, le positionnement de nos travaux est fondamentalement pluridisciplinaire, au confluent de la linguistique pour la formalisation de la mise en forme, de la psycholinguistique qui pointe les rôles de cette dernière dans la lecture/compréhension et nous donne des outils pour la mesure de ses effets, et enfin de l'ergonomie cognitive qui invite à prendre en compte le contexte d'utilisation du document et les besoins réels des utilisateurs. Bien que les résultats les plus directs de nos travaux concernent la conception d'interfaces homme-machine à destination des utilisateurs déficients visuels, l'étude des architectures textuelles et de la mise en forme a une portée plus large. Cet objet d'étude, en apparence anodin, renferme de nombreuses richesses et complexités ; nos travaux pourraient ainsi contribuer aussi bien à l'analyse automatique de documents qu'à la navigation non-visuelle au sein de textes, ou encore à la production de documents en braille sur papier.

### ► Vue d'ensemble du manuscrit :

Comme évoqué en introduction, nos travaux reposent sur le double constat, d'un côté, de la portée sémantique de la mise en forme (donnant lieu à des architectures textuelles) et de son importance pour le lecteur, et d'un autre côté, de l'existence de problèmes d'accessibilité liés au manque de prise en compte des architectures par les textuelles technologies d'assistance. Nous avons choisi de présenter dans le premier chapitre la mise en forme, ses modes d'expression, sa sémantique et les rôles qu'elle joue pour le lecteur, afin de mieux comprendre les différents problèmes d'accessibilité des contenus textuels et des technologies d'assistance pointés dans le chapitre 2.

Ainsi, dans le premier chapitre, nous définirons ce qu'est la mise en forme d'après le cadre théorique posé par le Modèle d'Architecture(s) Textuelle(s) (MAT). On évoquera notamment le fait que les marques de mise en forme ne sont pas toutes signifiantes, mais que celles qui le sont seraient issues, d'après le MAT, d'intentions auteuriales d'informer le lecteur quant à l'organisation du texte. Ce chapitre fera également le lien avec certaines études sur la mise en forme (appelée dans le cadre de la psycholinguistique « signaux visuels ») ayant montré ses différents rôles pour le lecteur.

Le chapitre suivant dresse un état de l'art de l'accessibilité des contenus numériques, et des technologies d'assistance utilisées par les déficients visuels pour accéder à ces contenus. En début de chapitre nous présenterons en préambule plusieurs réflexions sur l'accès à l'information dans des modes non-visuels, les contraintes de ces modes et leurs conséquences cognitives. Nous évoquerons ensuite les différentes approches existantes pour la création de contenus et interfaces accessibles, mais aussi dans quelle mesure la mise en forme est prise en compte dans ce cadre. Enfin, ce chapitre relate de l'étude exploratoire menée auprès d'élèves déficients visuels scolarisés « en inclusion », dont l'objectif était d'évaluer l'accessibilité de l'information pédagogique de manière générale. Cette partie de nos travaux s'inscrit dans une démarche ergonomique d'étude de la *situation* d'apprenants déficients visuels, des outils et contenus à leur disposition, ainsi que des contraintes auxquelles ils font face. Les résultats obtenus font écho aux différentes problématiques d'accessibilité évoquées dans ce chapitre, et notamment celle de l'accessibilité des architectures textuelles.

Le chapitre 3 constitue une application de la théorie du MAT à la formalisation de la mise en forme signifiante des documents. On y définira un ensemble de méthodes permettant de segmenter les phénomènes de mise en forme et d'extraire le sens qu'ils véhiculent. Ces méthodes sont définies dans une optique d'automatisation du processus de formalisation, et évoquent la possibilité de la mise en commun de connaissances sur la mise en forme à travers un modèle ontologique. Ce chapitre souligne la faisabilité de la généralisation de notre approche de restitution de la mise en forme des documents.

Dans le chapitre 4, nous présenterons un ensemble de réflexions sur les modalités orales utilisables dans le cadre de la restitution de la MFM, et comment les utiliser de manière optimale en prenant en compte la tâche de l'utilisateur et ses besoins. L'objectif de ce chapitre est de permettre la définition de stratégies de restitution des informations de mise en forme ainsi que des fonctions de lecture qui en dépendent, et plus généralement de fournir une base méthodologique pour concevoir un accès non-visuel à l'information dans un cadre numérique. Plus concrètement ce chapitre tente de répondre aux questions : *que restituer des informations de mise en forme précédemment formalisées ?* et *comment les restituer, quelles modalités utiliser pour quelles informations et/ou fonctions ?* La dernière partie du chapitre présente une plateforme logicielle que nous avons développée, dénommée READ (**R**Estitution de l'**A**rchitecture des **D**ocuments). Cette plateforme permet de définir différentes configurations d'accès non-visuel à des documents formatés en XML. Basé sur une oralisation du contenu par synthèse vocale, READ permet de configurer la lecture en y incorporant des d'icônes audio, du son spatialisé, du texte et différentes modulations des voix de synthèse pour la lecture. Le système est entièrement configurable et autorise la création d'extensions afin d'étendre les modalités utilisables en sortie (ex. un dispositif vibrant ou une plage braille). La flexibilité de READ fait que nous envisageons qu'il puisse être réutilisé comme plateforme expérimentale ; le logiciel sera donc prochainement publié sous une licence libre.

Enfin, le cinquième et dernier chapitre traite de l'évaluation de différentes stratégies de restitution de la mise en forme. Trois expérimentations ont été menées, dont on rapportera ici les hypothèses, méthodologies et résultats qui seront mis en perspectives ; les trois études expérimentales utilisaient une tâche de compréhension d'un texte expositif lu par une voix de synthèse, et dont la lecture était parfois augmentée, définissant ainsi plusieurs stratégies de restitution. La première expérimentation, menée auprès de 9 volontaires non-voyants, évaluait différentes stratégies de restitution des titres. La deuxième expérimentation évaluait 4 stratégies de restitution différentes avec une portée plus globale que la précédente, auprès de 60 sujets non-déficients, en termes de compréhension, charge cognitive et désorientation. Enfin, la dernière expérimentation reprenait 2 des 4 stratégies évaluées précédemment avec un protocole expérimental similaire, mais l'évaluation a été cette fois-ci menée auprès de 12 sujets non-voyants. Nos résultats ont mis en évidence que certaines stratégies permettent une meilleure compréhension des textes et une charge cognitive moindre, et donc une plus grande efficacité dans la compréhension.

Les différents chapitres du document suivent un déroulement cohérent et nous conseillons au lecteur de suivre l'ordre des chapitres. À noter que la lecture du premier chapitre conditionne fortement la compréhension des chapitres 3 et 4.

# Chapitre I - Aspects théoriques de la mise en forme des textes, et bilan empirique de ses rôles pour le lecteur

Résumé : Nous dresserons ici un panorama des travaux étudiant la mise en forme des textes, principalement en psycholinguistique et en psychologie des apprentissages, et montrerons quelles informations la mise en forme peut fournir, et quels rôles elle peut jouer pour le lecteur. Nous présenterons ensuite le Modèle d'Architectures Textuelles (MAT), qui pose un cadre théorique permettant d'analyser la mise en forme et son utilisation par l'auteur d'un texte ; ce modèle nous a permis de produire une méthodologie d'extraction de la sémantique de la mise en forme (voir Chapitre III). Cette partie pose ainsi les bases de nos travaux de recherche en tentant de cerner notre objet d'étude : la mise en forme signifiante des textes, ici appelée Mise en Forme Matérielle (MFM), qui, d'après les postulats du MAT, permet de mettre en place de véritables architectures (physiques et sémantiques) au sein des textes.

## **Sommaire du Chapitre I**

<b>1. La mise en forme des textes</b> .....	<b>8</b>
1.1 Définitions et bref historique .....	8
1.2 Le rôle de la mise en forme dans la compréhension des textes .....	9
<b>2. Le modèle d'architectures textuelles (MAT)</b> .....	<b>10</b>
2.1 Présentation .....	10
2.2 La Mise en Forme Matérielle (MFM).....	13
2.3 Analyse des intentions auteuriales.....	22
2.4 Application du MAT : exemples de modélisations d'architectures textuelles .....	23
<b>3. Conclusion</b> .....	<b>25</b>



# 1. La mise en forme des textes

---

---

## 1.1. Définitions et bref historique

---

La mise en forme d'un texte désigne, dans le langage courant, l'ensemble des propriétés visuelles du texte inscrit sur un médium d'affichage. Une définition plus formelle serait : « la *disposition graphique d'un contenu informationnel dans un espace donné* », ou encore, pour le terme synonyme mise en page : « *mise en place des éléments de composition et de typographie d'un document afin d'en obtenir un meilleur aspect visuel* »<sup>1</sup>. Si l'on s'en réfère à ces définitions on remarque que la nature de la mise en forme est purement visuelle, ce qui laisse entrevoir la difficulté de sa transposition dans des modalités non-visuelles. À noter que le terme *mise en forme* (ou *mise en page*) peut également désigner le processus de production des propriétés visuelles d'un texte ; dans nos travaux nous nous sommes intéressés non pas au processus engendrant ces propriétés, mais aux propriétés elles-mêmes.

Il est intéressant de revenir sur l'apparition historique de la mise en forme pour mieux comprendre sa nature et son importance ; cet historique est résumé par [Virbel 1985; Maurel 2006]. La mise en forme visuelle du langage écrit serait née d'un besoin (fonctionnel) de pouvoir traiter des volumes d'informations de plus en plus importants, et des contenus de plus en plus variés (dessins, schémas...) ; la mise en forme serait alors née en même temps que la démocratisation de l'écriture et que les volumes des manuscrits augmentaient. L'hypothèse selon laquelle la mise en forme permet d'aider au traitement du contenu est reprise par [Nespoulous et al. 2002], selon qui les limites des capacités cognitives humaines conduisent à *structurer* les contenus, que ce soit à l'écrit par des marques typographiques ou à l'oral par des marques prosodiques. Enfin, l'apparition de l'imprimerie ayant permis la diffusion de textes à grande échelle, elle a engendré le déplacement d'une culture de la connaissance orale vers une culture écrite, et la démocratisation de la « lecture silencieuse » [Chaytor 1945]. Le développement de la culture écrite aurait permis avec lui le développement et la complexification de la mise en forme des textes.

Nous avons évoqué le fait que la mise en forme permet aux auteurs de structurer leurs messages écrits, afin que leurs contenus puissent être plus facilement traités, si l'on en croit les raisons historiques de l'apparition de la mise en forme. Les auteurs doivent cependant respecter un ensemble de contraintes dans l'utilisation de la mise en forme, dont des contraintes éditoriales, celles du support, et surtout les principes d'utilisation de la mise en forme afin d'assurer son efficacité (voir 2.2.b), s'ils souhaitent être compris. Malgré les contraintes pesant sur la production de textes mis en forme, les auteurs disposent d'un espace de liberté plus ou moins étendu en fonction du contexte de production, et sont ainsi libres de choisir les paramètres qu'ils souhaitent (couleurs, polices, dispositions du texte, etc.) pour mettre leur texte en forme et donc lui donner une certaine structure. Cela se traduit par l'utilisation de différentes typographies et dispositions du texte et de ses différentes composantes, dont les variations sont quasi-infinies, permettant ainsi à l'auteur de structurer son texte de la manière dont il l'entend.

L'étude des phénomènes de mises en forme a donné lieu à quantité de travaux de recherches dans des domaines allant de la sémiologie à la psycholinguistique ou l'informatique. En psycholinguistique,

---

<sup>1</sup> Le Petit Robert 2007, éditions *Dictionnaires Le Robert*.

les *signaux visuels* ont été étudiés notamment pour leur aspect structurant, en pointant du doigt leur contribution à une structure *organisationnelle* du texte, aussi désignée suivant les auteurs par les termes *structure super-ordonnée*, *thématique*, *conceptuelle* ou encore *macrostructure* [Jonassen 1985; Britton et al. 1982]. Ces différents termes ne font pas forcément référence exactement aux mêmes concepts, mais incluent cependant tous l'idée de structuration de l'information textuelle. Les signaux sont qualifiés de signaux paralinguistiques par certains auteurs [Heurley 1994], ce terme mettant l'emphase sur la coexistence du visuel en parallèle du contenu linguistique ; on retrouve cette idée en linguistique et en sémiologie où l'on parle de paratexte [Peraya and Nyssen 1995; Peraya 1995], englobant cette fois les images et illustrations en plus de la mise en forme. Enfin, le terme anglais *layout* souvent retrouvé dans la littérature (dans différents domaines de recherche), semble désigner un sous-ensemble des propriétés de la mise en forme, en se focalisant sur sa propriété à effectuer des regroupements dans le texte [Wright 2000; Dyson 2004].

De par la définition de la mise en forme, et à travers les travaux de recherche l'étudiant, on entrevoit le rôle structurant qu'elle peut jouer dans les textes, avec cette idée qu'elle permet d'organiser, hiérarchiser, ou encore mettre en avant certaines informations pour permettre au lecteur de mieux interpréter le message écrit. Cette idée de structuration de l'information se retrouve d'ailleurs dans les travaux autour du **Modèle d'Architectures Textuelles** [Virbel 1985; Virbel 1989], modèle logico-linguistique sur lequel nous sommes appuyés pour nos travaux (voir section suivante 2).

## **1.2. Le rôle de la mise en forme dans la compréhension des textes**

---

Les modèles actuels de la compréhension de texte [Van Dijk and Kintsch 1983; Goldmann and Wolfe 2001] s'accordent sur le fait que comprendre un texte, c'est en construire une représentation mentale cohérente. Cela se produit à travers l'intégration de différents niveaux de représentation du texte (sa forme linguistique et sa sémantique) avec les connaissances antérieures du lecteur ; la cohérence de la représentation se construisant à la fois localement pendant la lecture pour relier les propositions adjacentes, et globalement pour relier l'ensemble aux connaissances.

Les phénomènes de mise en forme, aussi appelé *signaux* en psycholinguistique, ont pour fonctions principales du point de vue de la compréhension « d'indiquer au lecteur l'importance d'une idée, les relations que les idées du texte entretiennent, et d'aider le lecteur à sélectionner les éléments pertinents/importants du texte et à organiser ces éléments en mémoire en une structure hiérarchique ; ils constitueraient ainsi des guides pour le traitement cognitif » [Lemarié 2006a] (p.22). Des travaux ont effectivement montré que la présence de signaux dans un texte facilite la mémorisation et compréhension du texte, en aidant le lecteur à construire une représentation cohérente des thèmes d'un texte et de leur organisation [Lorch 1989]. Ils peuvent donc être vus comme des dispositifs facilitant le traitement cognitif du texte, en permettant au lecteur de plus facilement sélectionner et organiser mentalement son contenu [Kintsch 1994].

Malgré ce constat, les signaux visuels (et donc la mise en forme) n'ont bénéficié que d'un intérêt limité dans le cadre des modèles et théories de la compréhension des textes, qui considèrent le texte comme un ensemble organisé sans s'attarder sur les propriétés que leur inscription spatiale leur confère [Lemarié 2006a]. Cependant, sur le plan empirique on retrouve un ensemble de travaux autour des effets de signaux particuliers (phénomènes de mise en forme), notamment le soulignement (mise en emphase [Hartley et al. 1980]), les énumérations [Lemarié 2006a], les titres

[Wilhite 1989], ou encore l'influence des changements de paragraphes sur la vitesse de lecture (mesurée par oculométrie) [Schmid and Baccino 2001]. Ces travaux ont montré l'influence positive de la mise en forme sur la mémorisation et la compréhension ; cependant une compréhension globale des effets des signaux sur le lecteur et l'activité de lecture reste un champ de recherche à investiguer. Une avancée notable dans ce domaine a été la création d'un cadre théorique unifiant les propriétés des signaux (dites *fonctions informatives*) et leur influence sur le lecteur [Lemarié et al. 2008] ; ce modèle constituera une base de développement pour notre méthodologie de formalisation de la mise en forme matérielle (Chapitre III).

Quoi qu'il en soit, l'ensemble des travaux en psycholinguistique cités montrent que les phénomènes de mise en forme (*signaux visuels*) sont étroitement liés à la compréhension du contenu, ce qu'appuie notamment la thèse de [Schmid 2001] pour le cas des paragraphes. Cette base empirique a donc motivé nos travaux visant à donner accès à la mise en forme à l'oral pour les déficients visuels, et ce dans le but de servir, notamment, la compréhension du texte. Le Modèle d'Architectures Textuelles (MAT) a permis de formaliser la sémantique de la mise en forme et sa dimension *architecturante*, que nous allons détailler ci-après, constituant ainsi la clé de voûte soutenant notre démarche.

## 2. Le modèle d'architectures textuelles (MAT)

---

---

Dans cette partie du manuscrit sera présenté le MAT, un cadre théorique et analytique de la mise en forme des textes.

### 2.1. Présentation

---

D'après le **Modèle d'Architectures Textuelles (MAT)** [Luc and Virbel 2001; Luc et al. 2002; Virbel 1985; Virbel 1989; Pascual 1996], un texte est constitué d'un *contenu primaire* et d'une *mise en forme signifiante* (un contenu secondaire). Le contenu primaire correspond au *message* que l'auteur souhaite communiquer aux lecteurs ; il fait référence aux objets et événements du monde, et se voit *exprimé linguistiquement* d'une certaine manière à travers un choix de mots et de la syntaxe des phrases, pour former le contenu primaire. La Figure 1 ci-dessous résume les concepts importants du MAT et les liens qu'ils entretiennent.

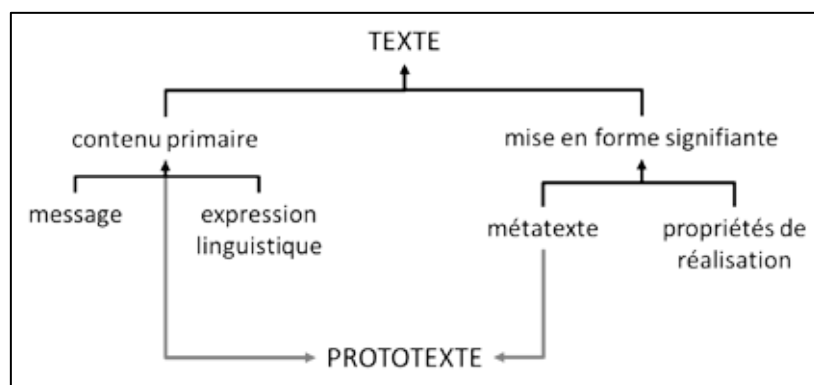


Figure 1 – Vue d'ensemble des concepts du MAT (adapté de [Lemarié et al. 2008])

La *mise en forme signifiante* des textes au sens du MAT (appelée Mise en Forme Matérielle comme nous verrons dans la section suivante) représente l'ensemble des *propriétés de réalisation* du *métatexte* au sein du texte. Nous allons développer ces deux concepts (en italique) qui méritent explication ; le concept de prototexte sera développé en suivant. Les *propriétés de réalisation* ou « marques » de réalisation incluent notamment les propriétés *typographiques* du texte, soit l'ensemble des propriétés visuelles des caractères écrits (couleur, taille, police, graisse, etc.), et les propriétés *dispositionnelles* du texte qui font référence à l'agencement spatial du texte sur la page (marges, espacements, alinéas, etc.). On parlera d'ailleurs de propriétés *typo-dispositionnelles* dans la suite du manuscrit, pour désigner l'ensemble des propriétés de réalisations typographiques et dispositionnelles du texte sur son support.

Le *métatexte* désigne simplement du texte faisant référence au texte lui-même ; ce concept est issu de la propriété de la langue à pouvoir parler d'elle-même (le métalangage au sens de [Harris 1971]). Par exemple, si je dis que « dans la phrase : je danse la samba, danser est un verbe de mouvement », la première partie de la phrase (*dans la phrase*) et la troisième (*danser est un verbe de mouvement*) peuvent être qualifiées de « métaphrases » et relèvent alors du métatexte. On pourrait également citer comme exemples de phrases relevant du métatexte : « ce manuscrit comporte 5 chapitres », ou encore « le titre de cette section est *Présentation* ». Le *métatexte* au sens du MAT désigne donc un ensemble de phrases faisant référence au texte et notamment à sa structure, son organisation, et aux relations entre différents éléments du contenu ; en un mot le *métatexte* décrit des *architectures textuelles* qui structurent et organisent le contenu primaire du texte. Les *propriétés de réalisation* permettent à l'auteur, d'après le MAT, d'exprimer la sémantique des métaphrases du métatexte, de la « réaliser », sous une forme plus ou moins typo-dispositionnelle ou plus ou moins discursive<sup>2</sup>, l'ensemble de ces formes ayant une sémantique équivalente au métatexte originel. Cela a des implications de premier plan quant à la représentation de la mise en forme du texte, dont la sémantique peut alors être exprimée sous la même forme que le contenu primaire, c'est-à-dire discursivement par des phrases.

La réalisation du métatexte à l'aide de marques typo-dispositionnelles engendre des *objets textuels*, c'est-à-dire des segments de texte qui peuvent être distingués visuellement du reste du texte par *contraste* (procédé détaillé en 2.2.b), comme des titres, des énumérations, des légendes, des notes de bas de page, etc. Lorsque la réalisation de l'ensemble des métaphrases du métatexte se fait dans le texte sous forme discursive, sans aucune marque typo-dispositionnelle, on dira alors que le texte est un *prototexte* (voir Figure 1 ci-avant), dans le sens où cette forme du texte est la forme initiale, la « première » forme (*proto* vient du grec *protos* qui signifie « premier »), à partir de laquelle un ensemble d'autres réalisations du texte incorporant différentes marques typo-dispositionnelles peuvent être dérivées.

---

<sup>2</sup> « Discursif » sera dans ce document employé au sens qu'on lui donne en linguistique c'est-à-dire : qui se rapporte au discours. On utilisera « discursif » plutôt que « verbal », ce dernier ayant plutôt trait au mode d'expression de l'information qu'à son caractère énonciatif.

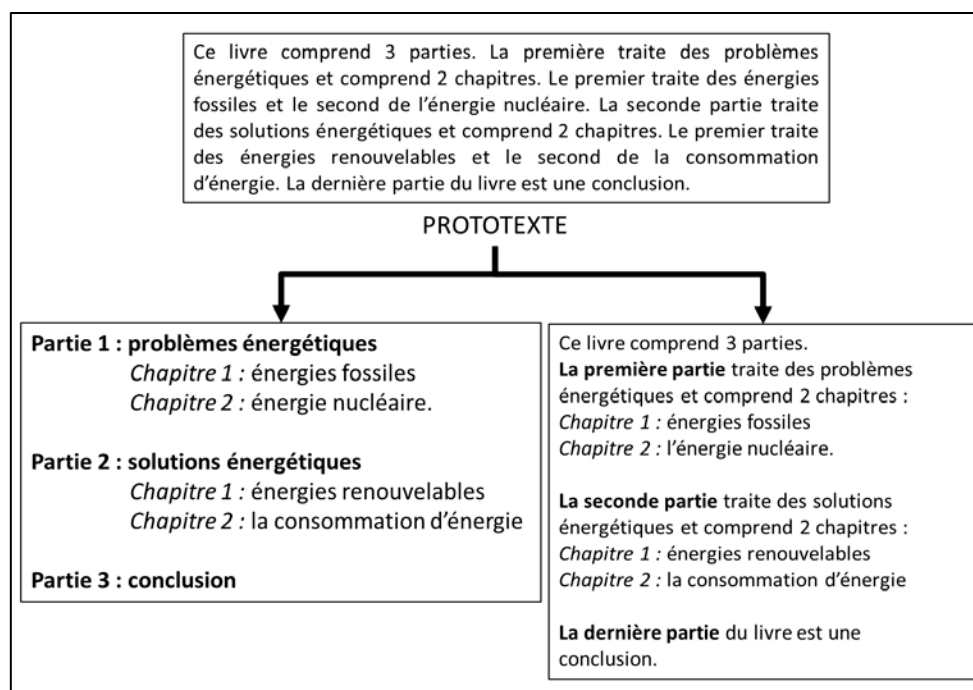


Figure 2 - Un prototexte et deux exemples dérivés mis en forme (adapté de [Lemarié et al. 2008])

Le postulat central du MAT, qui rejoint le concept de *prototexte*, est que le métatexte de l'auteur peut s'exprimer sous une quasi infinité de formes *réalisées* (à l'aide de marques typo-dispositionnelles) différentes, mais toutes équivalentes en termes d'informations fournies au lecteur. Ces formes peuvent être plus ou moins discursives ou plus ou moins typo-dispositionnelles ; ce qui est postulé est que le discursif peut être « réduit » sous forme typo-dispositionnelle, et les indices typo-dispositionnels peuvent être « développés » sous forme discursive. On parlera dans le premier cas de *réduction (du métatexte) au profit des traces (typo-dispositionnelles)*, et dans le second de *développement discursif (de la mise en forme)*. Pour illustrer ce postulat, ci-dessous un exemple de 4 formes équivalentes d'une même phrase comportant du métatexte :

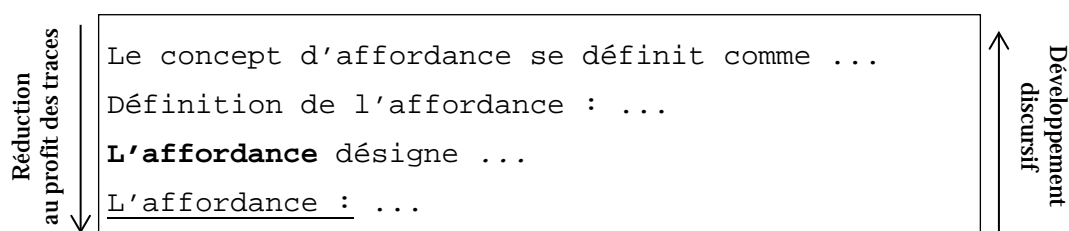


Figure 3 - Exemples de phénomènes de mise en forme équivalents

Dans l'exemple précédent, on retrouve des indices discursifs métatextuels dans plusieurs des formes (les termes « définition », « comme ») qui sont effacés par réduction au profit de marques typo-dispositionnelles (soulignement, et grasse). À noter que ces deux processus de réduction visuelle et de développement discursif permettent la production d'un ensemble de formes équivalentes mais qui ne sont pas forcément ordonnables, en d'autres termes dont le degré de développement ou de réduction ne peut être évalué avec objectivité.

Ainsi, le MAT postule que la mise en forme typo-dispositionnelle des textes est issue d'une réduction du métatexte [Virbel 1985] (qui est implicite<sup>3</sup> pour l'auteur pendant la rédaction), et qu'à l'inverse la mise en forme d'un texte peut être exprimée de manière discursive, dans les deux cas en conservant une équivalence informationnelle pour le lecteur, mais pas nécessairement une équivalence cognitive/fonctionnelle (nous le verrons dans la section 2.2). Cette propriété des marques typo-dispositionnelles à pouvoir être développées discursivement sous forme de métatexte nous a permis dans notre cas de les formaliser, même si cette propriété en elle-même était insuffisante pour ce faire (voir 2.3 et la méthodologie de formalisation de la MFM décrite au Chapitre III).

Un autre postulat important du MAT est que la signification du texte n'est pas déterminée uniquement par son contenu primaire mais aussi par son métatexte. Autrement dit, le MAT postule que la mise en forme (expression réalisée du métatexte) fait partie intégrante du texte et donc que le contenu d'un texte et sa réalisation sur son support sont indissociables [Virbel 1989; Pascual 1996]. On remarque d'ailleurs que [Luc 2000] a produit une modélisation des textes superposant les architectures textuelles du MAT et les *relations rhétoriques*<sup>4</sup> du contenu, avec la RST (Rhetorical Structure Theory [Mann and Thompson 1987]); ces travaux soulignent les liens forts entre le contenu propositionnel du texte (contenu primaire) et sa mise en forme. On en conclut qu'il est important d'analyser le discours du contenu primaire en parallèle des marques de mise en forme afin d'assurer une formalisation des architectures textuelles cohérente et complète. Cela paraît d'autant plus important que le lecteur utilise à la fois les marques de mise en forme, le contenu primaire ainsi que le contexte d'utilisation du document (type de document, média, etc.) pour reconnaître les intentions de l'auteur ; ce dernier postulat est discuté dans la section 2.3.

Pour terminer cette section nous citerons quelques applications du MAT dans des travaux antérieurs. Le MAT a été utilisé initialement pour de l'analyse linguistique des structures de textes [Virbel 1985; Virbel 1989] ainsi que pour l'analyse de certains objets comme les titres [Virbel 2002], les définitions [Pascual and Péry-woodley 1991] ou encore les énumérations [Luc et al. 1999]. Le MAT a également été utilisé pour la prise en compte de la mise en forme dans la génération informatique de texte [Pascual 1996; Chali 1997; Luc 2000]. Enfin, le MAT a été utilisé en psychologie cognitive et en psycholinguistique pour permettre d'étudier plus formellement la mise en forme et son impact cognitif sur le lecteur [Lemarié et al. 2008].

## **2.2. La Mise en Forme Matérielle (MFM)**

---

### ***2.2.a. Définitions et formalisation***

---

D'un point de vue descriptif, la Mise en Forme Matérielle (MFM<sup>5</sup>) désigne l'ensemble des propriétés signifiantes de réalisation du métatexte au sein du texte, à savoir des propriétés *typographiques*

---

<sup>3</sup> Cette réduction est le plus souvent implicite durant le processus de production d'un texte : l'auteur va utiliser directement une certaine mise en forme dans le texte sans passer par une forme discursive intermédiaire. Ses intentions (nous verrons dans la section suivante pourquoi l'on parle d'intentions) sont ainsi traduites directement sous forme de marques typographiques, dispositionnelles et/ou lexico-syntaxiques.

<sup>4</sup> Les relations rhétoriques représentent des relations entre segments du contenu primaire du texte sous-tendant la cohérence globale de sa sémantique. Par exemple, pour une affirmation suivie de la démonstration étayant cette affirmation, la RST postule une relation de "démonstration" entre les deux segments.

<sup>5</sup> L'acronyme sera utilisé de manière récurrente dans le reste du document pour désigner notre objet d'étude qu'est la mise en forme signifiante des textes.

(police, taille, graisse, etc.), *dispositionnelles* (alignement, alinéa, interligne, positionnement, etc.) mais aussi *lexico-syntaxiques*. Concernant le dernier type de propriétés de réalisation (lexicales et syntaxiques), que nous n'avons pas encore abordé, elles correspondent aux traces issues de la réalisation des métaphrases du métatexte, de la même façon qu'elle peut laisser des marques typo-dispositionnelles ; les deux ne sont pas mutuellement exclusifs, la réalisation d'une métaphore pouvant laisser à la fois des marques typo-dispositionnelles et lexico-syntaxiques. Dans l'exemple de prototexte présenté sur la Figure 1, où la réalisation des métaphrases est entièrement discursive, on retrouve de nombreux indices lexicaux et syntaxiques (ex. « première », « seconde partie traite de... »), que l'on retrouve pour certains dans les formes dérivées qui incluent des marques typo-dispositionnelles (ex. « Seconde partie : ... »). Quelle que soit leur nature, les marques de MFM (de réalisation du métatexte) permettent de véhiculer des architectures textuelles, qui dans le cas du prototexte de la Figure 1 correspond à une énumération des différents chapitres et parties d'un texte. Afin de resituer les nouveaux concepts introduits, la Figure 4 reprend la Figure 1 en y intégrant les concepts d'architectures textuelles, d'objets textuels et de marques de MFM.

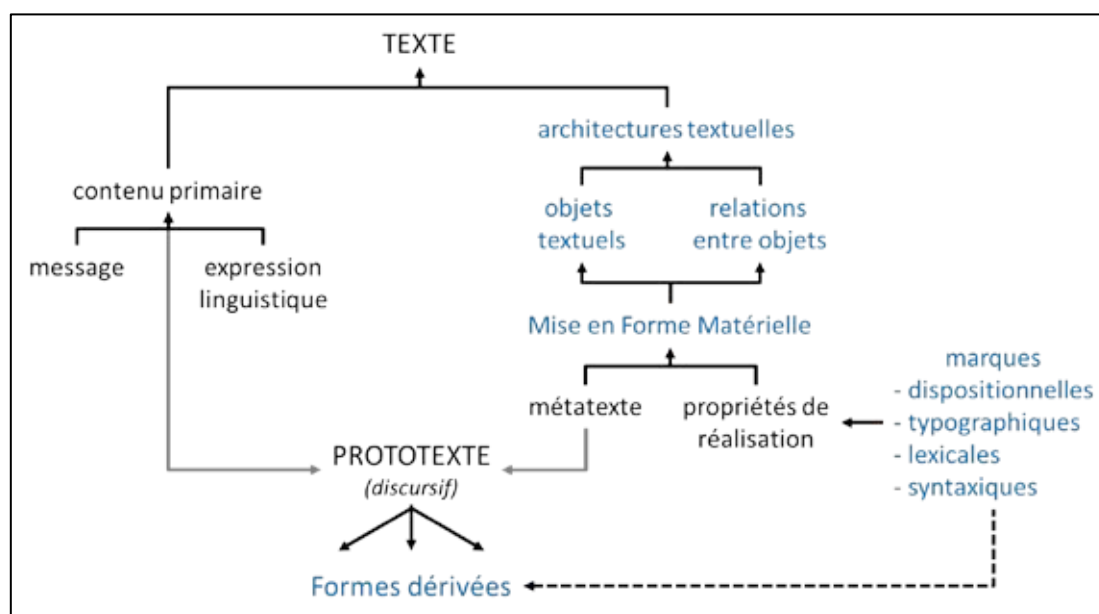


Figure 4 – Schéma des concepts importants du MAT : seconde version

Le terme « signifiant » a été souligné dans la définition de la MFM étant donné qu'il met l'accent sur un concept clé du modèle, qui est que les marques de MFM sont signifiantes car elles constituent pour l'auteur un moyen d'expression de ses intentions relatives à la *structuration* du contenu primaire de son texte. En d'autres termes, la MFM et les marques correspondantes représentent un moyen pour l'auteur de rendre visibles et interprétables des *architectures* dans son texte, c'est-à-dire des objets textuels (phénomènes de MFM) et des relations entre ces objets (ex. relations hiérarchiques). Le nom du modèle « *d'architectures textuelles* » vient donc du fait qu'il offre un cadre théorique pour l'analyse de la MFM, qui a une fonction architecturante au sein des textes. Par la suite, on utilisera l'acronyme MFM pour désigner le concept de manière générale, mais aussi dans d'autres expressions pour désigner des *marques* de MFM (indices typo-dispositionnels et/ou lexico-syntaxiques) présentées par les objets textuels et pouvant marquer des relations entre objets, ou encore pour parler de *la* MFM comme l'ensemble des *marques de MFM* d'un texte et des architectures textuelles correspondantes.

Attention cependant, toute marque typo-dispositionnelle du texte n'est pas forcément signifiante (ne participe pas à des *architectures textuelles*); en effet ces marques peuvent provenir de contraintes du support, d'erreurs diverses, de contraintes éditoriales. Nous mettrons de côté ces cas particuliers dans nos travaux pour considérer uniquement le cas prototypique où toute mise en forme est signifiante.

Les intentions exprimées par la MFM peuvent être explicitées par *développement discursif* sous forme de *métaphrases*. Les métaphrases constituent ainsi des équivalents discursifs aux phénomènes de MFM. Dans les termes de la théorie des actes de langages [Austin 1975] sur laquelle est basée le MAT, les métaphrases sont des énoncés *performatifs*<sup>6</sup> dont la force *illocutoire*<sup>7</sup> est tournée vers le texte ; elle décrivent donc une action de l'auteur sur le texte. Une métaphrase correspond donc à la forme discursive d'une marque de MFM du texte, exprimant l'action *architecturante* de l'auteur sous-jacente à cette réalisation. Une propriété importante des métaphrases est qu'elles constituent des équivalents aux formes réalisées qu'elles décrivent mais uniquement sur le plan informatif. En effet, la MFM d'un texte sur son support possède une dimension fonctionnelle qui n'est pas capturée par les métaphrases, que nous décrivons dans la section 2.2.c.

En résumé, lors de la rédaction l'auteur exprime son intention d'informer le lecteur de la structure de son contenu primaire à travers des marques de MFM constituant des architectures textuelles, pouvant prendre une forme plus ou moins discursive ou plus ou moins typo-dispositionnelle, qui constituent pour l'auteur un moyen d'expression de ses intentions. Le processus de développement discursif permet de reformuler ces intentions sous forme de métaphrases et donne ainsi un point de départ, une *proto-forme* de la MFM, qui peut être ensuite déclinée de différentes manières tout en conservant sa dimension informative. L'un des objectifs de nos travaux est justement d'explorer la reformulation des métaphrases dans des modalités non-visuelles, au-delà des marques typo-dispositionnelles et lexico-syntaxiques (voir Chapitre IV).

Un schéma syntaxique général a été défini pour les métaphrases du métatexte architectural décrivant la MFM des texte [Virbel 1985; Virbel 1989; Pascual and Virbel 1996], schéma permettant d'inventorier toutes les expressions discursives possibles associées aux architectures textuelles.

`<act> <V> <prep> <act> 'que' [<act> <Vpf> <d> <OT>]`

**Figure 5 - Schéma syntaxique général des métaphrases**

Ce schéma fait référence aux acteurs textuels (<act>), c'est-à-dire tous les intervenants dans la mise en forme d'un texte, ainsi que le lecteur. En effet l'auteur n'est pas le seul à intervenir sur la mise en forme ; dans un processus éditorial on peut notamment citer le typographe, l'éditeur ou encore le metteur en page [Mojahid 1990]. Cette prise en compte des différents acteurs était importante notamment dans le cadre de la génération de texte [Luc 2000; Pascual 1996] ou de l'analyse linguistique [Virbel 1985], mais ne l'est plus dans notre cas étant donné que nous nous intéressons à la compréhension du texte par le lecteur. Cela a pour conséquence que la partie du schéma signalée en bleu sur la Figure 5 ne nous concerne pas (ex. « *L'auteur indique au typographe que [...]* »). De

---

<sup>6</sup> Réalisant une action, par opposé aux énoncés constatifs.

<sup>7</sup> Austin distingue différents types d'actes qui peuvent coexister dans un même énoncé : les actes locutoires (produire un énoncé), les actes illocutoires (l'acte réalisé en disant l'énoncé, pouvant être implicite) et les actes perlocutoires (les effets de la réalisation de l'énoncé sur le récepteur du message).



plus, on peut considérer dans le cadre de nos travaux que tout élément de MFM est destiné au lecteur, de la part de l'auteur. Ainsi nous mettons de côté cette première partie des métaphrases en considérant qu'elles sont toutes de la forme : « L'auteur informe le lecteur que [...] ». Cette restriction a également été appliquée par [Pascual 1991] dans son recensement des métaphrases du MAT.

En revanche, la seconde partie du schéma syntaxique (<acteur>, <verbe performatif>, <lien de dépendance>, <objet textuel>), fait référence à une action d'un acteur (que nous considérerons être l'auteur) sur un objet textuel (<OT>), c'est-à-dire un élément du texte distinguable par sa MFM, à travers un verbe performatif. Par exemple : « L'auteur <développe> <un paragraphe P0> ». Des dérivées de ce schéma peuvent également être instanciées, par exemple pour mettre en relation deux ou plusieurs objets textuels : « L'auteur <annote l'image IMO par la légende LO> ». On notera que l'on fait référence aux objets textuels dans les métaphrases à l'aide d'identifiants définis dans le métatexte. Une dernière remarque importante est qu'il est possible de créer des objets textuels dans le métatexte qui ne seront pas réalisés dans le texte par de quelconques marques, et qui seront alors tout de même considérés comme des objets textuels, mais sans marques de MFM associées. Ce type d'objet textuel est issu de conventions particulières de structuration des textes, qui sont identifiables sans marques de MFM : par exemple une section ou un chapitre ne se distinguent pas visuellement du reste du texte (un chapitre s'étendra sur plusieurs pages), mais par convention ; on retrouve cependant ces objets dans le métatexte : « L'auteur <compose <sectionS0> par <paragraphe0> et <titre1> ».

Une autre caractéristique du métatexte du MAT est qu'il distingue le contenu textuel brut, des objets textuels auxquels il peut être rattaché ; dans les exemples précédents, on a distingué les objets textuels labellisés P0, IMO et LO sans faire référence au contenu associé à ces objets. Ainsi, en parallèle de la segmentation du texte en objets textuels d'après les marques de MFM, le texte est segmenté en *unités textuelles* qui sont ensuite rattachées aux objets textuels identifiés par la métaphore « L'auteur attache <identifiant d'une unité textuelle> à <identifiant d'un objet textuel > ». Ce cadre théorique comporte certaines similitudes avec des paradigmes informatiques comme la séparation des données de leur traitement et de leur présentation, ou encore les langages de description de ressources utilisés dans le web sémantique comme le RDF<sup>8</sup>. Ces similitudes pourraient faciliter la modélisation informatique de la mise en forme sur la base d'outils et techniques existants [Sorin et al. 2013].

---

<sup>8</sup> Resource Description Framework.

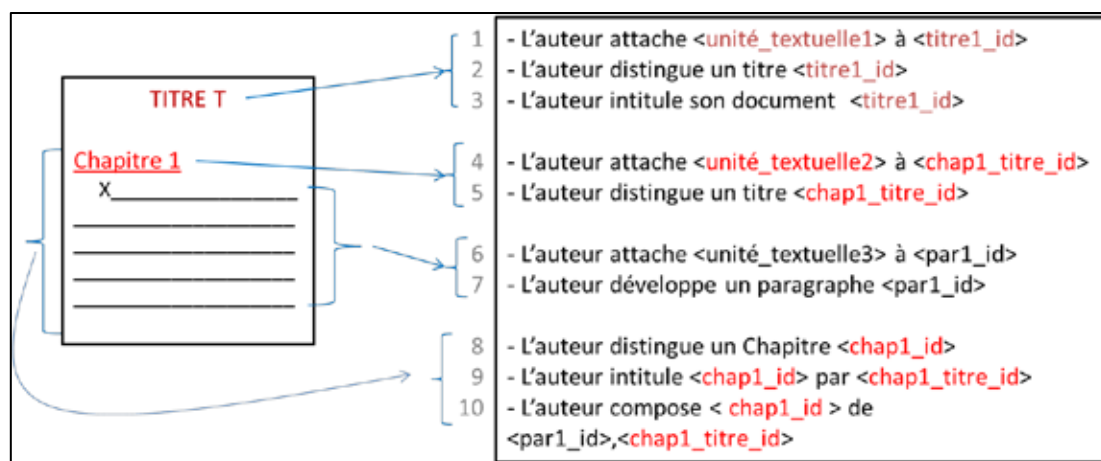


Figure 6 - Exemple de modélisation de la MFM sous forme de métaphrases

Dans cet exemple, les métaphrases décrivent les objets textuels marqués par leur MFM ainsi que les relations entre ces objets ; l'aspect *relationnel* entre objets textuels permet de décrire par exemple l'imbrication d'objets textuels (ex. un objet en compose un autre) ou d'autres types de relations comme le fait qu'un objet soit une illustration d'un autre. On remarquera enfin qu'un objet textuel est souvent décrit par plusieurs métaphrases, il n'y a à ce propos pas de règle particulière.

En ce qui concerne les métaphrases existantes, plusieurs classes de métaphrases avec leurs particularités ont déjà été définies [Luc 2000], chaque classe de métaphore représentant un ensemble de métaphrases équivalentes sur le plan sémantique. Pour autant, le modèle reste ouvert étant donné que la création de nouvelles architectures textuelles ou la modification d'architectures existantes, potentiellement génératrices de nouvelles classes de métaphrases, est chose possible ; la nécessité de conserver un modèle ouvert provient également de l'immense diversité d'architectures que l'on retrouve dans la nature, qu'il est difficile voire impossible de modéliser dans leur globalité. Par exemple des objets textuels ou des relations particulières peuvent apparaître dans des documents spécifiques à certains domaines de connaissance (ex. la biologie) ou certains contextes d'utilisation (ex. les notices de montage).

Enfin, il est intéressant de noter que développer discursivement les phénomènes de MFM sous forme de métaphrases, permet de s'affranchir d'une certaine complexité des marques de MFM issues de la réalisation du métatexte : en effet, cette complexité provient du fait que pour une même métaphore d'origine, on peut générer une quasi-infinité de variations de marques de MFM correspondantes. Or, d'après le MAT, l'ensemble de ces variations pourront être explicitées par la même métaphore. Les métaphrases permettent donc de réduire cette complexité en formalisant les architectures ; encore faut-il qu'elles puissent être correctement interprétées. Pour s'affranchir de cette difficulté dans le cadre de la thèse, nous avons développé une méthodologie semi-automatique de formalisation qui est l'objet du Chapitre III.

### **2.2.b. Les principes de contraste et d'identité régissant l'utilisation des marques de MFM**

Nous avons déjà mentionné le fait que l'auteur d'un texte disposait d'une certaine liberté quant à la réalisation de la mise en forme de son texte, dans le choix des marques de MFM à utiliser pour réaliser le métatexte, et véhiculer les architectures textuelles qu'il souhaite utiliser pour structurer son texte. Même si certains objets textuels répondent à des conventions vastement partagées comme les titres, certains phénomènes de MFM sont spécifiques à un genre de documents (ex.

articles scientifiques) voire « locales » à un seul document. Ainsi l'utilisation des marques de MFM n'est pas codifiée de manière rigide, et l'on peut retrouver des phénomènes particuliers de MFM spécifiques à un seul document, qui ne sont donc partagés par aucun autre. Leur interprétation est tout de même possible grâce aux **principes de contraste et d'identité** régissant l'utilisation des marques de MFM.

Le MAT stipule que l'utilisation de la MFM doit respecter les principes *d'identité* et de *contraste* entre objets textuels. Ces principes énoncent que des marques de mise en forme *identiques* doivent être utilisées pour désigner un même objet textuel ou une même relation entre objets, et à contrario que des marques différentes (*contrastées*) doivent être utilisées pour désigner des objets ou relations différentes. Formulé autrement, une même intention auctoriale doit être réalisée à l'aide des mêmes marques de MFM dans le texte, et des intentions différentes doivent être réalisées par des marques différentes. On retrouve la notion d'identité des phénomènes de MFM dans les logiciels d'édition de texte, qui permettent la création et l'utilisation de styles de texte afin justement d'assurer l'identité des objets textuels.

Le contraste et l'identité permettent donc d'assurer la cohérence des architectures textuelles utilisées par l'auteur et ainsi leur bonne interprétation par le lecteur ; ils sont d'ailleurs au cœur de la méthodologie de formalisation de la MFM développée.

### ***2.2.c. Les deux dimensions de la MFM pour les lecteurs***

---

L'utilisation par l'auteur de marques de MFM sert d'abord un objectif **informatif**, mais également dans certains cas un objectif **fonctionnel** ; le premier a trait à informer le lecteur des architectures textuelles, objets textuels et relations entre objets (informations décrites par les métaphrases), tandis que le second permet au lecteur de mettre en place certaines fonctions de lectures sur la base des marques de MFM. Formulé autrement, ces deux aspects correspondent à l'information donnée par la mise en forme et comment cette mise en forme peut être utilisée par le lecteur dans le cadre de son activité de lecture.

En effet le premier rôle des marques de MFM est d'informer le lecteur sur les architectures textuelles, ce qui dans le cadre du modèle se traduit par l'équivalence entre une marque de MFM et les métaphrases correspondantes. Mais cette équivalence se limite à une équivalence informationnelle, ce qui nous amène à la dimension fonctionnelle des marques de MFM. Le MAT stipule que certaines variations de réalisation du métatexte sont fonctionnelles [Lemarié 2006a], c'est-à-dire qu'à partir d'une même métaphore, toutes les réductions visuelles n'offriront pas les mêmes fonctions de lecture. Un exemple permet de comprendre aisément cette différence : un retour à la ligne doublé d'un espacement vertical à la fin de chaque paragraphe permet au lecteur de savoir rapidement combien il y a de paragraphes dans son champ de vision, la taille de ces paragraphes, lorsque la fin du paragraphe arrive pendant la lecture, etc. Cependant, si l'on signale l'enchaînement des thèmes par des indices discursifs (ex. « Et maintenant nous allons aborder... ») au lieu de marques typo-dispositionnelles pour distinguer des paragraphes (ex. alinéas et retours à la ligne), on conserve l'équivalence informative entre le métatexte et sa réalisation, mais l'équivalence fonctionnelle, elle, n'est pas effective. En effet le marquage discursif est suffisant pour percevoir les changements de thèmes lors de la lecture, mais ne permet pas une appréhension visuelle globale du texte, par exemple pour compter le nombre de paragraphes ou évaluer leurs tailles respectives. À travers cet exemple, on comprend que la dimension fonctionnelle a des conséquences sur la capacité

des lecteurs à mettre en place des stratégies de traitement de l'information, et a donc un impact cognitif.

Sur la base de l'existence d'une dimension fonctionnelle des marques de MFM, il est légitime de supposer que les intentions de l'auteur peuvent être de l'ordre du fonctionnel en marge de la dimension informative selon laquelle la MFM a été présentée ; par exemple l'auteur peut vouloir faire en sorte qu'il soit le plus aisé possible de trouver une information dans son texte, et choisir alors des marques typo-dispositionnelles plutôt que lexico-syntaxiques pour structurer son texte. Or, les fonctions de lecture s'appuyant sur la MFM *servent* certains objectifs de lecture, comme la recherche d'information ou la compréhension du texte, qui ne sont pas spécifiques aux lecteurs voyants et peuvent tout aussi bien être recherchés par une personne déficiente visuelle accédant au texte. Ainsi l'étude de la dimension fonctionnelle de la MFM semble pertinente dans le cadre de nos travaux. L'objectif n'est pas de mimer ces fonctions de lecture visuelle, mais de comprendre leurs rôles pour la lecture, ainsi que le rôle des marques de MFM dans la mise en œuvre par le lecteur de ces fonctions (i.e. quelles fonctions peuvent émerger de la présence d'une marque de MFM donnée?). Ultimement, nous souhaitons proposer des fonctions de lecture équivalentes dans des modes non-visuels, c'est-à-dire permettant d'atteindre les mêmes objectifs de lecture tout aussi efficacement. Dans les travaux rapportés ici nous nous sommes limités à certaines fonctions de lecture qui seront présentées dans le Chapitre IV. On présentera cependant dans la section suivante une liste non-exhaustive des fonctions de lecture basées sur la MFM.

Il n'existe pas à notre connaissance dans la littérature de recensement de telles fonctions de lecture en lien avec le système visuel<sup>9</sup> ; c'est-à-dire quelles sont les fonctions de traitement de l'information et les fonctions de navigation qui émergent de l'interaction entre le système visuel humain et les marques de MFM dénotant d'architectures textuelles. Nous savons cependant, comme cela a été évoqué précédemment, que sur le plan cognitif la mise en forme influence la compréhension et la rétention du contenu d'un texte [Lorch 1989; Lemarié 2006a] ; nous verrons également dans la section suivante que les marques de MFM permettent aux lecteurs de mettre en œuvre différentes « fonctions » de navigation visuelle dans les textes. Dans le cadre de nos travaux visant à restituer les architectures textuelles, il nous paraît important de prendre en compte non seulement la sémantique de la MFM à travers les métaphrases, mais aussi la dimension fonctionnelle de la MFM.

### ***2.2.d. L'aspect fonctionnel de la MFM & le système visuel humain***

---

Tout l'aspect fonctionnel des marques de MFM repose sur les propriétés du système visuel humain, ce qui semble raisonnable étant donné leur nature visuelle. L'on pourrait même formuler l'hypothèse que, la MFM ayant émergé (*à priori*, voir 1.1) d'un besoin fonctionnel de traitement, cette dernière se soit développée et raffinée au fil du temps en fonction des capacités du système visuel, pour justement servir des besoins fonctionnels. Cet argument expliquerait l'émergence de différents types d'objets et architectures textuelles spécifiques à certains domaines de connaissances ou cadres d'utilisation (ex. un menu de restaurant ne présentera pas les mêmes architectures et marques de MFM qu'un manuel de mathématiques ou qu'un roman). Dans cette section nous verrons comment la MFM est exploitée à des fins fonctionnelles par le système visuel, pour naviguer au sein des textes et pour optimiser le traitement de son contenu.

---

<sup>9</sup> Ce terme fait référence à l'ensemble des sous-systèmes de perception et de traitement de l'information visuelle chez l'humain.

La fonction principale de la MFM, émergeant de son exploitation par le système visuel, est sa capacité à effectuer des regroupements visuels de texte, à travers le *contraste* et *l'identité* (voir 2.2.b) entre les marques typo-dispositionnelles de MFM<sup>10</sup>, permettant de distinguer les objets textuels et de naviguer entre eux (entre regroupements). C'est bien grâce au système visuel que ces regroupements émergent si l'on s'en réfère aux lois du regroupement visuel du *Gestalt* [Wertheimer and Riezler 1944], qui stipulent notamment que la perception visuelle tend à regrouper des formes (des stimuli visuels) en fonction de leur similitude et de leur proximité ; d'autres lois sont à l'œuvre dans un cadre plus général mais seulement les deux lois précédentes nous intéressent dans le cadre de la mise en forme des textes (nous avons détaillé ces lois dans le Chapitre IV). Ainsi, en définissant des objets textuels par identité et contraste, l'auteur permet au lecteur d'effectuer des regroupements d'informations, ce qui a de nombreuses conséquences fonctionnelles<sup>11</sup> :

- § L'identification visuelle des regroupements (objets textuels) [Wright 2000], ainsi que leur typage et leur mise en relation à travers l'interprétation de leurs marques de MFM. Cette fonction est dérivée directement de la propriété principale des marques de MFM qui est d'informer le lecteur, à travers des indices typo-dispositionnelles ou lexico-syntaxiques. Elle permet au lecteur d'effectuer des inférences à un niveau macrostructural. Exemple : *en abordant ce paragraphe, de par sa MFM je sais qu'il peut être typé comme un item d'une liste à puces.*
- § Le typage des objets textuels (cf. point précédent de la liste) permet au lecteur de contextualiser l'information au cours de la lecture, grâce à sa vision périphérique et aux marques de MFM de la partie du texte lue. La fenêtre perceptive des propriétés typo-dispositionnelles du texte en lecture visuelle serait d'ailleurs plus étendue que la fenêtre perceptive à l'intérieur de laquelle l'information verbale utile est extraite [Cauchard 2008]. Exemples : *en lisant cette phrase je peux savoir qu'elle fait partie d'un item d'une liste à puces ; je peux repérer que j'arrive à la fin de la phrase qui est également la fin de l'item.*
- § Les objets textuels pouvant être imbriqués visuellement, une hiérarchie de différents types d'objets peut ainsi être définie. Exemple : *je peux inférer des marques de MFM de ce paragraphe que cette phrase est un exemple dans un item d'une liste à puces.*
- § La saillance des marques de MFM (définies par contraste) peut leur permettre de jouer le rôle d'ancres, de repères pour la navigation visuelle dans le texte. Exemple : *je peux rapidement naviguer vers le premier item de cette liste en utilisant les puces et l'indentation.* Les indices lexico-syntaxiques peuvent également jouer ce rôle d'ancres pour la navigation, par exemple pour revenir en début de phrase. De cette fonction en découlent plusieurs autres :
  - l'accès sélectif, c'est-à-dire le fait de pouvoir accéder un type d'objet textuel en particulier en fonction de ses marques de MFM. Exemple : *je peux chercher le titre le plus proche de ce segment de texte ;*
  - pour la lecture en diagonale (aussi appelée *skim-reading* ou *scanning* dans la littérature [O'Hara 1996]), les marques de MFM peuvent permettre divers types d'inférences. Exemple : *je peux savoir rapidement combien il y a d'items dans cette liste ; je peux savoir rapidement combien il y a de références dans le paragraphe*

---

<sup>10</sup> On exclue ici les marques lexico-syntaxiques qui n'ont pas la propriété de permettre au lecteur d'effectuer des regroupements visuels (au sens des lois du regroupement visuels du *Gestalt*).

<sup>11</sup> Le recensement des fonctions dans cette liste n'a que valeur d'exemple.

*suivant ; je peux également scanner la page pour voir quels types d'objets s'y trouvent.*

L'ensemble des fonctions citées précédemment permettent au lecteur d'effectuer des inférences et/ou de mettre en place certaines actions de navigation visuelle. Par exemple, la littérature montre que les marques de MFM ont des répercussions sur la recherche d'information [Waller 1987; Etcheverry et al. 2012]. Cependant ces inférences et actions de navigation ont des conséquences cognitives, sur le traitement et la représentation du contenu. Par exemple, parcourir les titres d'un document ou les mots soulignés dans les paragraphes est rendu possible grâce à leur MFM, mais ce parcours aura des conséquences importantes pour la représentation globale du contenu et donc pour la compréhension. Certaines études et résultats de la littérature mettent en lumière le versant *cognitif* de la dimension fonctionnelle de la MFM :

- § Le fait même que le texte ait une inscription spatiale, c'est-à-dire qu'il soit inscrit sur un support, permet un codage temporel de l'information (de séquentialiser les propositions du discours) et permet également une réinspection sélective efficace du texte [Baccino and Pynte 1991], les deux étant rendus possibles à travers un souvenir implicite de la localisation des informations du texte [Zechmeister et al. 1975].
- § La vision périphériques et les marques typo-dispositionnelles de MFM permettent de donner un contexte à l'information lue (voir la liste de fonctions précédente). [Cauchard 2008] a d'ailleurs montré que les marques de MFM permettent d'exploiter une fenêtre perceptive des propriétés typo-dispositionnelles du texte plus étendue que la fenêtre perceptive à l'intérieur de laquelle l'information verbale utile est extraite. Les marques de MFM typo-dispositionnelles permettent donc de donner un contexte « visuel » étendu à la lecture. Ce contexte visuel permet aux lecteurs d'anticiper la fin des paragraphes lorsqu'elle est marquée par un indice dispositionnel ; cette anticipation permet cognitivement de mettre en jeu des processus d'intégration de l'information du paragraphe lu avant de poursuivre la lecture [Schmid and Baccino 2001]. Ces auteurs ont montré la plus grande efficacité cognitive de cette configuration (paragraphes distingués par des marques typo-dispositionnelles) sur une configuration sans démarcation des paragraphes (un seul bloc de texte).
- § D'après le modèle en vigueur de la mémoire de travail (cf. modèle de [Baddeley 2003], présenté au Chapitre II), il existe des unités de traitement séparées pour l'information visuo-spatiale et l'information phonologique, qui fonctionnent en parallèle. On peut ainsi formuler l'hypothèse que la coexistence sur le support des marques typo-dispositionnelles et du texte permet une optimisation du traitement de l'ensemble en mémoire de travail [Maillet 2015], étant donné que l'information verbale (issu de l'interprétation du texte) et le visuel (interprétation des marques typo-dispositionnelles de MFM) sont traités par des entités séparées en mémoire de travail.

Il n'existe pas à notre connaissance une typologie des fonctions de lecture relatives à la MFM et leur impact sur la navigation et le traitement du contenu textuel. Cela pourrait s'expliquer tout d'abord par la difficulté de formalisation de cette dimension fonctionnelle et donc la difficulté de son évaluation : une marque de MFM peut servir plusieurs fonctions de lecture. De plus, les frontières entre les différentes fonctions telles que nous les avons présentées peuvent se recouper. Enfin, le caractère quelque peu implicite de ces fonctions fait que même les auteurs ne prennent pas

forcément cette dimension en compte dans leur mise en forme, ou alors de manière non-intentionnelle (ex. en mettant un mot en gras, l'intention initiale de l'auteur pouvait être de marquer un mot clé, et non de permettre au lecteur de scanner la page à la recherche des concepts importants). Une autre explication réside dans le manque d'applications de telles études portant sur la dimension fonctionnelle de la MFM ; en effet, la MFM représentait jusqu'ici un objet d'étude difficilement formalisable, dont on cherchait à comprendre l'influence sur plusieurs variables de performance relatives à la lecture (compréhension, mémorisation, charge cognitive, recherche d'information, etc.), pour pouvoir en optimiser l'utilisation, notamment pour l'apprentissage, mais dont les processus en lien avec le système visuel représentaient un niveau d'étude trop éloignés de ces applications pour être pertinent. Concernant ce dernier point, la restitution de la dimension fonctionnelle de la MFM dans des modalités non-visuelles, pour l'accessibilité, semble une application suffisante pour être moteur d'investigations futures. Nous proposons d'ailleurs au Chapitre IV une restitution de plusieurs fonctions de MFM dans des modes non-visuels.

En conclusion, il est clair que la MFM doit être considérée pour la restitution à la fois à un niveau informationnel (pour améliorer la disponibilité d'informations métatextuelles importantes pour la compréhension du discours) et à un niveau fonctionnel (pour améliorer l'accessibilité du contenu).

### **2.3. Analyse des intentions auteuriales**

---

Cette section vise à introduire la problématique de l'interprétation des marques de MFM, et ainsi montrer la nécessité de définir une méthodologie d'annotation (présentée au Chapitre III).

Nous avons vu qu'il était possible, dans le cadre du MAT, de ramener la sémantique des marques de MFM à une forme discursive équivalente : les métaphrases. Le modèle ne détaille cependant pas le processus d'interprétation des intentions de l'auteur, ayant donné lieu aux marques de MFM du texte. Il s'agit en effet d'un processus *d'interprétation*, étant donné qu'une même marque de MFM peut participer à la construction de différents objets textuels, et un même objet textuel peut être réalisé par différentes marques de MFM en fonction du texte. La reformulation discursive des objets textuels sous forme de métaphrases permet de s'affranchir de cette polysémie des marques de MFM, en explicitant les intentions architecturales de l'auteur sous-tendant leur utilisation. Les métaphrases représentent donc une forme explicite de la sémantique des architectures textuelles.

Du point de vue du lecteur, il est bon de noter que l'interprétation des marques de MFM n'est pas innée et nécessite un ensemble de connaissances et de compétences [Jamet et al. 2008], p.89 : « les signaux visuels des textes [*architectures textuelles*] ne sont pas interprétés de façon innée ; c'est bien l'usage de nombreux textes qui fait émerger chez le lecteur les comportements adéquats ». Ce constat rejoint le concept de *conventions* de MFM et le fait qu'il existe des conventions partagées qui sont déclinées, instanciées par les auteurs dans leurs textes à travers différentes architectures textuelles *conventionnées*, comme les titres et la hiérarchie définie par une titraille. L'interprétation des conventions de MFM par le lecteur ne peut se faire que si ce dernier possède les compétences nécessaires ; il paraît donc important pour la formalisation de la MFM (son interprétation « algorithmique ») que les conventions partagées aient été identifiées. Nous développerons ce point dans le Chapitre III.

L'interprétation de la MFM se base à la fois sur les marques typo-dispositionnelles et lexico-syntaxiques, qui sont le mode d'expression de la MFM, et à la fois sur le contenu primaire (les

« idées » de l'auteur exprimées linguistiquement, voir section 2.1). Ainsi, le lecteur peut interpréter la MFM du texte non seulement à l'aide des marques de MFM mais aussi à l'aide de la sémantique du contenu présentant des marques de MFM. Par exemple, si un objet textuel est ambigu (*est-ce un titre ?*), son contenu peut être utilisé pour désambigüiser ou simplement pour appuyer son interprétation (« *Introduction* », *c'est bien un titre*).

Le lien étroit entre MFM et contenu primaire est appuyé par [Payne and Reader 2006] qui ont montré que la compréhension d'un texte implique, en plus d'une représentation mentale de la sémantique des propositions du contenu et de leurs liens, une construction d'une représentation de sa structure physique en lien avec les propositions du contenu. On peut également citer les travaux portant sur le *codage spatial* du texte, qui ont montré que la localisation spatiale des mots et propositions du discours dans le texte était encodée en mémoire [Baccino and Pynte 1991; Rothkopf 1971].

La réalisation des objets textuels se faisant par contraste et identité entre les marques utilisées, une approche globale comparative semble nécessaire pour leur interprétation. L'idée est donc de comparer les différentes utilisations dans le texte des objets textuels présentant les mêmes marques, mais aussi les différentes utilisations d'objets textuels aux marques différentes. Il a d'ailleurs été montré que les lecteurs ne peuvent interpréter les marques de MFM que d'une manière globale [Lemarié 2006a]. L'interprétation des marques pour formaliser la MFM d'un texte doit donc suivre les principes de contraste et d'identité sur lesquels les objets textuels sont basés, et considérer le texte dans sa globalité.

## 2.4. Application du MAT : exemples de modélisations d'architectures textuelles

---

Cette section vise à donner quelques exemples d'architectures textuelles qu'il est possible de modéliser avec le cadre du MAT. Au-delà des plus répandues, il est possible de capturer des relations particulières et des types d'objets moins communs. Les architectures présentées dans cette section sont issues d'un corpus de manuels scolaires de mathématiques.

<b>&gt;&gt; Exercice :</b> Calculer $A = 2x^2 - 6x + 10$ pour $x = -5$ .	
<b>ÉTAPES</b>	<b>SOLUTION</b>
(1) J'attribue à $x$ la valeur donnée en rétablissant les signes « $\times$ » sous-entendus.	$A = 2 \times (-5)^2 - 6 \times (-5) + 10$
(2) J'applique les règles de priorité.	$A = 2 \times 25 - 6 \times (-5) + 10$
(3) J'achève le calcul.	$A = 50 + 30 + 10$
	$A = 90$

Figure 7 - Premier exemple d'architecture textuelle modélisable par le MAT

Cet exemple de double colonage est intéressant car chaque objet de la colonne « étapes » (discernable par les retours à la ligne et les numérotations) peut être mis en relation avec un ou plusieurs objets de la colonne « solution ». Au niveau de la modélisation, on peut en rendre compte à l'aide de métaphrases décrivant ces relations. Par exemple : (*l'auteur informe le lecteur que*) Le segment <étape\_1> <annote / illustre / correspond au> segment <solution1>. Le segment <étape\_2> <annote / illustre / correspond au> segments <solution2> et <solution3>.



Activités mentales  
Exercices rituels

Connaissance 1  
p. 14

- on détermine le signe du produit :
  - si on multiplie deux nombres de même signe,
  - si on multiplie deux nombres de signes différents.
- on ... leurs distances à zéro.

**b) Effectuer les calculs suivants.**  
 $A = (-5) \times (-2)$        $B = (+8) \times (-4)$   
 $C = (-7) \times (+3)$        $D = (+4) \times (+7)$

**c) Effectuer les calculs suivants.**  
 $E = -6 \times 5$                $F = 4 \times (-7)$   
 $G = -5 \times 8$                $H = 9 \times (-4)$

Figure 8 - Deuxième exemple d'architecture textuelle modélisable par le MAT

Ici, on s'intéressera aux renvois en marge : on repère visuellement que deux types de renvois différents sont présents. Le premier renvoi informe le lecteur que des activités sur format numériques sont disponibles pour l'ensemble de l'exercice correspondant (ce renvoi annote donc l'ensemble de l'exercice), tandis que le renvoi « Connaissance 1 » annote précisément la question « c » (si l'on s'en réfère à sa disposition spatiale et aux autres phénomènes similaires sur la page) en renvoyant à une connaissance permettant de résoudre cette question.

89 On remplit les cases en respectant la règle suivante :

Compléter les cases vides.

Figure 9 - Troisième exemple d'architecture textuelle modélisable par le MAT

Ce troisième phénomène est assez atypique, bien qu'assez proche d'une présentation tabulaire ; il est cependant possible de modéliser les relations entre les cases de la pyramide d'après la définition donnée au début de l'exercice, et pour chaque case faire correspondre plusieurs métaphrases renvoyant au statut de la case qui sera « <case\_0> <possède le statut de> <case\_A> relativement à la case X » et/ou « [...] <case\_B> relativement à la case Y » et /ou « [...] <case\_A\*B> relativement à la case Z ». La modélisation semble certes possible, mais pourrait ici entraver les objectifs pédagogiques de l'exercice qui comprennent, à priori, non seulement des calculs mais aussi la compréhension de la règle et son application à la configuration spatiale ; considérer les objectifs de la tâche est important quant à la restitution de la MFM modélisée comme nous le verrons plus en profondeur dans les chapitres suivants. Cet exemple a cela d'intéressant que les relations spatiales qu'il comprend sont assez semblables à celles qu'entretiennent les cases d'un tableau, auquel on pourrait donc également appliquer ce type de modélisation.

Pour conclure, même si le pouvoir descriptif du MAT permet de modéliser potentiellement toute architecture textuelle, il ne sera pas possible, ni même souhaitable, de tout retranscrire aux « lecteurs » déficients visuels à l'oral, à un niveau informatif comme fonctionnel. Toutefois, dans une optique d'accessibilité de l'information, pouvoir formaliser de la manière la plus exhaustive possible les architectures textuelles des documents, pour ensuite éventuellement n'en restituer qu'une partie

plus pertinente en fonction du contexte, reste une étape essentielle dans notre démarche. Nous verrons dans le Chapitre III comment mettre en place une telle formalisation.

### **3. Conclusion**

---

---

Nous avons vu que la mise en forme des textes, bien au-delà du simple fait de rendre le texte agréable à regarder et/ou à lire, avait une dimension informative et fonctionnelle dont l'impact sur le plan de la compréhension et sur la navigation visuelle dans les textes a déjà été démontré.

Le modèle d'architectures textuelles (MAT) offre un cadre théorique et analytique permettant de formaliser les phénomènes de mise en forme des textes (les objets textuels). Ce modèle, s'appuyant lui-même sur les travaux de [Harris 1971] qui considère que le langage naturel possède en son sein son propre métalangage, peut permettre de modéliser toute architecture textuelle, et ce dans un formalisme aisément implémentable avec des formats numériques répandus comme le RDF ou le XML [Sorin et al. 2013]. Sur la base du MAT, nous avons développé une méthodologie de formalisation des architectures textuelles, qui nous a ensuite permis de restituer ces phénomènes à l'oral, et d'évaluer l'impact de cette restitution sur la compréhension.

Au-delà de nos travaux, le MAT pourrait permettre une meilleure compréhension de ce qu'est la mise en forme des textes (ce qui a déjà été amorcé en psychologie cognitive [Lemarié et al. 2008]) et comment nous l'utilisons, et s'inscrit dans une perspective actuelle d'annotation (ajout de métadonnées) des données numériques [Berners-Lee et al. 2001] ; le texte étant le premier vecteur de sens des documents numériques, pouvoir capturer toutes ses dimensions semble une étape importante que ce soit pour l'accessibilité, la gestion de bases de données de documents ou encore la numérisation de documents papier et l'analyse automatique de leurs contenus.

# Chapitre II - Accessibilité des contenus textuels pour les déficients visuels : l'enjeu de la restitution de la mise en forme

Résumé : L'accessibilité des contenus numériques et des interfaces informatiques représente un enjeu majeur pour l'inclusion sociale et l'accès à la connaissance pour les personnes en situation de handicap. L'avènement des interfaces graphiques, et l'utilisation de structures visuelles riches dans les documents et interfaces, posent problème pour leur accessibilité notamment dans le cas des déficients visuels. Les caractéristiques inhérentes aux modes d'accès à l'information audio et tactile, profondément différentes de celles du visuel, représentent un premier facteur problématique pour l'accessibilité, que l'on présentera du point de vue de l'impact cognitif de l'accès à l'information dans des modes non-visuels. Nous verrons également dans ce chapitre comment les technologies d'assistance actuelles permettent l'accès aux contenus, en nous limitant aux contenus textuels, et dans quelle mesure la sémantique de la mise en forme est prise en compte dans ce cadre. En effet, en amont des considérations techniques sur les formats de fichiers ou le codage de l'information numérique, nous verrons pourquoi la prise en compte de la mise en forme des textes est nécessaire pour améliorer l'accessibilité non-visuelle des contenus.

## Sommaire du Chapitre II

<b>1. Introduction</b> .....	<b>28</b>
<b>2. Définir l'accessibilité pour mieux cerner ses enjeux</b> .....	<b>31</b>
2.1 Le parti pris de l'accessibilité numérique : une accessibilité de moyens universelle.....	31
2.2 De l'accessibilité à l'utilisabilité .....	32
<b>3. Comment les déficients visuels perçoivent et traitent l'information</b> .....	<b>34</b>
3.1 Conséquences cognitives de la déficience visuelle .....	35
3.2 Caractéristiques des modes audio et tactile comparativement au mode visuel : impact sur la cognition .....	36
3.3 Quelles solutions pour pallier aux contraintes des modes audio et tactile ?.....	40
3.4 Conclusion et implications pour l'accessibilité des contenus numériques .....	41
<b>4. Approches actuelles pour l'accessibilité des contenus numériques</b> .....	<b>44</b>
4.1 Approches générales pour la création de contenus accessibles .....	44
4.2 Panorama des formats numériques pour documents textuels.....	45
4.3 Les normes pour l'accessibilité des contenus numériques .....	48
4.4 Conclusion .....	53
<b>5. Les technologies d'assistance pour l'accès non-visuel aux contenus numériques : approches et limites</b>	<b>54</b>
5.1 Périphériques et modalités d'entrée et de sortie : les bases de l'accès (non-visuel) au numérique .....	54
5.2 Les lecteurs d'écran.....	61
5.3 Blocs-notes braille .....	69
5.4 Format et lecteurs DAISY.....	70
5.5 Conclusion .....	71

<b>6. Vers la restitution des architectures textuelles pour l'accessibilité des contenus numériques .....</b>	<b>73</b>
6.1 Prise en compte de la MFM par les lecteurs d'écran .....	73
6.2 Exemples empiriques d'apports de la MFM à l'accessibilité des contenus.....	75
6.3 De l'intérêt de la restitution de la sémantique de la MFM : exemples d'architectures textuelles non-restituées.....	76
<b>7. Étude en situation : accès et accessibilité de l'information pédagogique pour les élèves déficients visuels .....</b>	<b>78</b>
7.1 Contexte général : institutions, acteurs et moyens.....	78
7.2 Méthode et données d'investigation .....	79
7.3 Résultats : accès et accessibilité à l'information pédagogique .....	81
7.4 Récapitulatif des problèmes de disponibilité et d'accessibilité de l'information.....	86
7.5 Discussion .....	87
<b>8. Conclusion et problématique générale.....</b>	<b>89</b>

## 1. Introduction

---

---

L'accessibilité est un concept plus que jamais d'actualité, repris fréquemment dans les médias, en ce qui concerne l'accès aux infrastructures publiques pour les personnes en situation de handicap, avec notamment l'ordonnance n° 2014-1090 du 26 septembre 2014 repoussant les délais fixés par la loi du 11 Février 2005 sur *l'égalité des droits et des chances*. Le terme *accessibilité* renvoie dans le langage courant à la possibilité pour des personnes déficientes de pouvoir utiliser les mêmes services, outils et infrastructures que les personnes non déficientes, auquel cas on pourra qualifier ces services, outils ou infrastructures « *d'accessibles* ». La notion évoque donc la « possibilité d'accéder à ». Telle quelle, cette notion est extrêmement vaste, puisqu'elle englobe à la fois toutes les déficiences possibles, mais aussi des *situations* que peuvent rencontrer des personnes valides au-delà du monde du handicap (ex. « cet endroit est-il accessible en voiture ? »). Dans son acception courante, l'accessibilité ne présage en rien de la facilité avec laquelle l'accès est réalisé, des contraintes rencontrées par les utilisateurs, ou encore des frustrations ou sentiments d'exclusion potentiellement engendrées. Ces différentes notions et leurs liens avec l'accessibilité seront donc clarifiés dans la première partie de ce chapitre, afin de mieux asseoir les enjeux et délimiter le périmètre de nos travaux.

L'accessibilité touche donc en premier lieu les personnes *déficientes*, c'est-à-dire présentant une « lésion ou altération d'une structure ou fonction psychologique, physiologique ou anatomique, temporaire ou permanente » [Uzan 2005]. Cependant les aménagements ou développements réalisés pour ces populations visent à compenser les *incapacités fonctionnelles* qui sont les conséquences de la déficience ; or les mêmes incapacités peuvent survenir chez des personnes non déficientes dans certains contextes ou environnements. Par exemple, les bateaux aménagés sur les trottoirs bénéficient non seulement aux déficients moteurs, mais également aux poussettes ou aux livraisons. De la même façon, [Sperandio 2007] montre à travers de nombreux exemples que les difficultés rencontrées par des personnes déficientes sont souvent révélatrices de celles rencontrées dans certaines circonstances par des personnes valides. Nos travaux sur l'accessibilité des contenus numériques pour les déficients visuels ont ainsi une portée plus large, qui s'étend aux utilisateurs dans l'incapacité d'utiliser leur vue pour accéder à ces contenus.

L'objectif principal de ces travaux est d'améliorer l'accessibilité des documents numériques à travers la restitution de la sémantique de leur mise en forme, qui véhicule des architectures textuelles faisant partie intégrante de la sémantique du document. Afin de concevoir une restitution la plus adaptée possible aux utilisateurs déficients visuels, on commencera dans ce Chapitre par présenter les conséquences cognitives de la déficience visuelle mais aussi les caractéristiques des modes d'accès non-visuels à l'information (partie 3) ; il s'agit de mieux connaître les utilisateurs, mais aussi de comprendre comment une perception visuelle de l'information peut différer d'une perception des mêmes informations par d'autres sens que la vue. L'ensemble de ces réflexions nous permettront de guider la conception de la restitution des architectures textuelles, et soulèvent plus généralement des points importants impactant la conception d'interactions non-visuelles.

Actuellement, l'accessibilité de l'information numérique pour les déficients visuels repose sur deux axes. Le premier est l'accessibilité du contenu en lui-même, c'est-à-dire sa conformité relativement aux normes en vigueur, qui permet d'assurer une accessibilité de *moyens* c'est-à-dire la disponibilité

d'un maximum d'informations du contenu d'origine (ex. rendre accessible les images à l'aide d'alternatives textuelles, cf. partie 4). Le second est celui des technologies d'assistance logicielles, principalement les lecteurs d'écran (partie 5), qui jouent le rôle d'intermédiaires entre les contenus numériques et les synthèses vocales et plages braille qui permettent un accès non-visuel à l'information. Ces technologies sont nécessaires, étant donné que les systèmes d'exploitation et applications logicielles grand public sont conçus pour être utilisés « graphiquement ». Ainsi nous présenterons ci-après l'état de l'art des approches pour la disponibilité des contenus numériques, mais aussi celui des technologies d'assistance pour l'accès à ces contenus, afin de pouvoir positionner nos travaux mais aussi d'évaluer dans quelle mesure l'existant dans le domaine de l'accessibilité numérique prend en compte la mise en forme des textes.

Dans le premier chapitre de ce document, nous avons pu constater que la mise en forme, loin d'être un simple ornement du texte, est porteuse de sens, permet de structurer les textes, et permet au lecteur de mettre en place des stratégies de navigation visuelle dans le document ainsi que certaines optimisations cognitives pour la lecture. Toutes ces propriétés ont des conséquences majeures pour les activités mentales déployées par les lecteurs à partir de documents, notamment pour la compréhension, et justifient l'intérêt de restituer la MFM (la mise en forme signifiante des textes) aux déficients visuels afin de rendre les contenus et interactions plus accessibles. On parle bien de *restitution* de la MFM, des architectures textuelles qu'elle véhicule, et des fonctions de lecture qu'elle fait émerger, avec l'idée sous-jacente de proposer des équivalents non-visuels à ces informations et fonctions, mais des équivalents qui ne soient pas une simple retranscription (ou retransmission) de l'information dans des modalités non visuelles. Nous verrons que l'approche actuelle de l'accessibilité des contenus numériques, que l'on qualifiera d'*exhaustive*, consiste justement à retransmettre (rendre disponibles) un maximum d'informations pour que les utilisateurs déficients aient accès aux mêmes informations que les non déficients. Nous défendrons dans nos travaux une approche *holistique* (globale) de la conception d'interactions non-visuelles [Phipps and Kelly 2006; Giraud 2014], et qui s'oppose dans les faits à l'approche exhaustive, en cherchant à fournir une restitution *fonctionnelle* centrée sur la tâche et le contexte d'utilisation. Le but n'est alors plus de restituer un maximum d'informations mais de restituer ce qui est pertinent pour l'atteinte des objectifs, et ce de la manière la plus optimisée possible en termes cognitifs, et en termes de performance.

Nos travaux sont centrés autour de la restitution des architectures textuelles véhiculées par la MFM, dans le cas des documents *numériques*, et ce pour les possibilités offertes par ce contexte en termes de modalités et transformations de l'information, en entrée et sortie, mais aussi pour les nombreuses implications et impacts de l'accessibilité numérique mentionnés ci-avant. Plus précisément, nous nous intéressons au cas de l'accès au contenu textuel des documents numériques, ces contenus faisant l'objet de mises en forme de plus en plus riches depuis l'avènement des interfaces graphiques au milieu des années 90. Ce niveau d'étude peut sembler trivial à l'heure où les recherches en accessibilité portent sur les contenus complexes comme les mathématiques ou les jeux vidéo [Archambault 2010], ou encore les contenus dynamiques sur le web [Thiessen 2011]. Nous montrerons dans ce chapitre comment, au contraire, de nombreuses informations de mise en forme sont actuellement inaccessibles via les technologies d'assistance, et qu'ainsi une partie importante de la sémantique des documents est ignorée. L'accessibilité des contenus textuels n'est donc pas quelque chose d'acquis.

Nous présenterons tout d'abord dans ce chapitre les différentes définitions des notions touchant à l'accessibilité des contenus numériques et l'utilisabilité des technologies d'assistance, utilisées pour accéder à ces contenus, afin de mieux comprendre les enjeux de l'accessibilité associés aux différents termes et définitions existantes. La partie 3 étudie les caractéristiques de la déficience visuelle et celles de la perception audio et tactile d'un point de vue cognitif, posant ainsi les bases pour la conception d'interactions non-visuelles. Les deux parties suivantes dresseront l'état de l'art de l'accessibilité des documents numériques et des technologies d'assistance utilisées pour accéder à ces contenus, afin de remettre nos travaux en contexte. Enfin, nous présenterons comment et dans quelle mesure la mise en forme est actuellement prise en compte par les technologies, et comment une grande partie de sa sémantique reste ignorée à l'heure actuelle.

## 2. Définir l'accessibilité pour mieux cerner ses enjeux

---

---

Nous nous attacherons à définir et préciser plusieurs notions gravitant autour du concept d'accessibilité évoqué en introduction, en essayant de nous limiter au cadre des contenus numériques et interfaces informatiques. Pour rappel, l'accessibilité telle que nous l'appréhendons dans le langage courant renvoie au caractère accessible d'un outil, service ou contenu numérique, et donc à sa propension à pouvoir être utilisé par une population présentant ou non des déficiences. Cette définition n'est cependant pas consensuelle ; on ne retrouve d'ailleurs pas de définition de l'accessibilité dans les standards ISO<sup>12</sup>. Certains auteurs suggèrent que l'absence de définition consensuelle aurait des effets délétères [Persson et al. 2014], car l'utilisation trop permissive du terme limiterait la marge de progression qu'il pourrait y avoir dans ce domaine, en laissant penser qu'on ne peut aller « au-delà de ce qui est considéré comme accessible ». Les mêmes auteurs pointent cependant du doigt qu'une définition unique de l'accessibilité serait trop large pour être utilisée efficacement. Lorsque l'on parle d'accessibilité, il semble donc important de préciser quelles populations sont concernées et dans quel contexte.

Malgré les difficultés de mise en œuvre que l'on évoquera, les enjeux de l'accessibilité ont ouvert un vaste champ de recherche, où l'informatique semble un des domaines les plus propices pour voir émerger une accessibilité universelle. En effet, un des paradigmes fondamentaux en informatique est la séparation entre les données, le stockage de l'information, et la manière dont cette dernière est présentée (restituée). Ainsi, la restitution de l'information et sa manipulation (l'interaction entre l'homme et la machine) est adaptable, modulable, d'autant que les périphériques informatiques offrent un large panel de modalités d'interaction utilisables.

### 2.1. Le parti pris de l'accessibilité numérique : une accessibilité de moyens universelle

---

Le sens premier de l'accessibilité est celui d'une *accessibilité de moyens*, c'est-à-dire proposant à toute personne quelle que soit son handicap ou ses capacités, les mêmes moyens, les mêmes possibilités que les autres dans l'utilisation d'un produit, l'accès à un service ou à une infrastructure. Ainsi, quelque chose d'accessible dans l'absolu devrait l'être pour tout type de population. Selon cette définition il est très délicat d'évaluer l'accessibilité car cela impliquerait d'effectuer des tests auprès de tous types de déficients, en prenant en compte les polyhandicaps, les différentes tranches d'âge, etc., ce qui représente une tâche dantesque. Dans la littérature, le « problème » est déplacé sur la conception, et on parle alors de : *design for all*, *universal design*, *inclusive design*, ou encore *d'accessible design*. Ces concepts s'opposent à ce que l'on pourrait appeler le *specific design*, qui serait la conception pour un type de population cible en particulier. Les définitions de ces concepts, bien que légèrement différentes car produites par des organismes ou auteurs différents, renvoient toutes au caractère inclusif de la conception [Persson et al. 2014].

Malgré l'intérêt de l'étude du paradigme d'accessibilité universelle, la complexité de sa mise en pratique soulève quelque scepticisme quant à sa faisabilité. Les raisons sont multiples : difficulté de ciblage et recrutement des populations ciblées, possibles conflits entre l'utilité d'une conception

---

<sup>12</sup> International Organization for Standardization



pour un type de population qui pénaliserait une autre population, multiplicité des technologies d'assistance utilisées par les populations présentant une déficience, etc. [Newell and Gregor 2000] proposent d'ailleurs de passer d'un paradigme de *universal design* à un autre plus réaliste labellisé *User Sensitive Inclusive Design*, insistant sur le caractère inclusif de la conception qui dans certains cas prend en compte certaines spécificités de populations particulières.

Malgré les difficultés évoquées, c'est cette approche d'accessibilité (de moyens) *universelle* qui est actuellement adoptée pour assurer l'accessibilité des documents numériques : chaque document devrait ainsi respecter certaines normes de conception (présentées en 4.3) pour être accessible *au plus grand nombre*, sachant que l'accessibilité effective est ensuite laissée à la charge des technologies d'assistance gérant l'accès aux contenus. Le W3C (*World Wide Web Consortium*) définit ainsi un certain nombre de critères d'accessibilité sous le sigle WCAG (*Web Content Accessibility Guidelines*, voir 4.3 pour plus de détails), que les auteurs de contenus doivent respecter afin que les contenus produits soient les « plus accessibles pour les personnes déficientes »<sup>13</sup>. Le W3C donne d'ailleurs pour définition de l'accessibilité « la capacité pour les personnes déficientes d'utiliser le web [...] et plus précisément leur capacité à percevoir, comprendre, naviguer et interagir avec le Web » [Henry 2005]. L'accessibilité *numérique* semble donc pour le moment restreinte à une accessibilité de moyens qui se veut universelle, en passant par un ensemble de normes. Cependant, le fait qu'une page web respecte les WCAG ne présage en rien de l'utilisabilité du contenu et des interactions proposées par la page, qui dépend principalement de la technologie d'assistance utilisée.

## 2.2. De l'accessibilité à l'utilisabilité

---

L'accessibilité fait référence à la capacité d'accéder à quelque chose, mais ne laisse rien présager quant à l'efficacité de l'accès, de l'utilisation. Par exemple, on peut dire qu'un site web est accessible aux utilisateurs déficients visuels s'ils peuvent y accéder, quand bien même ce serait au prix de diverses frustrations et erreurs qui sont d'ailleurs courantes dans ce cadre comme souligné par [Lazar et al. 2007]. Dans le champ de l'éducation et des apprentissages, l'accessibilité est définie comme « la possibilité pour un élève d'accéder, de comprendre et de répondre à un contenu pédagogique » [Ketterlin-Geller and Tindal 2008]. Dans les deux cas la définition de l'accessibilité fait référence à une accessibilité de *moyens* sans aborder les questions des performances et des frustrations des utilisateurs (ou apprenants).

La notion d'utilisabilité vient ici compléter ce manque, car elle englobe l'efficacité, l'efficacité et la satisfaction avec laquelle un objet ou système est utilisé pour atteindre un objectif donné. Apparue au début des années 90 [Nielsen 1993], l'utilisabilité fait ainsi référence à la facilité, l'aisance avec laquelle un système (au sens large) peut permettre à son utilisateur d'atteindre un certain but dans un contexte donné. Certains auteurs montrent d'ailleurs comment les méthodes et techniques utilisées pour évaluer l'utilisabilité peuvent être réutilisées dans le cadre de l'accessibilité [Theofanos and Redish 2003; Leporini and Paternò 2008]. Le concept d'utilisabilité universelle est d'ailleurs apparu [Shneiderman 2000] comme une extension du caractère inclusif de l'accessibilité au concept d'utilisabilité : il s'agit alors de concevoir des outils non seulement accessibles au plus grand nombre mais également utilisables.

---

<sup>13</sup> Source : <http://www.w3.org/WAI/intro/wcag>

Il nous paraît important conceptuellement de distinguer l'accessibilité de l'utilisabilité, le second représentant une extension du premier. En effet le concept d'accessibilité est actuellement employé pour désigner une accessibilité de moyens c'est-à-dire une *disponibilité* d'informations, de systèmes ou de services pour une population donnée. Au-delà de cette disponibilité il semble essentiel de considérer l'efficacité, l'efficience et la satisfaction d'utilisation ou d'accès à ces informations ou services (l'utilisabilité au sens large). L'utilisabilité commence à rentrer en ligne de compte dans les travaux de recherche sur l'accessibilité, auquel cas on parlera alors d'accessibilité effective [Giraud 2014], d'accessibilité utilisable [Rohani Ghahari et al. 2012] ou bien d'utilisabilité universelle quand elle est étendue à virtuellement tout utilisateur, quelle que soit ses capacités physiques et cognitives. Au-delà de considérations terminologiques, cela nous permet de distinguer deux aspects distincts de nos travaux : le premier est de rendre *disponible* dans les documents numériques les architectures textuelles véhiculées par la mise en forme (voir Chapitre I) et ainsi contribuer dans une certaine mesure à l'accessibilité (de moyens) universelle ; le second est de rendre l'accès à cette sémantique la plus utilisable possible pour les déficients visuels (voir Chapitre IV traitant de la restitution de la mise en forme). Ainsi, le titre de ce manuscrit « contributions des architectures textuelles à l'accessibilité non-visuelle des documents numériques » fait référence aux potentiels bénéfiques retirés par les déficients visuels à travers ces deux niveaux d'accessibilité à savoir la disponibilité des architectures textuelles, et leur restitution pertinente (utilisable).

### 3. Comment les déficients visuels perçoivent et traitent l'information

---

---

Dans la population Française il y aurait 1 aveugle pour mille et 1.5% de malvoyants [Gendron 2010] ; ces chiffres seraient d'ailleurs en augmentation, malgré les progrès thérapeutiques, notamment à cause de l'allongement de la durée de la vie [Sander et al. 2005]. Les chiffres mondiaux sur le handicap quel qu'il soit sont d'ailleurs en augmentation [WHO - World Health Organization 2012], montrant la nécessité de développer un effort de recherche autour de l'accessibilité, d'autant que les polyhandicaps sont nombreux ; par exemple [Sander et al. 2005] rapportent qu'en France 4 déficients visuels sur 5 déclarent une ou plusieurs autres déficiences. Concernant les chiffres de la déficience visuelle, il existe un spectre étendu de pathologies visuelles différentes qui sont regroupées sous les termes « cécité » et « malvoyance », et l'on distingue parfois même la « malvoyance profonde » de la « malvoyance légère ». Dans la dernière étude citée, cette classification se fait sur des critères fonctionnels (ex. « pouvoir lire de près le journal »), tandis que la classification internationale des maladies utilisée par l'Organisation Mondiale de la Santé fixe certains seuils d'acuité visuelle<sup>14</sup> pour catégoriser les différents types de déficience visuelle. Dans notre cas nous utiliserons le terme déficients visuels pour désigner l'ensemble des personnes atteintes d'une pathologie du système visuel, le terme non-voyants pour désigner les personnes incapables d'accéder visuellement à du texte écrit et le terme malvoyants pour désigner les déficients visuels dont la vision permet tout de même un certain accès à l'information écrite. Nous utiliserons le terme « déficient visuel » le plus souvent dans ce manuscrit, car ces travaux sur l'accessibilité non visuelle ont des implications aussi bien pour les personnes atteintes de cécité que de malvoyance.

Dans cette partie, on fait référence à *l'information* en tant que donnée perceptive émise par l'environnement (au sens large) et captée par les déficients visuels. Cette transmission peut se faire selon plusieurs *modes*<sup>15</sup> [Awdé 2009] qui correspondent aux sens perceptifs humains ; on s'intéressera plus particulièrement aux modes audio et tactile<sup>16</sup> qui sont les deux vecteurs principaux d'information utilisables par cette population, en mettant à part les autres modes de transmission (olfactif, gustatif) qui sont utilisées de manière très marginale dans les interfaces informatiques.

Dans le cadre de nos travaux, étudions les possibilités de restitution de la MFM aux déficients visuels lors de l'accès aux documents textuels. Afin que cette restitution soit utilisable et plus généralement optimale, il nous semblait important de présenter tout d'abord les caractéristiques cognitives de la déficience visuelle, puis les spécificités des modes auditif et tactile ainsi que leurs conséquences sur la cognition. Nos travaux étant centrés sur la restitution de la mise en forme on ne s'intéressera ici

---

<sup>14</sup> Source : <http://apps.who.int/classifications/icd10/browse/2010/en#/H53-H54>

<sup>15</sup> Ne pas confondre avec les *modalités* de transmission qui sont les structures spécifiques de l'information pour un mode donné (ex. pour le mode audio on peut citer la parole et la musique), et le *média* qui constitue le support physique ou logique permettant la transmission de l'information (ex. l'écran pour les modalités visuelles)

<sup>16</sup> Le sens du toucher est souvent divisé en deux catégories [Pietrzak 2008](p.13) : « *ce que l'on ressent par les mécanorécepteurs de la peau (le tactile ou cutané) et ce que l'on ressent par les muscles, les tendons et les articulations (le sens kinesthésique)* ». Dans ce Chapitre on utilisera le terme *tactile* pour désigner le sens *cutané* ; certains dispositifs utilisant le sens *kinesthésique* seront toutefois présentés à titre d'exemples.

qu'aux modes permettant aux déficients visuels d'acquérir de l'information (de l'environnement ou la machine vers l'utilisateur), laissant de côté les modes et modalités d'entrée d'information par les personnes déficientes (de l'utilisateur vers la machine), qu'il serait toutefois intéressant d'étudier dans le cadre de la conception d'interactions multimodales.

#### **3.1. Conséquences cognitives de la déficience visuelle**

---

Sans la vision, il est plus difficile de faire l'expérience directe (sensorielle) des objets du monde, principalement pour des raisons pratiques : il est plus simple de présenter à un enfant l'image d'un cheval par exemple plutôt que de l'emmener dans un centre équestre pour qu'il fasse en fait l'expérience tactile, olfactive et auditive. Ainsi certaines représentations des déficients visuels peuvent s'avérer être uniquement verbales, car dépourvues de stimuli sensoriels associés [Chauvey et al. 2012] là où les voyants peuvent y associer des images du monde, plus facilement accessibles que des stimuli tactiles ou auditifs. Cependant, [Galiano and Baltenneck 2007] expliquent que l'acquisition de connaissances conceptuelles peut se faire sans problème à partir d'informations verbales, même sans expérience sensorielle directe associée à ces concepts, tout comme c'est le cas pour l'acquisition de concepts abstraits chez les voyants.

Sans la vision, certains mécanismes de compensation se mettent apparemment en place au niveau cortical pour traiter l'information orale et tactile [Sadato et al. 1996; Burton 2003] ; les aires visuelles du cortex occipital sont en effet réallouées au traitement des informations de ces canaux perceptifs. Ces mécanismes sont à l'origine de différences en termes de performances entre voyants et déficients visuels dans le traitement de l'information orale [Wan et al. 2010] ainsi que dans sa perception [Hugdahl et al. 2004]. La mémoire à court terme montre également des performances significativement plus élevées [Pring 2008] ; ce dernier auteur souligne cependant qu'il n'est pas possible de prédire les performances (ici chez les enfants) déficients visuels, et qu'en marge des performances observées, d'autres aspects du développement chez les enfants déficients visuels subissent souvent des retards. Enfin, les travaux cités montrent que les différences observées de ce type de performances entre voyants et déficients visuels ne se vérifient que dans le cas des non-voyants congénitaux où la déficience est apparue très tôt, voire dès la naissance. Les déficients visuels tardifs semblent eux avoir des capacités d'imagerie mentale<sup>17</sup> équivalentes aux voyants grâce à la présence en mémoire d'informations visuelles antérieures à leur déficience [Vanlierde and Wanet-Defalque 2005].

Même si le traitement de l'information audio-tactile semble optimisé chez les déficients visuels, ces performances n'ont été observées que dans le cas de tâches faisant appel uniquement à ces canaux perceptifs. La comparaison entre le mode visuel et les modes audio et tactile montrent que ces derniers présentent un certain nombre de limitations intrinsèques qui ont des conséquences sur le plan cognitif.

---

<sup>17</sup> L'imagerie mentale fait référence à la représentation et la manipulation mentales d'images

### 3.2. Caractéristiques des modes audio et tactile comparativement au mode visuel : impact sur la cognition

Afin d'illustrer les caractéristiques de ces modes, nous commencerons par présenter différentes modalités et dispositifs utilisés par les déficients visuels pour chacun de ces modes. Plus de détails sur leurs utilisations et caractéristiques seront données dans la section 5.1.

Le mode audio utilise principalement le discours oral, la musique et les sons non-verbaux pour transmettre de l'information ; ces modalités sont facilement réutilisables pour des interfaces informatiques avec notamment des voix de synthèse. Concernant le mode tactile, le braille est ici la principale modalité utilisée ; c'est un codage des lettres, chiffres et signes du langage écrit à l'aide d'une matrice de 6 picots en relief qui permettent ainsi de coder 64 caractères ( $2^6$ ). L'introduction de 2 picots supplémentaires en informatique permet de coder 256 caractères ( $2^8$ ). Le braille était initialement embossé sur papier mais l'avènement des afficheurs braille dynamiques (voir 5.1.b) a permis son utilisation pour l'accès aux contenus numériques. L'exploration de figures 2D en relief est également une modalité couramment utilisée par les déficients visuels notamment pour l'accès aux cartes tactiles ou aux formes géométriques, qui mérite d'être citée pour illustration. Enfin, il est à noter qu'on ne traitera pas dans ces travaux du mode de transmission kinesthésique, désignant la perception du corps dans l'espace et de ses mouvements, utilisé notamment avec des dispositifs à retour d'effort [Sullivan et al. 2014]. On se limitera donc au mode audio et au mode tactile limité au « cutané ». D'autres dispositifs utilisant les modes audio et tactile seront présentés en section 5.1.



Figure 10 – Alphabet braille (gauche) et figure en relief sur papier thermogonflable<sup>18</sup> (droite)

La première caractéristique commune aux modes audio et tactile est l'acquisition **séquentielle** de l'information. À l'oral, cette séquentialité n'offre aucun choix puisqu'elle dépend d'une « nécessité biomécanique » [Harris 1993]. La lecture braille est également séquentielle même si les techniques de lecture peuvent impliquer les index des deux mains, ces cas de figure impliquent uniquement une utilisation alternative des deux mains pour parcourir la ligne de caractères [Millar 2003] afin de limiter les mouvements des avant-bras. Cela est cependant à nuancer si l'on considère non pas un terminal braille informatique limité à une seule ligne, mais une page braille en relief où les deux mains peuvent être utilisées à des fins exploratoires ; [Ammar Ali 2006] met d'ailleurs en évidence dans sa thèse un niveau de lecture « globale » des documents en reliefs où les deux mains peuvent être espacées et utilisées conjointement pour « appréhender les différents blocs d'informations ». Cependant, même lors de ces phases de lecture ponctuelles où les deux mains sont utilisées l'accès reste séquentiel même si la rapidité et la résolution augmentent.

<sup>18</sup> Le thermogonflage est une technique utilisant un papier spécial contenant des capsules qui lorsqu'elles sont imprégnées d'encre gonflent avec la chaleur.

La séquentialité inhérente aux modes audio et tactile, opposée au caractère global/unifié <sup>19</sup> des stimuli visuels, implique tout d'abord un effort mnésique pour pouvoir retenir et organiser les séquences d'information (espacées dans le temps) dans un tout cohérent, ce qui impacte notamment la compréhension du discours dont on sait qu'elle implique la construction d'une représentation globale cohérente [Van Dijk and Kintsch 1983]. De plus, cette séquentialité ne permet pas de donner un contexte à l'information rentrante, comme c'est le cas avec la vision périphérique, ce qui rend la représentation globale plus difficile en l'absence notamment d'anticipation des changements. En lecture visuelle, il a été montré par exemple que les changements de paragraphe sont anticipés en vision périphérique et traités comme des marqueurs de ruptures thématiques et ce repérage anticipé provoque alors un ralentissement de la lecture en fin de paragraphe utile pour mettre en place des processus intégratifs [Hyönä 1995]. Même si la lecture visuelle d'un texte écrit peut elle aussi être qualifiée de séquentielle car répondant à l'ordre de présentation des mots, elle se différencie de la lecture braille et de l'oralisation du texte par l'inscription de l'information dans un contexte (un tout) visuel. Il est d'ailleurs intéressant de noter que lors de la lecture visuelle d'un long document sur ordinateur, contextualiser l'information en donnant une vue globale du document permet une plus grande efficacité de compréhension et satisfaction des utilisateurs [Hornbæk and Frøkjær 2001]. On touche ici du doigt un autre facteur permettant de différencier différents modes et modalités d'accès à l'information qui est la présence ou l'absence d'un *référentiel global* contenant les informations transmises. En lecture visuelle la page joue ce rôle et de la même façon en lecture braille sur papier ; cependant lors d'un accès à un texte par synthèse vocale ou à l'aide d'une plage braille ne permettant un accès que ligne à ligne, le référentiel global disparaît. Il serait d'ailleurs intéressant de comparer les conséquences cognitives de la perte de référentiel global en comparant une lecture braille papier à l'accès au même contenu sur une plage braille.

En restant sur l'exemple de l'accès à l'information écrite, on remarque également une différence de rapidité d'accès à l'information entre ces différents modes. On estime que la lecture braille oscille autour de 100/150 mots par minute [Legge et al. 1999] contre 250 à 300 pour la lecture visuelle [Maumet 2007; Millar 2003]. Ces chiffres ne représentent cependant qu'une estimation et varient considérablement en fonction de la population étudiée. L'écoute d'un texte oralisé par synthèse vocale se fait par défaut à environ 150 mots par minute, mais certains utilisateurs déficients visuels dépassent très largement les 300 mots par minute de la lecture silencieuse [Asakawa and Takagi 2003]. Ces chiffres sont à mettre en perspective avec la quantité d'information verbale nécessaire pour transposer l'information d'un mode à l'autre ; par exemple pour décrire un texte mis en forme visuellement, plusieurs autres termes et développements discursifs peuvent être nécessaires à l'oral ou en braille. De la même façon, une image ou une figure géométrique nécessiteront des descriptions verbales extensives ; dans le cas de la transcription sur du papier en relief, la densité maximale des traits étant faible à cause de la résolution de la perception tactile [Bou-Issa 2010], si l'objet s'avère complexe il nécessitera peut être plusieurs documents en relief différents pour être décrit. Ainsi non seulement les vitesses d'accès « brutes » sont différentes, mais ces différences sont accentuées par les contraintes de reformulation du visuel vers le verbal, rendant le temps total de l'accès à une même information bien plus important dans des modes non-visuels [Uzan 2005]. Par exemple, en navigation web, il est estimé qu'un non-voyant met au minimum 5 fois plus de temps

---

<sup>19</sup> On peut ici renvoyer au Gestaltisme ayant notamment étudié la perception visuelle, qui renvoie au fait qu'un stimulus visuel soit en lui-même un *gestalt*, un « percept global différent de la somme de ses parties » (pour paraphraser [Wertheimer and Riezler 1944])

que des personnes voyantes [Sperandio et al. 2004] pour réaliser la même tâche, allant jusqu'à un facteur 12 pour certains sites de commerce [Giraud 2014](p. 101).

La particularité de l'écrit est de pouvoir véhiculer graphiquement à la fois de l'information visuelle (la mise en forme et l'inscription spatiale) et de l'information verbale (le texte). D'après le Modèle d'Architecture Textuelle (MAT) la mise en forme visuelle des textes est un vecteur d'informations métatextuelles, qui permet d'alléger le contenu verbal en présentant des informations sur le texte lui-même non pas verbalement mais visuellement. Non seulement ce procédé diminue le volume de texte mais cela est particulièrement optimal en termes cognitifs, étant donné que la mémoire de travail est composée de plusieurs sous-systèmes esclaves traitant séparément l'information visuelle de l'information verbale [Baddeley and Hitch 1974].

En effet, d'après le modèle de Baddeley [Baddeley and Hitch 1974; Baddeley 2003], tout traitement cognitif de l'information utilise la Mémoire De Travail (MDT), historiquement appelée *mémoire à court terme*. La MDT est un système de mémoire transitoire permettant de traiter de petites quantités d'information en lien avec la mémoire à long terme (MLT) qui, elle, représente une zone de stockage permanente de l'information. La MDT est composée d'un système central qui contrôle et coordonne des sous-systèmes esclaves. Le modèle comprend dans sa version initiale 3 sous-systèmes esclaves [Baddeley 2003] dont le calepin visuo-spatiale, chargé de manipuler les informations visuelles et spatiales, et la boucle phonologique faisant de même avec les informations verbales. Les capacités de rétention de ces systèmes sont cependant limitées ; par exemple on mesure classiquement l'empan mnésique de la boucle phonologique avec des tâches de rétention de séries de nombres ou de mots (cet empan serait d'ailleurs de  $7 \pm 2$  items [Miller 1994]). Enfin, une dernière entité (le buffer épisodique) est chargée de la communication entre la MDT et la mémoire à long terme.

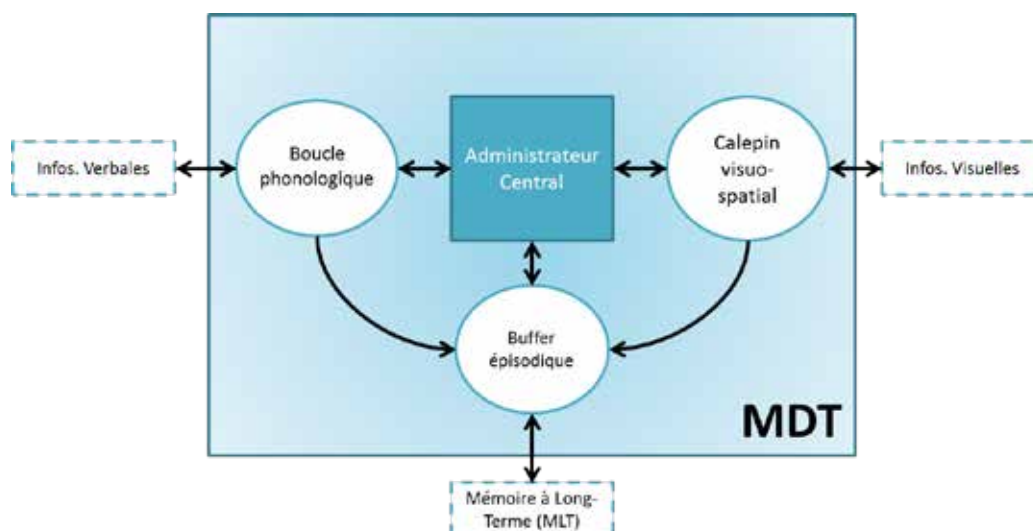


Figure 11 - Schéma de la mémoire de travail adapté de [Baddeley 2003]

D'après ce modèle de la mémoire de travail (MDT), actuellement en vigueur, lorsque je lis un titre mis en forme une entité dédiée se charge du traitement de l'information verbale (la boucle phonologique) pour interpréter le signifié du texte tandis qu'une autre entité (le calepin visuo-spatial) me permet d'inférer que ces propriétés visuelles sont celles d'un titre et donc que ce segment textuel a ce statut. Ainsi en l'absence de l'intervention du calepin visuo-spatial, si

l'ensemble de ces informations sont transmises verbalement (ex. par synthèse vocale) elles seront toutes traitées par une seule unité de traitement, la boucle phonologique, et demanderont un effort cognitif plus important. À noter que [Berz 1995] postule l'existence d'autres modules de traitement dédiés à la musique aux sons non-verbaux dans la mémoire de travail, ce qui aurait notamment pour conséquence l'optimisation du traitement du discours oral et de la prosodie associée.

De manière similaire à la complémentarité du visuel et du verbal dans le langage écrit, [Quek and Oliveira 2013] pointent la complémentarité du discours oral et de la gestuelle dans le dialogue. Ils mettent en avant le fait qu'en l'absence d'accès à l'expression corporelle du locuteur qui dans ce cas-ci est un professeur, l'apprentissage en classe est rendu particulièrement difficile pour les déficients visuels. Dans le cadre d'une lecture de texte par synthèse vocale, même si le discours est plus formalisé et limité au cadre du contenu on peut supposer que l'absence du visuel avec l'expression corporelle du locuteur puisse être pénalisante.

La fugacité du signifiant auditif s'oppose à la persistance, au caractère tangible du tactile et du visuel. Là où le visuel et le braille permettent une réinspection sélective (en lecture visuelle on sait notamment qu'un codage spatial de l'information permet la réinspection sélective de texte par des saccades oculaires régressives [Inhoff and Weger 2005]), l'oral va nécessiter un dialogue entre interlocuteurs afin d'assurer la transmission du message. Ainsi lorsque le dialogue oral n'est pas possible, par exemple lors de l'accès à un texte par synthèse vocale, l'utilisateur doit faire un effort attentionnel et mnésique supplémentaire pour éviter de perdre de l'information, où alors faire en sorte de pouvoir revenir en arrière (on peut illustrer cette remarque en citant les difficultés d'utilisation des serveurs vocaux). Seulement dans le cas des synthèses vocales, ces retours en arrière nécessitent des phases de navigation « exploratoires » pour atteindre l'emplacement désiré ; le terme *exploratoire* insiste sur le fait que l'utilisateur n'a pas de repère visuel ou tactile pour guider sa réinspection mais doit se baser sur le souvenir du discours oral et de sa temporalité et l'utilisation des fonctions de navigation implémentées. Ainsi la réinspection passe alors par des phases de navigation où certains morceaux d'information verbale doivent être traités jusqu'à trouver l'emplacement désiré, ce qui implique plus de ressources cognitives mises en jeu et un facteur temps plus important.

On peut également opposer l'audio au visuel et au tactile sur le plan du contrôle, des manipulations exercé par l'utilisateur pour accéder à l'information. En effet, en lecture tactile ou visuelle le lecteur contrôle le flux d'information à tout instant, tandis que l'audio impose à l'auditeur une certaine passivité dans l'écoute, dans le sens où il ne contrôle pas de manière fine ni le débit ni la réinspection comme nous l'avons vu, et mobilise ainsi plus sa mémoire de travail. Le *transient information effect* [Leahy and Sweller 2011] stipule que lorsqu'une information complexe est présentée de façon transitoire (par exemple à l'oral, ou une animation en visuel), l'apprentissage est réduit par rapport à une situation où la même information est présentée de façon permanente (par exemple à l'écrit). Les auteurs pointent du doigt la fugacité de l'information et la difficulté voire l'impossibilité de réinspecter et d'accéder de nouveau à l'information comme un facteur de surcharge de la mémoire de la travail particulièrement pénalisant lorsqu'un l'information est complexe et non familière.

Enfin, un dernier point à considérer pour comparer l'audio et le tactile est l'apprentissage nécessaire à l'usage de ces modalités. En effet le braille exige un contrôle psychomoteur particulièrement fin et la mise en place de stratégies de lecture adaptées pour être efficace [Millar 2003]. De plus, en amont



de son apprentissage l'accès à des structures spécialisées pour son enseignement et certaines barrières socioculturelles ont, entre autres, fait que le braille est très peu pratiqué : par seulement 10% des non-voyants aux États-Unis [National Federation of the Blind 2009], et par seulement 1% des déficients visuels (incluant malvoyants et non-voyants) en France[Sander et al. 2005]<sup>20</sup>. Il est toutefois bon de préciser qu'en fonction de l'âge d'apparition de la déficience, de sa sévérité visuelle ainsi que du milieu socio-économique des personnes en situation de handicap, le taux de littéracie braille peut varier d'un extrême à l'autre.

Le développement des technologies de synthèse vocale permettant l'accès à l'écrit sans apprentissage particulier est également mis en avant comme étant un facteur décourageant les apprenants. Le braille reste toutefois indispensable d'après les enseignants spécialisés [Wittenstein and Pardee 1996] qui l'utilisent préférentiellement pour l'enseignement [Tsiakali and Argyropoulos 2015], car il permet contrairement à l'oral de comprendre la ponctuation et l'orthographe, et de manière générale de mieux appréhender la construction du langage. Les enseignants spécialisés mettent souvent en avant le fait que le braille est indispensable pour la compréhension de la langue et le développement cognitif, même si aucune étude à notre connaissance ne compare le développement du langage et des capacités cognitives des apprenants utilisant principalement du braille comparativement à d'autres n'utilisant que la synthèse vocale. Les non-voyants braillistes seraient d'ailleurs plus employés et mieux éduqués que les non-braillistes [Ryles 1996] ; loin de montrer que l'apprentissage du braille est un facteur causal, ces chiffres soulignent l'importance de l'accès à l'éducation pour les déficients visuels.

### **3.3. Quelles solutions pour pallier aux contraintes des modes audio et tactile ?**

---

Nous présenterons dans cette section un panorama des travaux de recherche autour de l'accès non-visuel à l'information sous l'angle des limites des modes audio et tactile tels qu'ils ont été présentés, à savoir la séquentialité de l'accès à l'information, la vitesse d'accès, et la charge cognitive résultante.

Une première approche pour diminuer l'impact du facteur temps et de la séquentialité est d'augmenter sensiblement le débit des synthèses vocales [Trouvain 2005; Asakawa and Takagi 2003], qui se trouvent être compréhensibles à des débits plus élevés que ceux de la lecture visuelle « silencieuse ». Ces études ne prennent cependant pas en compte la charge cognitive induite par l'augmentation du débit des voix de synthèse.

Une autre approche consiste à utiliser le potentiel expressif des modes audio et tactile pour compresser l'information verbale sous une forme plus iconique. Par exemple on peut utiliser à l'oral la prosodie, les sons non-verbaux ou la spatialisation du son pour véhiculer de l'information (voir 5.1.a pour plus de détails). Par exemple [Hussain et al. 2014] remplacent dans l'oralisation d'un texte les items les plus fréquents par des sons non-verbaux. Les indices non-verbaux, avec ici le son spatialisé, semblent d'ailleurs plus efficaces en situation de charge cognitive élevée qu'une contrepartie verbale [Klatzky et al. 2006] pour guider des piétons non-voyants dans une tâche de navigation. En tactile, la définition de patterns iconiques permet de véhiculer rapidement une

---

<sup>20</sup> Ce chiffre est à mettre en perspective avec la proportion de cécités parmi les déficients visuels qui serait d'après cette étude d'environ 20%. De plus, cette même étude montre qu'en France 39% des déficients visuels auraient 75 ans ou plus et 61% auraient 60 ans ou plus.

information avec un nombre minimal de cellules braille [Pietrzak 2008] (voir 5.1.b). Enfin, pour réduire la quantité d'informations à transmettre, certains auteurs utilisent des méthodes de résumé de texte [Arne et al. 2008] ou de lecture en diagonale (*skim-reading* [Ahmed et al. 2012] qui permettent d'obtenir en peu de mots un aperçu d'un contenu. D'autres travaux se concentrent eux sur la catégorisation des contenus, afin de permettre de filtrer les contenus jugés non pertinents [Islam et al. 2012], réduisant ainsi la charge cognitive et le facteur temps. D'autres, enfin, proposent de fragmenter et réduire l'information au plus important et rassembler ces fragments dans une table des matières [Yesilada et al. 2007] dans le même objectif.

Une autre possibilité est d'utiliser plusieurs modalités en même temps pour transmettre l'information, en réservant certaines modalités pour certains types d'informations pour éviter de surcharger la mémoire de travail, ou en établissant certaines complémentarités entre les modalités afin d'assurer une interprétation plus facile de l'information [Awdé 2009], en accord avec le modèle de la mémoire de travail de Baddeley présenté précédemment. La multimodalité permet ainsi de limiter la charge cognitive en utilisant par exemple à la fois la parole, des sons non-verbaux et des informations tactiles avec des vibreurs comme l'a fait le dernier auteur cité.

L'effet *cocktail-party*<sup>21</sup> a déjà été utilisé en recherche pour proposer une solution au problème de la séquentialité de l'information à l'oral [Guerreiro and Gonçalves 2014]. L'idée est de jouer plusieurs pistes audio simultanément, ici plusieurs synthèses vocales lisant des textes différents, et de laisser soin à l'auditeur de centrer son attention sur la voix qui l'intéresse c'est-à-dire dont les mots clés qu'elle lit sont en rapport avec une information recherchée. Ces auteurs ont en effet montré que des auditeurs non-voyants pouvaient sélectionner (porter leur attention sur) un locuteur à l'oral parmi deux ou trois parlant simultanément, et ce en fonction de certains mots clés recherchés.

Enfin, la substitution sensorielle vise à rendre interprétable, à l'aide des mécanismes de plasticité cérébrale, des informations d'un canal sensoriel de départ transmises à travers un autre canal sensoriel. Par exemple, des images simples peuvent être retransmises par une matrices d'électrodes placées sur la langue et rendues interprétables avec l'entraînement, ou encore la retransmission d'images à travers une matrice vibrotactile [Bach-y-Rita et al. 2003]. Cependant, les dispositifs de substitution sensorielle restent marginalement utilisés et relativement peu ergonomiques [Lenay et al. 2003], et leurs applications restent limitées. Cette approche a cependant le mérite de couper court aux limites des modes audio et tactile en permettant de retransmettre des stimuli visuels.

### **3.4. Conclusion et implications pour l'accessibilité des contenus numériques**

---

Même si la plasticité cérébrale permet aux non-voyants congénitaux d'optimiser le traitement de l'information audio et tactile, ce ne semble pas être le cas avec d'autres types de déficiences visuelles. De plus les caractéristiques des modes audio et tactile font peser un certain nombre de contraintes pour l'acquisition d'information dans ces modes. Les deux contraintes principales que l'on peut mentionner sont une contrainte temporelle et une contrainte cognitive, induites par la séquentialité de l'acquisition d'information dans ces modalités, les débits d'acquisition plus lents, et

---

<sup>21</sup> L'effet « soirée cocktail » stipule qu'un auditeur humain est capable de porter son attention sur un seul locuteur à l'oral même dans une situation impliquant plusieurs locuteurs en simultanément et des bruits de fonds, comme par exemple pour écouter une seule personne lors d'une « soirée cocktails ». De plus le changement attentionnel d'un locuteur à l'autre peut se faire à l'écoute de certains mots clés (ex. le nom de l'auditeur)

l'absence de multimodalité permettant un traitement optimisé en mémoire de travail. On peut également citer la fugacité et la passivité devant l'information orale qui aggravent les facteurs temps et cognitif.

Un autre problème impactant indirectement ces contraintes est la transposition de l'information du mode visuel aux modes audio et tactile, ce que [Maurel 2004] désigne par le terme *transmodalité*, en précisant que la problématique première se résume à comment « transposer sans pertes ». La question est donc de savoir comment rendre l'ensemble (ou à défaut le plus possible) des informations visuelles *disponibles* dans d'autres modes, ce qui représente un premier niveau d'accessibilité de l'information mais pose pourtant un certain nombre de problèmes. C'est ce but de restitution *exhaustive* de l'information visuelle qui est recherché par la majorité des travaux appliqués au cas de la déficience visuelle, et c'est également ce qui est visé par les normes en vigueur relatives à l'accessibilité des contenus numériques (voir partie 4). Cependant, arriver à transposer l'information visuelle (verbalement par exemple) peut se faire au détriment du temps et de la charge cognitive ; par exemple une description verbale d'une figure géométrique ou de la structure d'une page ne sera probablement pas aussi efficace que sa forme visuelle initiale. Cette problématique rejoint un second niveau d'accessibilité qui englobe l'utilisabilité (voir 0), et qui doit être abordé de manière *holistique* [Giraud 2014; Phipps and Kelly 2006] c'est à dire en prenant en compte à la fois les caractéristiques des utilisateurs (de la déficience visuelle) mais aussi du contexte d'utilisation ; c'est cette approche que nous avons adoptée et défendons dans nos travaux.

En ce qui concerne l'efficacité de la transposition de l'information visuelle, on subodore déjà à ce stade que plus l'objet à transposer fera une utilisation intensive des caractéristiques spécifiques au mode visuel (disposition spatiale, couleur, taille, forme, etc.), plus cette dernière sera complexe. Les interfaces « graphiques » qui permettent d'interagir avec les périphériques informatiques font justement un usage intensif de l'espace et des propriétés du mode visuel pour organiser l'information. [Dourish 2001] illustre d'ailleurs cet usage en disant que le passage des interfaces textuelles aux interfaces graphiques a transformé (p.11) « la tâche de gestion de l'information en gestion de l'espace ». Ce dernier auteur cite d'ailleurs quelques propriétés des interfaces graphiques exploitant les capacités humaines à traiter l'information visuelle :

- § La *reconnaissance de patterns et le raisonnement spatial* permettent des regroupements et une organisation spatiale de l'information, ce qui fait écho à ce que l'on peut appeler l'accessibilité de la structure de l'information [Archambault 2010](p.13) et pour laquelle la théorie de la mise en forme donne plusieurs pistes pour formaliser cette information (voir Chapitre I).
- § La *densité informationnelle* permise par le visuel où la disposition spatiale, les images, symboles et le texte se complètent pour former un message. Cette densité pose problème au regard des contraintes des modes audio et tactile mentionnées.
- § *L'attention périphérique* est utilisée pour gérer l'espace informationnel (ex. les notifications dans la barre des tâches) ; un équivalent à l'oral ou en tactile semble problématique étant donné la nature séquentielle de ces modes.
- § *Les métaphores visuelles* sont utilisées pour faciliter l'appréhension des interfaces graphiques, conjointement avec la manipulation directe des objets utilisant ces métaphores. La question d'une transposition efficace de ces métaphores à l'oral ou en tactile reste une question ouverte.

Ainsi, les interfaces graphiques utilisent de manière extensive les propriétés visuelles pour structurer et manipuler l'information. Les technologies d'assistance (voir 5), en transposant cet accès dans des modes non-visuel, engendrent bien souvent les mêmes problèmes que ceux évoqués précédemment à savoir un facteur temps pouvant être « 2 à 6 fois supérieur pour une même opération si l'on prend en compte les actions préparatoires et périphériques » [Uzan 2005]<sup>22</sup>, des contraintes cognitives plus importantes [Giraud 2014] et des problèmes de transposition de la structure visuelle de l'information [Francisco-Revilla and Crow 2010; Lazar et al. 2007]. À noter que ces différents problèmes sont également présents dans d'autres situations telles que l'apprentissage dans un cadre inclusif des élèves déficients visuels (voir l'étude en situation rapportée dans la partie 7).

En conclusion, l'accès à l'information pour les déficients visuels revêt d'un problème double : comment rendre disponible des informations visuelles dans des modes non-visuels, et comment faire en sorte que cette restitution soit efficace malgré les contraintes des modes audio et tactile ? Ces problèmes peuvent être vus comme constituant deux étapes à franchir pour tendre vers une utilisabilité universelle de l'information. Le cadre de ces travaux se limitant à l'accessibilité des contenus numériques (et plus précisément du texte et de sa mise en forme visuelle), les parties suivantes présenteront l'état de l'art de la disponibilité des contenus numériques d'un côté et l'état de l'art de l'accessibilité de ces contenus permise par les technologies d'assistance actuelles. Pour terminer, même si l'accès aux contenus numériques semble contraignant au regard de ce qui a été mentionné, certaines solutions sont pointées du doigt pour améliorer les performances et limiter la charge mentale comme utiliser la multimodalité mêlant tactile, parole et sons [Awdé 2009; Maurel 2004], ou sélectionner les informations présentées sur la base de la pertinence avec les objectifs de la tâche [Giraud 2014], dont nous nous sommes d'ailleurs inspiré pour concevoir une restitution efficiente de la mise en forme (voir Chapitre IV).

---

<sup>22</sup> Dans une étude comparative entre voyants et déficients visuels sur les mêmes tâches impliquant des pages web (ex. acheter un album de musique), [Giraud 2014] a montré les temps d'exécution des déficients visuels pouvaient être jusqu'à 12 fois supérieurs aux temps d'exécution des utilisateurs voyants.

## **4. Approches actuelles pour l'accessibilité des contenus numériques**

---

---

On présentera dans cette partie les différentes approches et normes actuellement en vigueur pour rendre un contenu « accessible », c'est-à-dire pour faire en sorte que les informations qu'il véhicule et que les interactions qu'il propose (par exemple pour les pages web) soient disponibles également aux utilisateurs déficients. L'accessibilité visée est ici une accessibilité de moyens, c'est-à-dire qu'on essaie d'assurer que les informations et interactions visuelles (mise en forme, contenus dynamiques, images, etc.) soient retranscrites et donc disponibles.

Tout d'abord on verra quelles sont les approches générales que peuvent adopter les créateurs de contenus pour produire des contenus accessibles. On présentera ensuite les différents formats de contenus numériques existants, en nous limitant aux documents à visée textuelle (HTML, XML, PDF, EPUB). Le but est de présenter les différentes approches utilisées pour l'accessibilité de ces formats, et de voir dans quelle mesure ces derniers prennent en compte la mise en forme matérielle (la sémantique de la mise en forme visuelle en quelque sorte) ou pourraient l'intégrer dans leur codage de l'information. Enfin, on introduira les normes existantes pour l'accessibilité des contenus numériques ainsi que leurs limites.

### **4.1. Approches générales pour la création de contenus accessibles**

---

On distingue deux approches générales pour donner accès aux non-voyants aux contenus numériques.

La première approche, qualifiable de *normative*, consiste à enrichir le document d'origine en rajoutant certaines informations (ou méta-informations) en respectant un ensemble de règles quant à sa structuration et au langage interne qui permet de le décrire. Le but est de rendre le contenu accessible via diverses aides techniques, matérielles ou logicielles. Au lieu de produire une version alternative d'un document, l'auteur devra par exemple respecter au sein de son document d'origine les WCAG (voir 4.3.a), requérant notamment la présence d'une alternative textuelle pour chaque image. Cette solution assure la disponibilité de l'information, c'est-à-dire qu'une personne en situation de handicap visuel peut éventuellement y accéder, mais ne garantit pas que cet accès se fera de manière efficace et dépourvue d'erreurs comme nous l'avons vu précédemment. Comme déjà évoqué, des outils évaluant automatiquement la conformité des contenus produits avec les normes en vigueur existent, mais ne permettent pas d'évaluer la qualité de l'accessibilité en général. Cette approche garantit donc en partie la disponibilité de l'information, si elle est correctement appliquée et sous réserve de la pertinence des normes à cet égard, mais c'est aux technologies d'assistance de garantir que l'interaction soit efficace, efficiente et satisfaisante.

La seconde est de produire ces documents dédiés, c'est-à-dire de créer une nouvelle version de chaque document, média ou page web, conçues alors spécifiquement pour une population cible, par exemple les non-voyants. C'est le cas lorsqu'une transcription braille ou une version audio dédiée est proposée pour un document ou un ouvrage. Cependant, cette stratégie est coûteuse et donc rarement utilisée, et semble donc difficilement généralisable au web au regard de la quantité de documents numériques mis en ligne chaque jour, d'autant que les normes d'accessibilité sont déjà

assez peu appliquées (voir section précédente). De plus, la création de versions dédiées aux utilisateurs déficients peut être perçue comme une mise à l'écart [Nevile 2005] jouant en défaveur du sentiment d'inclusion sociale. Cette solution peut tout de même être viable lorsque les pages ou documents d'origine présentent une trop grande complexité graphique, auquel cas même le respect des normes d'accessibilité peut ne pas être suffisant et la production d'une version simplifiée peut s'avérer plus accessible [Lunn et al. 2008]. On retrouve tout de même des exemples de cette approche sur la toile, avec notamment le site de la ligue de football européenne qui existe en version dite « accessible »<sup>23</sup> pour laquelle le contenu a été épuré. Enfin, une initiative a vu le jour en 2014 qui propose des versions entièrement remaniées de plusieurs centaines de site web. VoxiWeb<sup>24</sup> propose à ses abonnés d'accéder à des versions « nettoyées, structurées et accessibles » de ces sites, incluant nativement une synthèse vocale pour la lecture des pages. Le lancement récent des applications mobiles sous Android et iOS témoigne d'un certain succès de l'offre, ce qui tend à confirmer les limites de la première approche « normative ».

Enfin, on peut citer une approche mêlant les deux précédentes, développée par des chercheurs de l'Université de Manchester, qu'ils dénomment le *transcodage* de pages Web. Étant donné que produire des versions dédiées de pages est coûteux pour les développeurs, l'idée est de réorganiser automatiquement les pages web sur la base d'informations sémantiques rajoutées (annotées manuellement ici) dans les pages, afin d'en proposer des versions plus accessibles aux technologies d'assistance. Leur approche dénommée DANTE [Yesilada et al. 2007] est basée sur l'utilisation d'une ontologie<sup>25</sup> spécifique décrivant les rôles et les types d'éléments retrouvés dans les pages Web pour les annoter ; les pages Web annotées sont ensuite *transcodées*, c'est-à-dire fragmentées, filtrées et réorganisées avec un ensemble de méthodes décrites dans leur article. Une première implémentation de cette approche a été proposée par [Bechhofer et al. 2006], annotant les fichiers CSS<sup>26</sup> des pages HTML à l'aide de l'ontologie qu'ils ont développée, pour ensuite *transcoder* ces pages en filtrant les informations périphériques, réduisant les menus de navigation et plaçant les informations importantes en début de page ; cette implémentation n'a pas été évaluée. Une autre implémentation de *transcodage* a été testée, basée sur l'annotation directe des fichiers HTML, qui fragmente chaque page Web en plusieurs pages avec peu de contenu et dont les contenus périphériques ont été supprimés, et reliant ces différentes « sous-pages » à l'aide d'une table des matières. Cette approche a montré des résultats encourageants, en permettant d'alléger la charge cognitive des utilisateurs et en augmentant leur mobilité comparativement aux pages d'origine, dans le cadre d'une tâche demandant de rechercher une page et de résumer son contenu.

### 4.2. Panorama des formats numériques pour documents textuels

---

Nous ferons mention des principaux formats de fichiers numériques existant et comment ils ont été ou pourraient être rendus accessibles. Cette partie a pour objectif de mieux comprendre les logiques et contraintes de ces langages sous l'angle de l'accessibilité, ainsi que leurs capacités à intégrer les informations de MFM.

---

<sup>23</sup> Site de l'UEFA classique : <http://fr.uefa.com/> ; et en version accessible : <http://ac.uefa.com/fr/>

<sup>24</sup> Voir <https://www.voxiweb.com/>

<sup>25</sup> Une ontologie en Informatique est une structure représentant des connaissances. Une ontologie regroupe un ensemble de concepts ainsi que les relations (au sens large) entre ces concepts. (ex. un « homme » est un « individu »). En ce sens ce sont des « spécifications de conceptualisations » [Gruber 1995].

<sup>26</sup> *Cascading Style Sheets*, fichiers définissant des styles de mise en forme utilisés dans les pages HTML

#### ***4.2.a. Niveaux de description du contenu : de la présentation à la sémantique***

---

Cette partie vise à introduire les différents niveaux de description du contenu que les formats numériques peuvent intégrer et auxquels on fera référence dans les parties suivantes. La première logique est celle de présentation de l'information ; par exemple le titre précédent peut être décrit dans le fichier par ses propriétés visuelles (*Police Times New Roman, 12 pts, Italique*). Cela est suffisant pour produire l'affichage du contenu. Cependant, dès lors que l'on souhaite manipuler cet objet, pour extraire de l'information du document, générer une table des matières, ou si un utilisateur non-voyant souhaite y accéder, alors cette logique de présentation ne suffit pas car elle est dépourvue du *signifié* de ces propriétés visuelles qui désignent cet objet en tant que titre (avec un certain niveau hiérarchique).

Cependant le niveau « sémantique » peut être décrit de bien des façons différentes. Par exemple, sur une page web, on peut décrire un même objet en fonction de : ses propriétés visuelles, son rôle pour la navigation, son rôle au sein du discours, ce qu'il représente, etc. De plus, la description peut se faire à plusieurs niveaux de segmentation, allant de la description des caractéristiques du document à celle de son contenu propositionnel. Cela est fort bien illustré par les différents formats de métadonnées permettant de décrire les documents web et leurs contenus, développés dans le cadre du web sémantique [Berners-Lee et al. 2001] dont l'objectif premier est d'extraire des connaissances des documents et de les mettre en relation à un niveau global. On peut par exemple citer le Dublin-Core qui est un format de métadonnées décrivant les caractéristiques des documents (ex. auteur, titre, date de publication, etc.) et l'ontologie FOAF (*Friend Of A Friend*) qui permet de décrire les individus mentionnés dans le contenu et les relations qu'ils entretiennent.

#### ***4.2.b. Langages de markup et formats courants***

---

Nombre de format de fichiers partagent la même logique de balisage (ou *markup*) : le fichier est formée d'un ensemble de balises pouvant être imbriquées dans d'autres balises et contenir du texte ou d'autres balises (exemple : <balise> texte </balise>). On retrouve ces langages dans les pages HTML sur le web, dans les fichiers PDF<sup>27</sup> ou encore dans les e-books (livres & documents électroniques) au format EPub<sup>28</sup>. Les balises pouvant englober du texte, ces dernières permettent de typer et de définir certaines propriétés (d'affichage notamment) à appliquer au texte ou encore spécifier sa nature ou son rôle (ce qu'on appelle des métadonnées). L'idée derrière ces langages de balisage est de séparer la logique structurelle interne du document de la façon dont il sera présenté.

#### **XML : eXtensible Markup Language**

XML est un langage de balises, et désigne aussi le format de fichiers du même nom. XML n'a cependant du langage que les contraintes grammaticales, car il est extensible dans le sens où le vocabulaire utilisable pour les balises n'est pas fixe, mais défini par l'auteur. C'est donc un langage générique pour lequel il existe une multitude de vocabulaire définis, notamment pour certains formats de fichiers comme le PDF ou les fichiers MS-WORD qui utilisent du XML dans la construction du document final.

#### **HTML : Hypertext Markup Language**

---

<sup>27</sup> Portable Document Format

<sup>28</sup> Electronic Publication

HTML est le langage utilisé pour représenter les pages web, « HTML » désignant également le format du même nom. HTML peut être vu comme une implémentation particulière du XML, avec des règles et un vocabulaire particulier. Le vocabulaire utilisé dans HTML comprend des balises génériques utilisées pour la présentation (ex. <div>, une division de la page) et des balises à portée sémantique (ex. <h1>, un titre de niveau 1). Le HTML ne gère pas les caractéristiques de l'affichage de la page, ces propriétés sont définies dans des feuilles de style à part spécifiant le rendu de chaque balise, mais simplement la logique structurelle de la page. Tout le problème pour l'accessibilité est que cette logique mélange justement présentation et sémantique, et que l'usage des balises est laissé au soin des auteurs de contenus. Ainsi des balises à portée sémantique peuvent voir leur utilisation détournée (ex. <h1> utilisé pour une citation et pas un titre ; un autre exemple assez répandu est l'utilisation de tableaux <table> pour découper une page en zones et non pas pour un tableau de données), et inversement des balises génériques de présentation peuvent être utilisées à la place des balises à portée sémantiques initialement prévues (ex. une <div> spécifique peut être utilisée pour un titre). Cette liberté donnée aux auteurs dans l'utilisation des balises a engendré une grande hétérogénéité des contenus mêmes, malgré l'existence de recommandations et bonnes pratiques à suivre qui ne sont ni toujours connues ni toujours suivies. Tout le problème pour l'accessibilité est donc que les balises soient utilisées correctement, ce qui renvoie aux normes d'accessibilité définies par le WCAG (voir 4.3), qui ne sont à l'heure actuelle que marginalement appliquées.

En ce qui concerne les balises à portée sémantique, elles sont en nombre limité pour ce qui est des objets textuels mis en forme : on peut citer les balises de titre, paragraphe, liste, tableau, emphase et citation. Nous avons vu dans le Chapitre I que de nombreux autres architectures textuelles, porteuses de sens, existaient. Cependant, l'ajout de métadonnées au code HTML est possible à travers différents attributs des balises.

### **MathML : Mathematical Markup Language**

MathML est un vocabulaire XML utilisé pour décrire les formules mathématiques, et peut notamment être utilisé au sein d'une page HTML depuis sa version 5. Même si nos travaux ne portent pas sur les mathématiques, il est intéressant de voir que les expressions mathématiques illustrent le problème de l'accessibilité des contenus. MathML comprend deux sous-langages, un langage de présentation et un langage sémantique. Le premier assure que l'expression soit affichée comme spécifié dans le code XML, alors que le langage sémantique conserve uniquement la formule mathématique en elle-même et laisse le soin de l'affichage aux applications utilisateur.

### **EPub : Electronic Publication & DAISY**

EPub est un standard utilisé pour les livres électroniques, basé notamment sur des fichiers XML. Il est utilisé au sein d'un autre standard dénommé DAISY (Digital Accessible Information SYstem) pour la publication de livres audio. EPub a été créé pour la distribution de publications numériques sous format unique, avec la plus grande interopérabilité possible. En termes de marquage sémantique il offre les mêmes possibilités que le langage HTML. L'avantage du standard DAISY est notamment de réguler l'utilisation du marquage dans les fichiers EPub afin d'assurer sa cohérence.

### **PDF : Portable Document Format**

PDF est un format de fichier extrêmement répandu dont le principal atout est de conserver la mise en forme du document quelle que soit le périphérique de lecture utilisé, offrant ainsi un format de document « portable ». Étant donné que ce format était initialement prévu pour de l'affichage



uniquement, le marquage sémantique dans les documents est limité aux en-têtes et pieds de page ainsi qu'à certains marques pages (signets) insérables dans le document.

## Conclusion

Les formats de fichiers présentés laissent tous à l'auteur la liberté d'utiliser les balises qu'il souhaite, et ainsi ne permettent pas d'assurer la cohérence de la sémantique interne des documents. Pour cela certaines normes ont été créées (voir partie suivante) afin que le marquage sémantique soit cohérent, et plus généralement pour assurer la disponibilité (l'accessibilité) des informations dans ces formats. La tâche est d'autant plus difficile que ces formats ont été initialement créés pour produire des documents destinés à une lecture visuelle. Enfin, nous avons vu que ces formats ne prenaient en compte qu'une partie des informations significatives de MFM dans leurs vocabulaires.

### 4.2.c. *GeM : Un langage de description unifié*

---

GeM (*Genre and Multimodality*) est une méthodologie d'annotation utilisée pour décrire sur un même plan les regroupements graphiques et les relations rhétoriques du discours, afin d'étudier les variations des utilisations du verbal et du visuel à travers les différents genres de documents [Bateman 2008]. Le principe est basé sur le découpage du document en unités de bas (phrases, images, icônes), la segmentation se faisant sur la base des ruptures graphiques, et de constituer ensuite les regroupements graphiques d'une part et les différents segments propositionnels du discours d'autres part (sur la base de la théorie des structures rhétoriques [Mann and Thompson 1987]), sur la base de ces unités. On retrouve aussi une annotation particulière pour les éléments de navigation dans le document (pointeurs, numéros de pages, références croisées, etc.), et une annotation des propriétés graphiques des regroupements (taille, couleur, disposition, etc.).

GeM permet d'illustrer ici la complexité d'un document dans sa forme visuelle et donc la complexité de son accessibilité dans des modes non-visuels. Le point central d'étude dans GeM est le lien entre l'inscription graphique du discours et sa structure rhétorique, les deux étant intrinsèquement liés, et chacun complétant l'autre [Luc 2000]. Le premier challenge, qui est ici partiellement adressé dans ces travaux, est de savoir comment restituer le discours sans le visuel tout en conservant la sémantique initiale ? Mais en marge de cette sémantique, il semble également important que les informations visuelle initiales (combien y-a-t-il de zones ? où se trouve le titre principal ?) puissent être accessibles notamment pour que les déficients puissent avoir un référentiel commun avec les non-déficients.

## 4.3. Les normes pour l'accessibilité des contenus numériques

---

### 4.3.a. *Portée et limites des « Web Content Accessibility Guidelines »*

---

Comme évoqué précédemment, le *World Wide Web Consortium (W3C)* est à l'origine de la production de normes d'accessibilité pour les documents web, incluant les formats et technologies web (HTML, CSS, JavaScript et Flash), les fichiers images (PNG, SVG) et les documents PDF. Ces normes, labellisées *WCAG* pour *Web Content Accessibility Guidelines* définissent un ensemble de critères ainsi que leurs implémentations techniques afin d'assurer l'accessibilité (la disponibilité de l'information) du web aux personnes déficientes ; à noter que même si certains formats particuliers sont visés, ces normes se veulent les plus génériques possibles indépendamment de leur mise en œuvre technique.

Une première version des WCAG a été éditée en 1999, puis une seconde plus récemment en 2008. Ces normes (dans leur version 2.0) définissent 4 critères généraux constituant 4 grandes classes de critères d'accessibilité ; le contenu et les interactions proposées doivent ainsi être : *Perceptibles*, *Fonctionnelles (opérables)*, *Compréhensibles* et *Robustes*. Ci-dessous une copie des critères généraux définis par les WCAG<sup>29</sup>.

### Perceptible

- § Fournir des **alternatives textuelles** aux contenus non-textuels
- § Fournir des **légendes** et autres alternatives pour les contenus multimédia
- § Le contenu doit pouvoir être **présenté de différentes façons**, notamment par les technologies d'assistance, sans perte de sa signification
- § Faire en sorte que le contenu soit facilement accessible **en visuel comme à l'oral**

### Fonctionnel

- § Toutes les fonctionnalités doivent être **accessibles à l'aide du clavier**
- § Donner **assez de temps** aux utilisateurs pour lire et utiliser le contenu
- § Ne pas utiliser de contenu provoquant **d'attaques** (épileptiques)
- § Guider les utilisateurs pour **naviguer et trouver** les contenus

### Compréhensible

- § Faire en sorte que le texte soit **lisible et compréhensible**
- § Faire en sorte que les contenus se comportent **de manière prédictible**
- § Aider les utilisateurs à **éviter et corriger les erreurs**

### Robuste

- § Maximiser la **compatibilité** avec les outils actuels et futurs.

Pour chaque format de fichier et technologie, ces critères sont développés en un certain nombre de recommandations techniques qui sont en fait l'implémentation de ces critères. Par exemple, on retrouve 59 recommandations pour les fichiers HTML et XHTML<sup>30</sup>, incluant notamment la définition d'alternatives textuelles ou l'utilisation de certains types d'attributs pour les balises. À noter que les WCAG définissent trois niveaux de conformité aux recommandations (A, AA et AAA), dont l'atteinte est conditionnée par des « Critères de Succès » (*Success Criteria*) qui sont définis indépendamment de l'implémentation technique. Malgré un certain succès technique et politique [Kelly et al. 2007], le modèle d'accessibilité adopté a été largement critiqué et montre en effet certaines limites tant dans l'implémentation des recommandations que dans leur efficacité.

Dans un journal en ligne destiné aux développeurs Web, [Clark 2006] a publié un article (au titre évocateur : *To Hell With WCAG 2*) mettant l'accent sur une série de limitations des WCAG, en citant notamment : la taille extensive de la documentation technique, le langage parfois difficilement déchiffrable utilisé dans une optique de généricité par rapport aux technologies, l'absence de standards de validation pour la conformité aux recommandations, et enfin l'omission des déficiences cognitives dans les recommandations. Ces problèmes semblent bien réels si l'on regarde le taux d'application des WCAG sur le web : en 2004, la *Disability Rights Comissions* a montré que seul 19% des 1000 sites web inspectés présentaient un niveau de conformité minimal (niveau A) avec la première version des WCAG (1.0) [Commission 2004]. Une étude longitudinale sur une période de 10

<sup>29</sup> Disponibles ici : <http://www.w3.org/WAI/WCAG20/glance/>

<sup>30</sup> Disponibles ici : <http://www.w3.org/TR/2014/NOTE-WCAG20-TECHS-20140408/html.html>

ans portant sur 6000 pages a montré que même si le taux d'application des WCAG 1.0 était en augmentation, il est passé de 2 à seulement 10% pour l'échantillon étudié [Harper and Chen 2012] ; ces derniers auteurs précisent que l'apparition de la version 2.0 des WCAG n'aura probablement pas d'effet sur le taux d'application des recommandations, bien qu'elles aient le mérite de présenter des recommandations plus évolutives au regard des nouvelles technologies. Enfin, au-delà des problèmes liés aux recommandations elles-mêmes, leur faible utilisation s'explique aussi par le manque de sensibilisation des développeurs et webmestres à l'accessibilité. [Lazar et al. 2004] ont en effet montré que même si les créateurs de contenu web étaient de plus en plus sensibles à l'accessibilité, 22% des propriétaires de sites web ignoraient l'existence des recommandations d'accessibilité du W3C.

Un autre aspect important est que le respect des WCAG ne suffit pas pour assurer l'accessibilité des contenus. L'étude de la *Disability Rights Comissions* [Commission 2004] a mis en évidence l'absence de relation entre le nombre de violations des WCAG 1.0 et les mesures subjectives ou objectives des capacités des participants à utiliser les sites web correspondants. De plus, [Petrie et al. 2005] ont étudié plusieurs sites Web de Muséums qui, bien que respectant les WCAG, n'étaient pas perçus comme accessibles par les utilisateurs. Plus récemment, [Power et al. 2012] ont montré que seuls la moitié des problèmes rencontrés par les utilisateurs non-voyants pouvaient être résolus par l'application de certaines des recommandations du W3C. Ainsi, certains auteurs recommandent une approche plus holistique de l'accessibilité, qui au-delà des critères techniques définis par les WCAG prenne en compte l'utilisateur et le contexte d'utilisation dans la création des contenus [Kelly et al. 2007; Newell and Gregor 2000; Giraud 2014], rejoignant ainsi une démarche ergonomique. Cependant, cette première approche normative peut sembler d'autre part nécessaire, étant donné l'hétérogénéité des contenus web (voir 4.2.b). Une étude menée auprès de 1465 déficients visuels en 2014<sup>31</sup> a montré que seulement 36% d'entre eux considéraient qu'il y avait eu des progrès en ce qui concerne l'accessibilité Web au cours de l'année précédente.

#### ***4.3.b. Des normes basées sur l'annotation sémantique du contenu : « Accessible Rich Internet Applications »***

---

Un avancement notable dans les normes d'accessibilités aux contenus numérique a été la création début 2014 des normes **ARIA** (*Accessible Rich Internet Applications*)<sup>32</sup> par le W3C. Ces normes permettent un marquage sémantique de la structure, des composants (*widgets*) et des comportements des pages Web à l'aide d'un vocabulaire défini. La nature sémantique du marquage implique que ce dernier dépend du rôle que joue véritablement les éléments marqués au sein de la page, au-delà de la logique de présentation du marquage HTML. Trois paradigmes ont été introduits dans le modèle proposé par les normes ARIA : le *rôle* ainsi que les *propriétés* et *états* des objets de la page.

Les propriétés et les états utilisent un vocabulaire décrivant des informations qui ne sont pas accessibles autrement que visuellement, permettant notamment la gestion de la mise à jour dynamique de zones de contenu dans une page. Concrètement, les propriétés et les états ne se distinguent que par la fréquence de mise à jour de ces informations, les propriétés étant dédiées à

---

<sup>31</sup> <http://webaim.org/projects/screenreadersurvey5/#progress>

<sup>32</sup> Voir <http://www.w3.org/TR/wai-aria/>. La première ébauche (*draft*) des WAI-ARIA date de 2006 [http://www.w3.org/TR/wai-aria-roadmap/#desktop\\_timeline](http://www.w3.org/TR/wai-aria-roadmap/#desktop_timeline)

des informations plus statiques et les états à des informations plus dynamiques. Les propriétés permettent par exemple de décrire le niveau hiérarchique d'un élément dans la structure du document (*aria-level*), ou encore certaines relations entre éléments, comme le fait qu'un élément labellise un ensemble d'autres éléments (*aria-labelledby*) ou qu'un élément soit composé d'un ensemble d'autres éléments (*aria-owns*, qui définit des relations parent/enfant de nature visuelle). On peut citer comme exemples d'états le caractère valide ou non d'une entrée de formulaire (*aria-invalid*) ou le caractère invisible d'un élément (*aria-hidden*), susceptible d'induire en erreur les utilisateurs n'ayant pas accès à cette information visuelle<sup>33</sup>. Les normes ARIA comprennent ainsi 35 propriétés et états attribuables aux éléments des pages Web, divisées en 4 catégories : relations entre éléments, contenus mis à jour dynamiquement, *drag and drop* et enfin les *widgets* (éléments interactifs des pages web utilisant des entrées utilisateur).

Le paradigme central des WAI-ARIA est celui de *rôle* des éléments des pages Web. Les rôles permettent de typer les éléments d'une page web et définissent également les états et propriétés pouvant être associées à chaque rôle. Les deux catégories principales de rôles sont les rôles des *widgets* et ceux des éléments la *structure* du document, séparant ainsi le contenu structuré du document des éléments interactifs<sup>34</sup>. Les rôles de type widget représentent donc la nature des éléments interactifs des pages web (ex. liste déroulante, bouton cliquable, etc.), permettant ainsi avec les propriétés associées de désambigüiser leur usage. Les rôles de type *structure* permettent eux de représenter la nature des structures visuelles du document et notamment de typer les regroupements ou zones du document. Les principaux rôles pour les zones du document sont : bannière (zone de présentation), zone de navigation, *main* (zone incluant le contenu principal), zone de recherche, article (contenu périphérique indépendant), zone de contenu *complémentaire* au contenu principal et zones d'*informations diverses* (bas de page, copyrights, etc.). L'utilisation de ce marquage pour définir les zones du document rend obsolète l'utilisation de liens renvoyant vers le contenu principal (*Skip to Main Content*). D'autres rôles définissent des types d'objets que l'on retrouve fréquemment dans les documents : les définitions, le contenu mathématique, les images, les tableaux et les titres. Il est d'ailleurs intéressant de noter que ces rôles sont des redites de balises HTML déjà existantes, ce qui dénote d'une volonté d'indépendance de la sémantique du contenu et de sa logique interne dans le code HTML, afin de résoudre les problèmes déjà mentionnés posés par ce dernier.

Ainsi, les normes ARIA définissent un vocabulaire pour le marquage sémantique des éléments des pages web, et permettent également la gestion des propriétés visuelles de ces éléments et de leur mise à jour dynamique. Ces normes et l'approche utilisée ici sera comparée à la méthodologie d'annotation de la mise en forme que nous avons définie dans le Chapitre III. La première version finale des normes ARIA étant datée de Mars 2014, aucune étude à notre connaissance n'a évalué son impact sur l'accessibilité ou l'utilisabilité des pages web ; cependant, l'approche adoptée basée sur la sémantique visuelle des contenus nous paraît des plus intéressante. Toutefois, ARIA rajoute une couche de complexité dans le développement des sites web, ce qui semble fortement en augmenter les coûts [Puzis et al. 2015] et rendre ARIA difficilement applicable à l'ensemble du Web. Néanmoins,

---

<sup>33</sup> Il est fréquent que les développeurs utilisent des objets invisibles à des fins fonctionnelles ; par exemple un élément textuel peut être rendu invisible pour que sa valeur soit réutilisée pour l'affichage à la suite d'une action utilisateur (ex. « Réponse enregistrée » ou « Champ Valide »).

<sup>34</sup> Les rôles ARIA suivent d'ailleurs une hiérarchie d'héritage, voir : [http://www.w3.org/TR/wai-aria/rdf\\_model.png](http://www.w3.org/TR/wai-aria/rdf_model.png)

l'utilisation de certains marquages ARIA comme celui des régions semble être de plus en plus utilisé<sup>35</sup>.

#### **4.3.c. Vérification de la conformité des documents aux normes du W3C**

---

En parallèle des normes d'accessibilité définies par les WCAG, le W3C propose une liste d'implémentations techniques à utiliser dans chacun des formats concernés, et ce pour chacun des critères normatifs définis. Il est donc relativement aisé de définir des outils automatiques ou semi-automatiques permettant de vérifier la conformité d'une page ou de tout un site Web avec les critères du WCAG, en se basant sur les implémentations techniques correspondantes. Le W3C propose d'ailleurs une liste d'outils, commerciaux ou libres, permettant d'évaluer l'accessibilité : <http://www.w3.org/WAI/ER/tools/>. Certains des outils listés par le W3C mais aussi certains travaux de recherche définissent des métriques d'accessibilité basées sur les WCAG, afin de rendre l'évaluation des critères qu'elles définissent plus fine, plus étendue ou plus automatique. [Vigo and Brajnik 2011] en étudient 10 différentes dans le but de définir un cadre d'évaluation unifié des métriques d'accessibilité basées sur les WCAG.

L'évaluation automatique de la conformité aux normes montre toutefois ses limites ; par exemple, on peut évaluer automatiquement plutôt facilement si une alternative textuelle est descriptive ou représente un simple label, mais il devient plus difficile d'évaluer la pertinence de la description faite par l'alternative textuelle<sup>36</sup>. Un autre exemple peut être plus parlant serait celui d'un site à la structure logique atypique, mais respectant tout de même les normes d'accessibilité. Plusieurs méthodes alternatives existent : évaluation par des experts, tests utilisateurs et ballades « cognitives » en sont quelques exemples, et semblent d'ailleurs être plus valides que la conformité aux normes dans certains cas [Brajnik 2008].

Concernant les normes ARIA, l'automatisation de l'évaluation de la conformité à ces normes semble délicate. Dans la liste d'outils de vérification automatique du W3C, un seul outil inclut dans sa description la conformité aux normes ARIA, mais il nous a été impossible de recueillir des informations sur son fonctionnement<sup>37</sup>. Une remarque d'importance concernant ARIA est que les normes peuvent être utilisées à plusieurs niveaux : par exemple uniquement pour gérer la mise à jour dynamique de contenu [Brown and Harper 2013], ou encore pour marquer certains objets ciblés comme les figures et leur labels correspondant [Freedom Scientific 2013], allant jusqu'à un marquage étendu de la sémantique de la page permis par les balises ARIA. De plus, la pertinence du marquage sémantique de l'information, au-delà de sa présence dans le code source, rejoint les difficultés rencontrées dans l'évaluation de la conformité aux WCAG avec des problématiques du traitement du sens global du document (de son visuel, du discours et de son contexte). Ces derniers points forment un début d'explication à l'absence de « checklist » pour vérifier l'application des normes ARIA sur le site du W3C, contrairement à ce qui a été fait pour le WCAG.

---

<sup>35</sup> Voir <http://webaim.org/projects/screenreadersurvey5/#landmarks>

<sup>36</sup> À noter que certains outils comme OCAWA, basés sur des systèmes experts, tendent à répondre à cette problématique. Voir : <http://www.ocawa.com>

<sup>37</sup> Le site internet correspondant [www.odellus.com](http://www.odellus.com) semble ne plus être disponible

### 4.4. Conclusion

---

Globalement, la création de contenus est très permissive en terme de sémantique du langage utilisé, étant donné que celle-ci n'impacte pas le rendu visuel final, ce qui a des effets négatifs sur l'accessibilité. Pour pallier cela, des normes ont vu le jour afin d'assurer une cohérence de la logique de structuration du contenu, et afin que le sens général puisse être conservé lors d'un accès non-visuel. Les normes les plus répandues sont les WCAG (*Web Content Accessibility Guidelines*), qui sont les normes d'accessibilité Web définies par le W3C (*World Wide Web Consortium*), ainsi que les normes ARIA (*Accessible Rich Internet Application*) qui ont été produites plus récemment par le même organisme. D'autres standards existent, principalement des standards nationaux définis pour l'accessibilité des sites web d'institutions ou de services publics ; en France, on peut citer le RGAA<sup>38</sup> (*Référentiel Général d'Accessibilité des Administrations*) qui définit des recommandations similaires à celles des WCAG pour la conception des sites Web ; aux États-Unis, il existe des recommandations techniques pour l'accessibilité définies par le gouvernement dans la *Section 508 Amendment* ; en Allemagne, on retrouve un standard similaire dénommé BITV ; d'autres normes et standards existent pour les données numériques publiques au Japon, en Italie et en Irlande.

Cependant, ces normes ne semblent pas suffisantes pour résoudre les principaux problèmes d'accessibilité rencontrés par les déficients visuels, en témoignent plusieurs études [Commission 2004; Power et al. 2012] et l'apparition de solutions proposant de remanier entièrement les pages web pour assurer leur accessibilité (VoxiWeb). De plus, même si l'on considère que les normes permettant une certaine disponibilité des contenus dans des modes non-visuels, l'utilisabilité des contenus numériques dépend tout de même des technologies d'assistance utilisées pour y accéder. Ainsi l'accessibilité des contenus numériques ne se résume pas qu'à la disponibilité des contenus, mais inclut également l'utilisabilité des technologies d'assistance, qui sont les garants de l'efficacité, de l'efficience et de la satisfaction avec lesquelles se fait l'accès aux contenus numériques.

Enfin, la disponibilité de l'information permise par les normes d'accessibilité ne semble pas être totale. Par exemple, une alternative textuelle ne peut que difficilement être aussi informative qu'une image elle-même, et même pire dans certains cas fournir une interprétation erronée ou au mieux occultant les autres interprétations possibles de l'image [Kelly et al. 2007]. Sur un autre plan, au sens de la théorie de la MFM de nombreuses informations ne sont pas disponibles actuellement dans des contenus numériques respectant les normes du W3C. On peut citer notamment les relations entre objets textuels comme un encadré jouant le rôle d'illustration d'un concept développé dans le texte, ou encore certains statuts particuliers comme le fait qu'une emphase soit utilisée dans le texte pour marquer les mots clés, ou encore le fait que deux paragraphes soient regroupés visuellement sur la page. Plus de détails sur l'état actuel de l'accessibilité des informations de MFM seront donnés dans la partie 6.

---

<sup>38</sup> <https://references.modernisation.gouv.fr/rgaa-3-0>

## **5. Les technologies d'assistance pour l'accès non-visuel aux contenus numériques : approches et limites**

---

---

Dans le domaine de l'informatique, « technologie d'assistance » (ou *AT* pour *Assistive Technologies*) est un terme générique utilisé pour désigner l'ensemble des solutions matérielles et logicielles permettant à des utilisateurs déficients d'accéder aux contenus numériques et de les utiliser. Ces technologies ont émergé de la nécessité de proposer de nouveaux moyens d'interaction en marge des interfaces graphiques et des périphériques d'entrée/sortie habituels, qui ne permettent pas à certains utilisateurs d'accéder aux contenus. Elles jouent un rôle prépondérant dans l'accessibilité du numérique en général étant donné qu'elles gèrent les périphériques d'entrée et de sortie d'information, le système d'exploitation et les fonctions d'accès et d'interaction avec les contenus, entre autres. Nous présenterons les technologies principales utilisées par les personnes déficientes visuelles, les fonctions qu'elles proposent et les différents travaux de recherche autour de leur enrichissement ou leur évaluation, afin de mieux comprendre les approches actuelles de l'accessibilité. Nous nous concentrerons cependant sur l'accès des contenus numériques et plus particulièrement des contenus textuels, dans le but d'évaluer la prise en compte de la mise en forme matérielle par les technologies d'assistance. De plus, nous nous restreindrons aux technologies d'assistance dédiées à un accès non-visuel, laissant de côté les outils d'agrandissement et de modification des couleurs destinés à d'autres pathologies du système visuel que la cécité. Cette dernière restriction rejoint un choix méthodologique de conception pour la restitution de la MFM : nous avons en effet choisi de nous limiter dans un premier temps à un accès uniquement non-visuel, afin de nous affranchir de la complexité liée à la diversité des pathologies visuelles existantes et des stratégies de compensation correspondantes mises en place par les malvoyants.

L'objectif de cette partie est de dresser un panorama des technologies existantes, leurs modes de fonctionnement, leurs utilisations et utilité, afin de comprendre comment les non-voyants accèdent à l'information numérique.

### **5.1. Périphériques et modalités d'entrée et de sortie : les bases de l'accès (non-visuel) au numérique**

---

Classiquement sur un ordinateur, l'écran et les enceintes (ou le casque audio) sont les périphériques physiques de *sortie*, car (du point de vue de l'outil) ils permettent de communiquer de l'information vers l'utilisateur ; le clavier et la souris constituent eux les périphériques *d'entrée* permettant de communiquer de l'information de l'utilisateur vers l'outil. Ces périphériques sont essentiels car ils permettent la communication homme-machine, et donc l'échange d'information.

À la manière de l'écran pour les interfaces graphiques, la **plage braille** et la **synthèse vocale** à travers le périphérique de sortie audio, sont les deux vecteurs d'information principaux pour les non-voyants utilisant un ordinateur ou autre outil informatique. Ce sont les périphériques de *sortie* les plus utilisés dans ce cadre d'interaction. Pour les périphériques d'entrée, le clavier d'ordinateur est majoritairement utilisé avec une variante « braille » ; d'autres périphériques existent mais sont actuellement plutôt limités au cadre de la recherche. Cette section vise à dresser un panorama des périphériques et des modalités d'entrée et de sortie utilisables par les déficients visuels, de leurs

caractéristiques, leur utilisation dans les technologies grand-public actuelles ainsi que des travaux de recherche sur le sujet. Une attention particulière sera portée aux modalités audio étant donné que nous avons choisi d'étudier la restitution de la MFM dans ce mode-ci (voir Chapitre IV).

### **5.1.a. Modalités audio en sortie**

---

Nous étudierons dans cette section trois modalités audio de sortie pour communiquer de l'information aux utilisateurs déficients visuels : la **synthèse vocale**, les **icônes audio** et la **spatialisation du son**, tous les trois utilisables à travers un périphérique tel qu'un casque audio.

La **synthèse vocale**, abrégée en anglais TTS (pour *Text-To-Speech*), est une technique informatique permettant la production de parole à partir de texte. Les premiers travaux portant sur la production de sons parlés remontent à la fin du XVIII<sup>ème</sup> siècle avec la machine à parler de *Von Kempelen*, qui produisait des voyelles à l'aide d'un système mécanique mimant le fonctionnement de l'appareil phonatoire humain. Plusieurs phases d'évolution majeures ont permis le développement de la production de parole, jusqu'à l'introduction d'une composante émotionnelle avec la prosodie qui a permis au début des années 90 de rendre les voix plus expressives (voir [Schröder 2009] pour un historique de l'évolution des technologies de synthèse vocale). Les voix de synthèses actuelles permettent un rendu extrêmement réaliste, restant toutefois limitées aux bases de règles (certes extensives) et aux fonctions de traitement automatique des langues sur lesquelles elles reposent. La synthèse vocale permet donc l'accès aux contenus numériques pour les personnes déficientes visuelles à travers la lecture vocale de texte.

La synthèse vocale présente plusieurs avantages : un coût réduit, un débit de lecture supérieur à la lecture tactile (voir 3.2), et enfin elle ne nécessite aucun apprentissage particulier étant donné qu'elle restitue du langage naturel parlé. Concernant le coût des synthèses vocales, il est bon de préciser qu'il existe de nombreuses voix de synthèse gratuites ou disponibles librement sur internet<sup>39</sup> avec des qualités de rendu variables, mais aussi des voix propriétaires présentant souvent un meilleur rendu, vendues habituellement par lot pour une centaine d'euros. Concernant le débit de lecture en synthèse vocale, des utilisateurs aveugles peuvent aisément comprendre les voix à des débits équivalents à ceux d'une lecture visuelle silencieuse, c'est-à-dire autour de 300 mots par minute [Asakawa and Takagi 2003], là où l'on considère qu'une vitesse moyenne de lecture tactile braille se situe autour des 100 mots par minute [Millar 2003; Legge et al. 1999]. Des utilisateurs moins avancés utiliseront plutôt le débit par défaut des synthèses qui oscille habituellement autour de 150 à 180 mots par minute.

En marge de la parole, la sortie audio d'un ordinateur permet de véhiculer de l'information d'une autre manière à travers le son. On peut citer par exemple les sons associés aux messages d'erreur ou aux notifications couramment utilisés. Ces sons courts, utilisés pour notifier d'un événement ou pour véhiculer une information particulière, existent sous plusieurs formes sous l'appellation **icônes audio** (*earcons*). D'une durée la plupart du temps inférieure à 1.5 secondes, les icônes audio étendent le concept d'iconicité du visuel à l'audio afin d'être porteuses de sens [Blattner et al. 1989]. On distingue deux approches existantes dans la conception de ces sons iconiques. La première est d'utiliser des sons dits « naturels » comme des enregistrements afin qu'ils soient le plus

---

<sup>39</sup> Le projet MBROLA propose de nombreuses voix gratuitement dans une trentaine de langues : <http://tcts.fpms.ac.be/synthesis/mbrola.html>



représentatifs possibles (on les retrouvera plutôt sous le terme *auditory icons* dans la littérature), par exemple un son de froissement de papier pour signaler la suppression d'un fichier. La seconde consiste à utiliser des sons « abstraits » (les *earcons*) dont la signification doit être apprise par les utilisateurs, qui peuvent être partiellement représentatifs afin de faciliter leur apprentissage (ex. une courte mélodie joyeuse associée à la réussite d'une action). En jouant sur les composantes des icônes audio abstraites comme le timbre, la hauteur, le rythme ou encore le volume, on peut définir des familles d'icônes audio pour faciliter leur interprétation [Brewster et al. 1993]. Enfin, en augmentant sensiblement le débit de l'enregistrement d'une phrase énoncée, on peut créer des *spearcons* (*speech earcons*). Leur principal avantage est de se départir de la nécessité de concevoir des sons uniques et de les associer à des concepts abstraits, puisque chaque item à sonoriser peut être associé à une description verbale déjà unique (ex. « Enregistrer sous ») [Walker et al. 2006]. Même si le signifié langagier n'est plus interprétable tel quel, l'apprentissage est facilité par la possibilité d'associer la série d'intonations (en théorie) unique de chaque son à sa contrepartie langagière.

Ainsi, les icônes audio sont un vecteur d'information alternatif à la parole, moyennant apprentissage ; leur principal avantage réside dans une restitution plus succincte que la parole pour une information donnée, ce qui en fait une modalité de restitution pertinente pour des informations aux fréquences d'occurrence élevées. Les principales utilisations des icônes audio en recherche ont été pour les menus audio afin d'identifier les items et leur hiérarchie [Yalla and Walker 2008], également pour l'accès à certains contenus complexes comme les mathématiques où l'on retrouve certains symboles et structures fréquemment [Bates and Fitzpatrick 2010] ou encore en remplacement de certains items récurrents dans un texte [Hussain et al. 2012].

Enfin, une troisième modalité audio utilisable par les déficients visuels est la spatialisation du son. Dans la nature, tout son est émis depuis un certain emplacement, une source localisée dans l'espace. Les recherches en psycho-acoustique ont montré que le système auditif humain était capable de localiser la source spatiale des sons en utilisant des indices tels que le délai interaural de perception du son<sup>40</sup>, la différence interaurale d'intensité du son, le volume, l'écho ainsi que d'autres indices spectraux [Rumsey 2001]. La discrimination des sources sonores dans l'espace est plus ou moins efficace selon la position et la distance à l'auditeur ; on sait notamment que la localisation sur l'axe horizontal est plus précise que sur l'axe vertical [Searle et al. 1976], et que sur l'axe horizontal la discrimination est meilleure quand les sources sonores sont placées en face des sujets (azimut tendant vers zéro) [Grantham and Wilkerson 1986]. Il est possible de simuler la spatialisation du son dans un environnement informatique en reproduisant les fonctions de transfert acoustiques entre la source sonore et les oreilles de l'auditeur, cela incluant les différents indices perceptifs cités ci-avant. L'origine spatiale d'un son peut alors être simulée quand celui-ci est joué dans un casque audio binaural, ce qui s'appelle de la *synthèse binaurale*. À noter que la qualité du rendu dépend directement de la fonction de transfert acoustique utilisée ; en effet, cette dernière est basée sur de nombreux paramètres liés à la morphologie de l'auditeur comme l'écart interaural ou la forme du pavillon de l'oreille. Une fonction de transfert<sup>41</sup> utilisant des paramètres morphologiques moyens peut être utilisée, en risquant de déformer la perception de la localisation du son pour certains

---

<sup>40</sup> Un son provenant de la gauche d'un auditeur sera d'abord perçu par son oreille gauche avant de l'être par son oreille droite ; c'est cette différence temporelle qui est ici utilisée.

<sup>41</sup> Cette fonction de transfert propre à chaque morphologie correspond à ce qui est appelé la *HRTF*, pour *Head-Related Transfer Function*

auditeurs. Une autre possibilité est de mesurer cette fonction de transfert, à travers une batterie de tests utilisant un matériel dédié coûteux durant lesquels le sujet doit rester immobile.

L'intérêt du son spatialisé est de pouvoir rajouter une information supplémentaire (de nature spatiale donc), au son qui est joué. Dans le cadre d'interfaces non-visuelles, cette information spatiale peut être utilisée en tant que telle pour retranscrire la disposition (spatiale) d'informations visuelles par correspondance directe entre les sources sonores dans l'espace et les informations à restituer. Par exemple, [Cofino et al. 2014] utilisent des sources sonores spatialisées pour indiquer la localisation d'informations dans un tableau (ex. la lecture de la colonne de gauche se fait depuis une source sonore située sur la gauche de l'auditeur) ; on peut également citer [Murphy et al. 2010] qui spatialisent la lecture d'expressions mathématiques sur un axe gauche-droite pour indiquer la position de lecture dans cette dernière. Il est également possible, comme pour les icônes audio, de définir de manière arbitraire une correspondance entre une source sonore et un type d'information ; ainsi [Parente 2008] a développé un environnement audio dans lequel certains types d'informations étaient lus depuis des sources spatiales dédiées (ex. les informations pertinentes étaient lu depuis une source sonore localisée sur la gauche de l'auditeur, tandis que le contenu non pertinent était lu depuis une source localisée sur la droite de l'auditeur). Les résultats de ces études ont montré que son spatialisé pouvait être utilisé avec succès lorsque la finesse de la localisation spatiale demandée aux utilisateurs était raisonnable. Ainsi, la spatialisation du son et de l'audio en général, incluant les son non-parlés et les voix synthétiques, semble être une dimension expressive intéressante pour les interfaces non-visuelles, que nous avons d'ailleurs utilisé pour restituer certaines propriétés de la MFM.

En conclusion, en ce qui concerne les technologies d'assistance actuelle, la synthèse vocale reste la modalité la plus utilisée des trois présentées. Même si certaines icônes audio sont incluses, et même si leur utilisation est configurable dans le cas d'un lecteur d'écran par exemple, leur utilisation reste assez marginale pour le grand public au même titre que pour les interfaces graphiques. Par exemple en prenant le cas d'un système d'exploitation Windows récent, on ne retrouve que très peu d'icônes audio utilisées par défaut. Enfin, l'utilisation du son spatialisé semble être cantonnée à la recherche pour le moment, bien que certaines utilisations dans le cadre du divertissement multimédia aient vu le jour.

### **5.1.b. Terminaux braille**

---

Les *terminaux braille*, aussi communément appelés *afficheurs braille* ou *plages braille*, sont des matrices de picots piézoélectriques (i.e. déformables sous l'action d'un champ électrique) permettant de former des caractères braille en plaçant chaque picot en position haute ou basse par rapport à une grille. Les terminaux braille vendus dans le commerce forment une ligne de 40 ou 80 caractères pour les modèles les plus répandus. Ces terminaux présentent le plus souvent des touches sur les côtés permettant de naviguer entre les lignes du texte retranscrit, et d'autres touches au-dessus de chaque caractère braille permettant de désigner le caractère (d'y placer le curseur virtuel) dans le cadre d'une sélection de texte par exemple. Une fois connecté à un ordinateur, ces terminaux permettent donc « d'afficher » une ligne de texte en braille, et plus généralement ils permettent un accès braille aux contenus et applications numériques. Leur utilisation conjointe avec un ordinateur est relativement répandue, et ils sont intégrés dans la plupart des bloc-notes braille (voir 5.3). Ils constituent, avec la synthèse vocale, les deux principales modalités de sortie d'information dans le

cadre de l'interaction des non-voyants avec un ordinateur. À noter que l'affichage sur certaines plages braille récentes peut être commandé via un smartphone [Hong 2012].



**Figure 12 - Terminal braille** (*image libre de droits*)

L'utilisation de ces dispositifs est contrainte par l'apprentissage du braille, assez peu répandu dans la population globale des déficients visuels (voir 3.2 pour plus de détails). Les principaux inconvénients des terminaux braille résident dans leur coût (aux alentours des 2000USD pour 40 caractères et le double pour 80 caractères<sup>42</sup>), et dans la résolution tactile limitée qu'ils proposent. Cette faible résolution empêche par exemple la retranscription d'images ou l'exploration de la globalité d'une page de texte (ex. nombre de paragraphes, taille, etc.). Le coût des matrices de picots est d'ailleurs une des causes de la taille limitée des afficheurs, la complexité de la gestion d'un « grand » affichage tactile transposant un contenu visuel en est probablement une autre ; il faut par exemple gérer la densité des lignes qui doit être moindre en tactile pour permettre leur discrimination (3mm minimum entre deux traits [Bou-Issa 2010]) et donc transformer, simplifier les contenus visuels pour en tenir compte.

Le coût des terminaux braille est lié à celui des actuateurs piézoélectriques ainsi qu'à leur procédé d'assemblage principalement manuel. Plusieurs travaux de recherche ont été menés pour trouver des alternatives plus abordables. Il y a une vingtaine d'années, [Fricke and Baehring 1993] avaient eu l'idée d'une actuation des picots contrôlée à l'aide d'un circuit hydraulique contenant un fluide spécial à particules conductrices<sup>43</sup> ; plus récemment, [Wilhelm et al. 2014] ont utilisé des actuateurs à membrane moins coûteux que les piézoélectriques. Les avancées de la recherche ont ainsi permis de diminuer le coût des terminaux braille environ de moitié depuis le début des années 2000. De plus, certaines aides existent pour financer l'achat de ce type de dispositifs d'assistance ; en France on peut citer l'AGEFIPH (Association de Gestion du Fonds pour l'Insertion Professionnelle des Personnes Handicapées), qui utilise les fonds reversés par les entreprises privées ne respectant pas les quotas instaurés de travailleurs handicapés<sup>44</sup>.

Les limites de taille des afficheurs braille font écho à certains prototypes de recherche. Des matrices allant de quelques centaines à plusieurs milliers de picots ont ainsi été développées [Shinohara et al. 1998; Prescher et al. 2010], celui de Shinohara *et al* ayant la particularité de pouvoir placer les picots à plusieurs hauteurs pour ainsi former des volumes. À noter que certains dispositifs présentant des matrices de plus de 10000 picots sont disponibles à la vente<sup>45</sup>, mais leur coût prohibitif, les problèmes de compatibilité et la rareté des applications logicielles utilisables en font des dispositifs assez marginaux. Une autre approche est de placer un nombre limité de cellules braille sur un objet

---

<sup>42</sup> <http://www.perkinsproducts.org/store/en/39-braille-displays>

<sup>43</sup> Fluide électrorhéologique

<sup>44</sup> Loi n° 87-517 du 10 juillet 1987 en faveur de l'emploi des travailleurs handicapés

<sup>45</sup> Exemple : <http://www.tactisplay.com/shops>

mobile (comme une souris), et de laisser l'utilisateur explorer une surface avec cet objet, sur laquelle est virtuellement projetée un texte ou un graphique, et en produisant un affichage sur les cellules correspondant à la zone sous laquelle l'objet est placé. Le dispositif *Tactos* utilise ce principe qui a fait ses preuves pour la reconnaissance de formes, notamment des caractères latins [Hanneton et al. 1998], puis plus tard des formes géométriques simples [Gapenne et al. 2003]. Le principe a été repris et implémenté sur une *trackball* et un stylo par [Lecolinet and Mouret 2005], et sur une souris d'ordinateur [Headley et al. 2011].

Enfin, une dernière approche pour pallier les problèmes de résolution limitée du braille est de définir arbitrairement des patterns de picots permettant ainsi de représenter iconiquement un concept. Dans sa thèse, [Pietrzak 2008] a notamment montré que des patterns de picots dynamiques (avec « clignotement » des picots) pouvaient être utilisés pour reconnaître des directions avec des taux d'erreurs très faibles (environs de 5% pour les meilleurs jeux de patterns).

### **5.1.c. Dispositifs haptiques particuliers**

---

Le terme « haptique » (du grec *haptomai* qui signifie *je touche*), désigne tout ce qui est relatif au sens du toucher c'est-à-dire au sens tactile tel qu'on l'a ici défini (incluant le sens *cutané* et le sens *kinesthésique*). Les plages braille peuvent donc être qualifiées de dispositifs *haptiques*. Cependant, on introduira dans cette section deux dispositifs particuliers, relativement répandus dans la littérature tous deux sous l'appellation *haptique* : les dispositifs vibratoires (utilisant le sens cutané) et les dispositifs à retour d'effort, utilisant eux le sens kinesthésique.

Les dispositifs vibratoires fonctionnent généralement soit en émettant du son à la manière d'un haut-parleur, soit en utilisant un arbre à came décentré monté sur l'axe d'un moteur. Ces dispositifs sont assez communs et on les retrouve notamment dans les téléphones portables et les manettes de jeux-vidéos. Les vibrations sont habituellement associées à certains événements sans constituer de messages informatifs à elles seules ; [Brewster and Brown 2004] ont cependant défini des patterns de vibrations variant en rugosité et en rythme, avec des résultats prometteurs montrant des taux de reconnaissance entre 70 et 90% [Brown et al. 2005]. En associant des patterns vibratoires à certains niveaux de gris, [Maurel et al. 2012] ont montré qu'il était possible d'explorer un document structuré (dont l'image est convertie en niveaux de gris) sur une tablette à l'aide d'un gant comportant un moteur vibrant sous l'index. Enfin, la combinaison de plusieurs moteurs vibrants permet de définir des patterns plus complexes : [Tan et al. 2003], à l'aide d'une matrice de 9 moteurs montés sur le dos d'une chaise, montrent qu'il est possible de reconnaître des directions avec un taux d'erreur d'environ 10%.

D'autres dispositifs haptiques utilisés en recherche sont les dispositifs à retour d'effort. Ils sont fortement représentés par les bras articulés à retour d'effort, comme le PHANTOM Omni de la société Sensable<sup>46</sup>. Le principe est de monter un stylet ou d'attacher la main de l'utilisateur sur un bras articulé à plusieurs degrés de liberté relié à un moteur, afin que l'utilisateur puisse percevoir des stimuli haptiques lors de l'exploration de l'espace atteignable par le bras, à travers la résistance et les vibrations produites par le moteur. Ce type de dispositif a notamment été utilisé pour l'exploration de graphes et plus particulièrement de courbes dans un repère cartésien [Yu et al. 2001]. A noter que ce type de périphérique est à la fois un périphérique d'entrée et de sortie, car il permet de donner de

---

<sup>46</sup> <http://www.sensable.com/>

l'information à l'utilisateur à travers le retour de force, mais utilise également les entrées utilisateurs à travers le déplacement du bras articulé.



Figure 13 - Le PHANTOM Omni de Sensable

D'autres dispositifs haptiques plus marginaux existent ; [Stone 2001] en citent un certain nombre dans l'historique qu'ils ont dressé, avec par exemple un gant produisant des sensations de pression sur la main à l'aide de zones gonflables ou encore des exosquelettes à retour d'effort. Pour terminer on peut citer [Levesque and Hayward 2003] qui ont développé et testé un prototype compressant la peau des doigts posés sur le dispositif, afin de simuler la sensation tactile de creux et de bosses.

#### ***5.1.d. Périphériques d'entrée***

---

Le périphérique d'entrée le plus couramment utilisé par les déficients visuels pour interagir avec un ordinateur est le clavier. Certains repères en reliefs classiquement utilisés pour la dactylographie sont utilisés pour connaître la position des touches, en plus de la forme générale du clavier. Un autre type de clavier utilisé par les non-voyants est le clavier braille (voir Figure 14), qu'on retrouve notamment sur les blocs-notes braille (voir 5.3). Ce clavier fonctionne sur un principe « d'accords », comme en musique, où chaque caractère braille correspond à une combinaison de touches unique. Sur l'image ci-dessous on peut voir 3 touches pour chaque main, pour les 6 points d'un caractère braille informatique. Ainsi, pour former les caractères braille par « accords », chaque touche de la main gauche est associée à chacun des 3 points de la partie gauche d'un caractère, et de la même façon les 3 touches de la main droite sont associées aux 3 points de la colonne de droite du caractère. Ce clavier était déjà utilisé historiquement sur les emboisseuses braille<sup>47</sup> qui permettent de mettre une feuille de papier en relief pour y inscrire un texte braille ; les touches aux extrémités droite et gauche sont apparues avec l'évolution numérique du clavier et sont habituellement associées au retour à la ligne et au retour arrière (*Backspace*) pour le traitement de texte. À noter que les claviers braille ont été très peu étudiés dans la littérature ; une étude par [Lee et al. 2002] montre cependant que ces claviers observent des vitesses de frappe inférieures à celles des claviers QWERTY classiques, bien que cette étude n'inclue pas d'utilisateurs experts de ce type de clavier. Ces résultats sont d'ailleurs étonnants si l'on considère le fait que les claviers braille laissent les mains statiques pendant la frappe, ce que certaines dispositions de claviers classiques d'ordinateur comme le BEPO<sup>48</sup> ou le DVORAK<sup>49</sup> cherchent justement à optimiser pour augmenter la vitesse de frappe et maximiser le confort. Quoi qu'il en soit, le clavier braille permet la saisie de texte mais nécessite pour

---

<sup>47</sup> Les plus populaires sont les emboisseuses *Perkins* : <http://www.perkinsproducts.org/about-perkins-brailleurs>

<sup>48</sup> Disposition de clavier optimisée pour la frappe de texte en Français, dont le nom provient de la disposition de la première ligne de touches alphabétiques

<sup>49</sup> Disposition de clavier optimisée pour la frappe de texte en Anglais, du nom de son inventeur August Dvorak

interagir avec le système d'exploitation de définir de nouvelles combinaisons de touches en plus de celles utilisées pour les caractères et/ou l'ajout de touches supplémentaires.



Figure 14 - Clavier QWERTY (gauche) ; clavier braille et caractère braille (droite) (images libres de droits)

Une autre modalité d'entrée répandue est le toucher, qui n'est cependant utilisée que sur les *smartphones* et tablettes tactiles. Le contrôle de l'appareil se fait alors à travers un ensemble de mouvements et de contacts d'un ou plusieurs doigts en contact avec l'écran. Les non-voyants semblent plébisciter les *smartphones* d'*Apple* utilisant le système d'exploitation *iOS*<sup>50</sup>. Ses utilisateurs ont la possibilité d'utiliser la commande vocale pour interagir avec leur périphérique.

La reconnaissance vocale, autre modalité d'entrée utilisable, a fait des progrès immenses avec le développement des réseaux de neurones et plus récemment avec l'avènement du *deep learning*, qui est une technique d'apprentissage automatique permettant la modélisation d'ensemble complexes de données. *Google* a d'ailleurs annoncé que le taux d'erreur de la reconnaissance vocale proposée sur les appareils *Android* a chuté à 8%<sup>51</sup>. Cependant dans le cadre des interactions non-visuelles, un taux d'erreur même faible aura probablement un impact négatif sur le facteur « temps » des interactions qui semble déjà problématique dans ce cadre (voir 3.2), en obligeant les utilisateurs à vérifier que leurs requêtes ont bien été comprises, et le cas échéant à les reformuler. De plus, utiliser la reconnaissance vocale pour interagir avec un périphérique informatique peut être problématique au regard de la confidentialité des interactions. Le même problème a déjà été soulevé pour les synthèses vocales, qui, utilisées sans casque audio permettent à d'autres personnes à proximité de « jeter un œil » sur ce que l'utilisateur est en train de faire [Ahmed 2015]. Ce dernier auteur pointe notamment du doigt le fait que les non-voyants rencontrent des problèmes de confidentialité au quotidien, par exemple lorsqu'ils ont besoin d'une aide extérieure pour lire un courrier ou déboguer une machine si l'utilisateur est novice. La reconnaissance vocale a cependant le mérite de ne nécessiter aucun dispositif externe et d'utiliser du langage naturel pour l'interaction.

D'autres modalités d'entrée existent, comme les gestes des mains<sup>52</sup> ou du corps, aucune n'ayant été à notre connaissance déjà utilisée dans le cadre d'interactions non-visuelles.

## 5.2. Les lecteurs d'écran

---

### 5.2.a. Présentation générale des outils existants

---

Les lecteurs d'écran sont des applications logicielles permettant d'interagir avec le système d'exploitation du périphérique, avec ses applications et permet aussi de prendre connaissance des

---

<sup>50</sup> La section suivante détaillera ce constat. Pour un aperçu de VoiceOver : <http://www.apple.com/accessibility/ios/>

<sup>51</sup> Sources : [venturebeat.com](http://venturebeat.com) et [Speech Technology Magazine](http://SpeechTechnologyMagazine)

<sup>52</sup> *LeapMotion* commercialise un dispositif populaire d'entrées gestuelles : <https://www.leapmotion.com/>

contenus ; leur nom reflète leur usage qui est de permettre de redonner un accès à toute information visuelle affichée à l'écran. Les lecteurs d'écran représentent donc une interface entre les périphériques, avec leurs interfaces graphiques natives, et les technologies d'assistance comme la synthèse vocale et la plage braille, permettant ainsi aux déficients visuels d'utiliser les périphériques informatiques sur lesquels ils sont installés. Les lecteurs d'écran doivent ainsi gérer toute la complexité des interactions et des contenus visuels initialement implémentés et présentés par les applications et le système d'exploitation (abrégé *SE*), pour les rendre accessibles dans des modalités non-visuelles. Ils doivent ainsi faire face à des problèmes de compatibilité avec le système d'exploitation et les applications<sup>53</sup>, des problèmes d'accessibilité du contenu (contenus dynamiques, graphiques, cryptés, etc.) et doivent également gérer les contraintes des modalités non-visuelles comme leur séquentialité (voir 3.2).

Nous verrons dans cette section quels sont les principaux lecteurs d'écran existants, leur fonctionnement en ce qui concerne l'accès aux documents textuels et nous dresserons un bilan des études portant sur leur usage. Dans cette section on fera plusieurs fois référence à un sondage récent du groupe WebAIM, un organisme privé œuvrant dans le domaine de l'accessibilité Web. Dans ce sondage, daté de Janvier 2014, 1465 utilisateurs de lecteurs d'écrans étaient interrogés sur leurs pratiques ([disponible ici](#)).

Les 3 lecteurs d'écrans les plus utilisés sont JAWS, NVDA et VoiceOver utilisés respectivement par 50%, 19% et 10% des sondés. D'autres lecteurs d'écran existent notamment Window-Eyes de *GW Micro* tout de même utilisé par 14% des sondés mais dont l'utilisation est en déclin ; on peut également citer Orca, lecteur d'écran gratuit et open source fonctionnant sur les systèmes Unix. JAWS (*Job Access With Speech*) est un lecteur d'écran commercialisé par *Freedom Scientific* depuis 1995 dont la licence pour particuliers est vendue environ 1 000 USD ; JAWS fonctionne exclusivement sous Windows. Malgré son coût, il reste le lecteur d'écran le plus majoritairement utilisé par les déficients visuels, peut-être à cause de l'avantage de son ancienneté<sup>54</sup>. NVDA et VoiceOver ont eux vu le jour respectivement en 2005 et 2006. NVDA (*Non-Visual Access Desktop*) est un lecteur d'écran fonctionnant sous Windows dont le principal avantage est d'être gratuit et open source. Enfin, VoiceOver est le lecteur d'écran développé par *Apple*, et se trouve être inclus dans les différents systèmes d'exploitation de la marque depuis quelques années (OSX pour ordinateurs et iOS pour dispositifs mobiles). L'utilisation de dispositifs mobiles par les déficients visuels a d'ailleurs explosé depuis 2009 ; 82% des sondés par WebAIM disent utiliser un dispositif mobile. Les dispositifs mobiles Apple fonctionnant avec VoiceOver semblent d'ailleurs plébiscités par les déficients visuels (utilisés par 65% des sondés sur dispositifs mobiles), loin devant le lecteur d'écran *TalkBack* d'Android (21%). Sur dispositifs mobiles ou sur ordinateur, le principe de fonctionnement des lecteurs d'écran reste inchangé, la différence principale étant que ces derniers sont contrôlés à l'aide du clavier sur ordinateur, et via un ensemble de gestes en contact avec l'écran sur les dispositifs mobiles. [WebAIM](#) recommande d'ailleurs aux développeurs de se conformer aux standards d'accessibilité en vigueur

---

<sup>53</sup> Chaque application peut potentiellement encoder différemment l'information, utiliser des normes particulières de communication des données, protéger son code interne, etc.

<sup>54</sup> Ce n'est qu'une hypothèse mais la résistance au changement, liée au fait que les lecteurs d'écrans sont des systèmes complexes nécessitant un apprentissage long incluant de multiples raccourcis clavier, reste une explication plausible au maintien de l'utilisation de JAWS. Cette hypothèse est supportée par les travaux de [McCarthy et al. 2013] montrant que les utilisateurs de lecteurs d'écran changent très rarement de logiciel bien qu'ils soient très similaires.

## 5. Les technologies d'assistance pour l'accès non-visuel aux contenus numériques : approches et limites

sans se préoccuper des différences entre lecteurs d'écran, qui semblent présenter une certaine homogénéité dans leur fonctionnement.

Les lecteurs d'écran visent donc à retransmettre et retranscrire l'information visuelle dans des modalités non-visuelles avec notamment la parole, à travers la synthèse vocale, mais aussi avec le braille quand une page braille est disponible. Pour comprendre le fonctionnement général des lecteurs d'écran, nous présenterons les principales catégories de commandes proposées par JAWS [Freedom Scientific 2014].

Catégorie de commandes	Exemples de sous-catégories	Exemples de commandes	Raccourcis clavier
<i>Système d'exploitation</i>	Explorateur de fichiers	Remonter d'un niveau dans l'arbre hiérarchique	<i>Backspace</i>
	Lecture de texte	Paragraphe suivant	<i>CTRL + SHIFT + Arrow up/down</i>
	Boîtes de dialogue	Aller au contrôle suivant de la boîte de dialogue	<i>TAB</i>
<i>Navigateur : Firefox ou Internet Explorer</i>	Navigation entre onglets	Fermer l'onglet courant	<i>CTRL + W</i>
	Fonctions spécifiques	Ouvrir le gestionnaire de téléchargements	<i>CTRL + J</i>
<i>Pages web et HTML</i>	Commandes générales	Aller au prochain élément cliquable	<i>SLASH</i>
	Titres	Lister les titres de la page	<i>INSERT + F6</i>
	Formulaires	Aller au premier élément de formulaire sur la page	<i>INSERT + CTRL + HOME</i>
	Navigation	Recherche du texte dans la page	<i>CTRL + F</i>
	Tableau	Aller à la ligne suivante du tableau	<i>WINDOWS KEY + ALT + DOWN ARROW</i>
<i>Commandes standard</i>	Commandes de lecture	Lire le mot précédent	<i>INSERT + LEFT ARROW</i>
	Commandes contextuelles	Lire le titre de la fenêtre active	<i>INSERT + T</i>
	Gestion du curseur virtuel et de la souris	Placer la souris sur le curseur virtuel de JAWS	<i>INSERT + NUM_PAD MINUS</i>
	Boîtes de dialogue	Lire le bouton par défaut de la boîte de dialogue	<i>INSERT + E</i>
	Gestion de la synthèse vocale	Augmenter le débit de la synthèse vocale	<i>PAGE_UP</i> (pendant la lecture)
<i>Fonctions spécifiques à JAWS</i>	Lecture	<i>Skim Reading</i> (lecture "en diagonale") ; lit le début de chacun des paragraphes	<i>CAPS LOCK + CTRL + DOWN ARROW</i>
	Numérisation de texte	Lancer l'OCR ( <i>Optical Character Recognition</i> ) sur tout l'écran	<i>CAPS LOCK + SPACEBAR</i> suivi de <i>O</i> puis <i>S</i>
	Aide	Lancer l'aide contextuelle	<i>CAPS LOCK + F1</i>

**Tableau 1 – Extrait des commandes listées dans le manuel de JAWS**

On remarquera qu'afin de ne pas interagir avec les raccourcis et actions déjà définis au sein des applications et du SE, la plupart des raccourcis sont basés sur les touches CAPS\_LOCK (verrouillage majuscule) et la touche INSERT qui sont utilisées pour le traitement de texte et n'ont jamais d'action directe dans les applications. À noter que les deux premières catégories de commande du tableau se réfèrent à des commandes déjà implémentées respectivement par Windows et les navigateurs web, et ne sont donc pas spécifiques à JAWS ; cependant, elles ont été incluses dans le manuel proposé par *Freedom Scientific* pour leur importance quant à l'utilisation de l'ordinateur en général. Les raccourcis clavier permettent donc virtuellement de réaliser toutes les actions nécessaires pour interagir avec les contenus numériques : gestion des applications, de la lecture, du système d'exploitation, de la navigation dans les contenus, etc. L'ensemble des commandes donnent lieu à un feedback vocal ou braille suivant la configuration.



### ***5.2.b. Accès aux contenus numériques par les lecteurs d'écran***

---

Étant donné que nos travaux s'inscrivent dans le cadre de l'accessibilité des documents numériques, nous allons examiner quelles sont les méthodes et moyens d'accès aux contenus proposés par JAWS et les lecteurs d'écran en général. Nous avons regroupé les principales fonctions associées ci-dessous ; à noter que, sauf exception, ces fonctions sont utilisables pour tous les formats de documents présentant un marquage adéquat dans leur code interne :

- § Lecture exhaustive : les lecteurs d'écran permettent d'effectuer une lecture exhaustive des documents (fonction *read all*), avec un ordre de lecture défini par défaut (de haut en bas et de gauche à droite). La lecture peut aussi être réalisée ligne par ligne (via la touche TABULATION), ou d'objet en objet pour ceux ayant un indice de tabulation défini.
- § Accès par type d'objets : cette catégorie regroupe la majorité des commandes utilisées pour la navigation au sein d'un document structuré. On y retrouve un ensemble de commandes permettant le déplacement du curseur virtuel sur le premier, le dernier, le précédent ou le prochain objet d'un certain type. Les principaux types d'objets utilisés pour la navigation dans ce cadre sont : les titres, les liens hypertexte, les paragraphes, les tableaux, les éléments cliquables, les formulaires (cela incluant les zones de saisie de texte, les listes déroulantes, les listes à cocher, les boutons, etc.) et les listes (à puces ou numérotées). Ainsi, on retrouvera par exemple une commande pour aller au premier champ du formulaire de la page ou encore une autre pour aller au prochain titre. Les commandes permettant de naviguer entre les titres représentent d'ailleurs la première méthode utilisée pour trouver de l'information dans une page<sup>55</sup>. Il est à noter que l'efficacité de l'utilisation de ces commandes est totalement dépendante du marquage du contenu ; un objet marqué en tant que titre dans le code HTML d'une page peut ne pas représenter un titre pour le contenu de la page. On en revient encore une fois à l'importance du respect des normes d'accessibilité pendant la conception des documents. De plus, les auteurs de contenu sont libres dans l'utilisation des différents types d'objets utilisés pour la navigation (ex. naviguer par titre ou par lien) et peuvent l'utiliser dans une logique de présentation et non pas selon la sémantique définie pour chaque type, et ainsi le type des objets ne suffit pas à inférer leur rôle dans le contexte de la page ; ce problème est partiellement résolu par les normes ARIA d'accessibilité qui proposent un marquage sémantique du contenu (voir 4.3.b).
- § Accès par région : début 2014 ont été créées les normes *ARIA (Accessible Rich Internet Applications)* par le *W3C*, qui permettent de typer les zones de contenus des pages web (ex. zone principale de contenu, zone périphérique, zone de navigation, etc.). Ainsi il est possible pour les utilisateurs de lister les différentes zones définies sur la page et naviguer entre ces zones pour accéder au contenu qu'elles contiennent [Freedom Scientific 2013]. Bien que les normes ARIA soient récentes, la navigation par zones est relativement utilisée par les déficients visuels, utilisation qui est d'ailleurs en augmentation<sup>56</sup>.
- § Accès par signets : Les utilisateurs peuvent insérer des signets (ou *place-markers*) dans les pages web afin de s'en servir de repères de navigation. Il est ainsi possible de naviguer entre les signets définis dans une page donnée. Les signets sont conservés même après la fermeture des pages.

---

<sup>55</sup> Voir <http://webaim.org/projects/screenreadersurvey5/#finding>

<sup>56</sup> Voir <http://webaim.org/projects/screenreadersurvey5/#landmarks>

- § **Autres** : La fonction de recherche de texte dans une page, qui n'est d'ailleurs pas spécifique aux lecteurs d'écran, est utilisée par nombre d'utilisateur pour trouver de l'information sur le web<sup>57</sup>. Enfin, JAWS implémente une fonction de « lecture en diagonale » (*skim-reading*), lisant « la première ligne ou phrase de chaque paragraphe du document »<sup>58</sup>.

Il est intéressant de noter que d'après plusieurs études [Yesilada et al. 2007; Murphy, Kuber, McAllister, et al. 2007], plus les déficients visuels gagnent en expertise d'utilisation du lecteur d'écran, moins ils explorent les pages linéairement et plus ils utilisent les fonctions d'accès « transverses » telles que l'accès par type d'objet ou la fonction de recherche de texte. De plus, d'après leurs résultats la navigation des utilisateurs experts est plus basée sur la structure des documents dans le sens où ils arrivent à la reconstituer à partir des informations récoltées pendant leur navigation « transverse ». Cela rejoint une problématique commune des modes non-visuels et de la compréhension de contenus qui est que la séquentialité de ces modes d'accès à l'information, et l'absence d'un référentiel global, rend plus difficile la construction d'une représentation globale. Nous pensons que l'étude de la MFM permettra une formalisation plus exhaustive et une restitution plus pertinente de cette structure globale dans des modes non-visuels.

On remarquera de par les principales catégories de fonction présentées que les lecteurs d'écran prennent assez peu en compte la mise en forme du document. Le typage des objets des documents repose principalement sur les types d'objets HTML (titre, lien, liste), alors que nombre d'objets différents existent dont les sous-titres, les définitions, les notes marginales, et les citations pour en citer quelques-uns. De la même façon, les relations de mise en forme ne sont pas prises en compte, la navigation étant basée sur les types d'objets HTML mais pas sur les relations de mise en forme ou sémantiques qu'ils entretiennent ; par exemple un renvoi en marge va être en relation avec le paragraphe en vis-à-vis. ARIA introduit cependant un premier typage sémantique et quelques relations entre objets (utilisées principalement pour labelliser les formulaires par JAWS [Freedom Scientific 2013]) ; nous verrons dans le Chapitre III les limites d'ARIA en ce qui concerne la formalisation de la mise en forme. À noter également que JAWS propose un langage de script à ses utilisateurs, qui leur permet de programmer des comportements de navigation et de lecture personnalisés ; le développement de scripts n'est cependant accessible qu'à une minorité d'utilisateurs experts dans le fonctionnement du lecteur d'écran et ayant des notions de programmation. Les lecteurs d'écran offrent cependant une configuration assez poussée : quels contenus lire, comment, quel contenu lire avec la plage braille ou oraliser via la synthèse vocale, la verbosité de la synthèse vocale, etc.<sup>59</sup>. Par exemple un utilisateur peut configurer JAWS pour signaler certains éléments d'une page avec voix différente ou encore certains « sons » (icônes audio)<sup>60</sup>.

### **5.2.c. Accessibilité et utilisabilité des lecteurs d'écran**

---

Le principal moyen d'interagir avec un lecteur d'écran, et donc avec l'ordinateur pour les déficients visuels, est d'apprendre les raccourcis clavier pour pouvoir utiliser les commandes associées. Le manuel des commandes de JAWS contient plus de 300 références de raccourcis clavier associés à diverses fonctions, dont de nombreux raccourcis utilisant 3 touches et des appuis particuliers (ex.

---

<sup>57</sup> <http://webaim.org/projects/screenreadersurvey5/#finding>

<sup>58</sup> Voir [http://www.freedomscientific.com/Training/Surfs-Up/Skim\\_Reading.htm](http://www.freedomscientific.com/Training/Surfs-Up/Skim_Reading.htm)

<sup>59</sup> Voir [http://www.freedomscientific.com/Training/Surfs-Up/Personalized\\_Settings.htm](http://www.freedomscientific.com/Training/Surfs-Up/Personalized_Settings.htm)

<sup>60</sup> Voir : [http://www.freedomscientific.com/Training/Surfs-Up/Speech\\_Sounds\\_Schemes.htm](http://www.freedomscientific.com/Training/Surfs-Up/Speech_Sounds_Schemes.htm)

CTRL+ALT + 2 appuis rapides sur ESPACE). Malgré cette liste extensive de raccourcis et de la même façon que pour une interface graphique, il n'est pas nécessaire de connaître l'ensemble des commandes pour utiliser un lecteur d'écran : là où les menus visuels permettent l'accès aux fonctions logicielles sans connaître aucun raccourci pour les interfaces graphiques, connaître les raccourcis donnant accès au menu d'aide ou aux menus contextuels peut être suffisant pour une utilisation basique des lecteurs d'écran.

Cependant, il a déjà été démontré que l'utilisation de raccourcis clavier dans les interfaces graphique amène une plus grande efficacité d'utilisation [Lane et al. 2005] ; dans le cas d'interfaces non-visuelles, les contraintes des modalités pénalisent d'autant plus l'efficacité, ce qui rend l'apprentissage des raccourcis clavier d'autant plus important. La dernière étude citée montre d'ailleurs que même des utilisateurs experts (de MS-Word ici) apprennent peu de raccourcis clavier pourtant plus efficaces ; ces auteurs expliquent cet effet par l'habitude des méthodes graphiques apprises en premier, ou simplement par le fait que les utilisateurs ne souhaitent pas y consacrer du temps et un effort d'apprentissage en pensant (à tort) que cet effort ne sera pas utile ou trop important. On peut donc supposer que si ces résultats sont transposables à des utilisateurs déficients visuels, ces derniers vont donc apprendre un nombre « minimal » de commandes, pénalisant d'autant plus leur efficacité que l'accès séquentiel imposé par les modes non-visuels est moins rapide qu'un accès visuel. Le « nombre minimal » de commandes à apprendre pour pouvoir utiliser un lecteur d'écran semble toutefois relativement élevé, en prenant pour exemple [le guide de WebAIM](#) des commandes les plus importantes pour naviguer sur internet avec JAWS qui en comprend plus de 80. Une étude auprès de 11 utilisateurs de lecteurs d'écran [Giraud 2014](p.121) a recensé 30 raccourcis différents utilisés pour des tâches de navigation web « quotidiennes », dans le cas où les participants devaient « apprendre à naviguer à un apprenti, c'est-à-dire à une personne novice qui ne sait pas naviguer sur le Web ». Quoiqu'il en soit, l'étude de l'apprentissage et de l'utilisation des raccourcis clavier par les utilisateurs de lecteurs d'écrans semble particulièrement intéressante afin justement d'évaluer et d'optimiser l'utilisabilité de ces interfaces ; pourtant à notre connaissance rien de tel n'existe dans la littérature.

Il est assez frappant de ne retrouver dans la littérature aucune étude évaluant l'utilisabilité ou la facilité d'apprentissage des lecteurs d'écran. Les études impliquant les utilisateurs de lecteurs d'écran évaluent leurs frustrations et les problèmes qu'ils rencontrent [Lazar et al. 2007; Leporini and Paternò 2004; Power et al. 2012], les comportements et tactiques qu'ils emploient pour faire face à certains problèmes ou situations [Vigo and Harper 2013; Borodin et al. 2010], proposent des nouvelles techniques pour améliorer la navigation non-visuelle [Rohani Ghahari et al. 2012; Vigo et al. 2009; Yesilada et al. 2007], évaluent l'accessibilité du contenu en lui-même [Bowman 2002] ou encore testent de nouvelles méthodes d'évaluation de l'utilisabilité en psychologie cognitive [Federici et al. 2010]. De plus, les trois études citées visant à recenser les problèmes et frustrations rencontrés par les utilisateurs de lecteurs d'écran mettent en avant seulement deux grandes catégories de problèmes : ceux liés aux contenus et à leur conception, et les problèmes techniques de compatibilité et d'erreurs/bugs des technologies. Ainsi, les lecteurs d'écrans sont principalement remis en question sur le volet technique de la compatibilité et des bugs ; les études ne mettent pas en perspective les problèmes d'accessibilité aux contenus relevés (ex. complexité de la structure de la page) avec les fonctions proposées par les lecteurs d'écran pour accéder à ces contenus.

## 5. Les technologies d'assistance pour l'accès non-visuel aux contenus numériques : approches et limites

Étude	Ensemble des problèmes et frustrations relevés	
	Liés aux contenus	Liés aux lecteurs d'écran / aux modalités et médias de présentation
[Lazar et al. 2007]	Alternatives textuelles, liens hypertextes, formulaires mal conçus, structure de la page confuse, formats inaccessibles, fichiers de contenu trop lents à charger.	Erreurs, crashes, conflits avec l'application ou le contenu.
[Leporini and Paternò 2004]	Liens hypertexte, frames, boutons et champs de formulaires, menus de navigation, titres et structure du contenu.	<i>Néant.</i>
[Power et al. 2012]	Contenu mal structuré, problèmes de formats de fichier, fichiers trop lents à charger, contenus non pertinents avec la tâche, contenu introuvable, liens hypertextes obsolètes	Feedback insuffisant pour connaître l'effet des actions, fonctionnalité attendue absente, fonctionnalités ne se comportant pas comme attendu. <i>À noter que ces problèmes sont, d'après les auteurs, partiellement résolus par la modification du contenu utilisé.</i>
[Bowman 2002]	<i>Définis en termes de possibilité de réaliser une action pour une tâche de recherche d'information.</i> Trouver un champ de formulaire, affiner la recherche, lire les résultats de la recherche, télécharger un document, sauvegarder les résultats de la recherche.	<i>Néant</i>
[Murphy, Kuber, McAllister, et al. 2007]	Liens hypertextes, formulaires, présence et pertinence des alternatives textuelles, problèmes de formats de fichier.	Pas d'information sur la localisation spatiale du contenu lu, difficulté de la construction d'une représentation globale d'une page, charge mentale importante, perte de temps notamment pour filtrer les contenus non pertinents.

**Tableau 2- Études sur l'utilisabilité des lecteurs d'écran**

La lecture du Tableau 2 confirme que les lecteurs d'écrans ne sont pas évalués de manière spécifique au-delà de leurs limites techniques. [Power et al. 2012] rapportent des classes de problèmes rencontrés par 32 utilisateurs de lecteurs d'écran sans préciser lesquels sont imputables aux contenus et lesquels le sont aux technologies d'assistance elles-mêmes. [Murphy, Kuber, McAllister, et al. 2007] rapportent les résultats de 30 interviews semi-structurées qui mettent en lumière deux limites des lecteurs d'écrans : le manque d'information sur la localisation spatiale de l'information pénalisant l'orientation et la collaboration avec des personnes voyantes, et la difficulté de construire une représentation globale de la page (pour les utilisateurs novices uniquement), résultat appuyé également par [Baumgartner et al. 2010] expliquant que les lecteurs d'écran fournissent une information trop morcelée, fragmentaire ; la charge mentale et le facteurs temps ont également été mentionnés mais sont imputables en grande parties aux contraintes de l'audio (voir 3.2). Ces derniers résultats prouvent bien que les lecteurs d'écrans sont améliorables, ce qui est d'ailleurs illustré par une remarque d'un utilisateur avancé rapportée dans cette étude (p.89) : « on essaie de tirer le meilleur parti des technologies disponibles en mettant en place des stratégies adéquates, mais ces technologies doivent encore être développées pour constituer une solution utilisable et accessible ». Bien qu'ayant de multiples implications, cette étude qualitative de [Murphy, Kuber, McAllister, et al. 2007] a surtout mis en avant le manque d'information spatiale dans la restitution des lecteurs d'écrans, chose à laquelle ils proposent de remédier avec un prototype multimodal audio et à retour d'effort [Murphy, Kuber, Strain, et al. 2007], sans mentionner quels autres fonctions et paradigmes utilisés par les lecteurs d'écran pourraient être améliorés ou modifiés. Pourtant, de nombreux travaux proposent de nouvelles méthodes de navigation non-visuelles [Rohani Ghahari et al. 2012; Yang et al. 2013; Vigo et al. 2009].

Ainsi, à notre connaissance, la littérature ne comporte pas d'études évaluant l'utilisabilité des lecteurs d'écran, que ce soit dans le domaine de l'accès aux contenus ou de l'interaction avec le système d'exploitation ; de tels travaux permettraient pourtant d'améliorer l'accessibilité numérique de manière générale, notamment pour intégrer les résultats de recherche épars aux technologies actuelles. Ce manque peut s'expliquer par l'absence de point de comparaison : les lecteurs d'écrans existent en nombre limité et présentent tous des méthodes similaires pour accéder aux contenus et interagir avec le système d'exploitation. Cette absence d'alternative comme moyen de comparaison se retrouve également chez les utilisateurs : [Lazar et al. 2007](p.265) « les utilisateurs n'ont pas rapporté de problèmes liés aux lecteurs d'écrans susceptibles d'améliorer leur utilisabilité [...] mis à part la correction des erreurs techniques » ; cette étude a pourtant montré que l'accès aux web avec les lecteurs d'écrans était source de nombreuses frustrations. Ce deuxième point constitue un autre élément explicatif de l'absence de critique des lecteurs d'écrans en vue de leur amélioration, car les études rapportant les problèmes et suggestions des utilisateurs sont en effet limitées à ce qu'ils connaissent. Cependant, une approche plus exploratoire [Murphy, Kuber, McAllister, et al. 2007] permet de pointer les limites et manques des technologies. Enfin, une autre explication proviendrait du lien fort entre l'accessibilité effective et l'accessibilité du contenu accédé ; en effet, même les méthodes et fonctions d'accès les plus performantes ne rendront un contenu mal conçu et non annoté que peu accessible. Cela rend également l'évaluation des technologies d'accès (les lecteurs d'écran par exemple) à grande échelle complexe, étant donné que les résultats seront forcément dépendants de l'accessibilité des contenus accédés ; le simple fait d'évaluer les performances et l'utilisabilité d'un lecteur d'écran sur deux pages différentes pose déjà problème. Cependant, cela reste faisable, puisque nous avons testé différentes méthodes de restitution de la mise en forme dans le Chapitre IV.

#### ***5.2.d. Les navigateurs Web « audio »***

---

Plusieurs navigateurs Web ou extensions de navigateurs Web existants proposent un accès non-visuel aux contenus, avec un fonctionnement similaire aux lecteurs d'écran mais ici limité à la navigation Web. On peut citer ChromeVox<sup>61</sup> et FireVox<sup>62</sup> qui sont tous deux des extensions aux navigateurs Web Chrome et Firefox, destinées au grand public. À noter que ces extensions n'utilisent, contrairement aux lecteurs d'écran, que le mode audio en sortie.

Plusieurs prototypes de recherche ont vu le jour, reprenant le même principe de proposer une navigation « audio » du Web, en rajoutant parfois certaines fonctionnalités. *HearSay* [Borodin and Bigham 2008] est un navigateur Web pour les déficients visuels proposant une annotation collaborative des pages (les utilisateurs peuvent apposer des labels sur différents éléments des pages web, qui seront consultables et modifiables par les autres utilisateurs) et un support multi-langues. *WebAnywhere* propose lui une interaction simplifiée et surtout un accès Web sans installation locale ; les résultats de l'étude utilisateur sont cependant mitigés [Bigham et al. 2008]. Enfin, *CSurf* [Mahmud 2007] reprend les fonctionnalités classiques des navigateurs audio en rajoutant des informations de contexte à la navigation hypertexte à l'aide de techniques d'apprentissage automatique, qui identifient le thème de la page courante pour mieux guider la navigation sur les pages accédées ensuite à travers les liens hypertextes.

---

<sup>61</sup> <https://chrome.google.com/webstore/detail/chromevox/kgejglhpjiefppelpmljqlcjbjhboiplfn?hl=fr>

<sup>62</sup> <http://firevox.clcworld.net> ; FireVox n'est cependant plus compatible avec les versions récentes de Mozilla Firefox

Ce type de logiciel présente deux avantages : la possibilité de fournir une interface simplifiée pour des utilisateurs novices, et la portabilité de l'application. La dernière étude en date de WebAim montre que leur utilisation est très marginale : seuls 0.4% des 1400 sondés utilisaient un navigateur audio, ici ChromeVox<sup>63</sup>.

### 5.3. Blocs-notes braille

Les blocs-notes braille (ou *braille notetakers*) sont des ordinateurs portables sans écran dédiés à l'accès non-visuel des contenus, dont le petit format les rend mobiles et notamment utilisés pour la prise de note en cours (voir étude en situation, partie 7). Les blocs-notes braille sont à minima composé d'un clavier (braille ou alphanumérique), d'une sortie vocale et de touches fonctionnelles. Les plus haut de gamme incluent une plage braille dynamique pour la lecture du braille. Ces dispositifs intègrent des logiciels similaires aux lecteurs d'écrans pour permettre aux déficients visuels d'interagir avec le système, utilisant toutefois des raccourcis claviers et des touches différentes étant donné que les périphériques d'entrée sont différents (touches fonctionnelles supplémentaires, et clavier braille remplaçant parfois le clavier alphanumérique utilisé en entrée sur ordinateur avec les lecteurs d'écran). De plus, les fonctions proposées ne sont pas exactement les mêmes et l'ensemble des fonctions des lecteurs d'écran peuvent ne pas être implémentées, ceci étant lié au fait que les systèmes d'exploitation utilisés par les blocs-notes sont différents de ceux des ordinateurs classiques.



Figure 15 – Bloc-notes braille « Apex BT 32 » de HumanWare

Les systèmes d'exploitations intégrés (souvent propriétaires et dédiés, adaptés) là encore varient en fonction de la gamme du produit ; parfois basés sur des SE existants ils proposent ainsi des fonctionnalités variées en fonction du modèle. Les blocs-notes les plus basiques incluent des fonctions de lecture et écriture de fichiers texte, une calculatrice et un calendrier (à la manière des premiers PDA<sup>64</sup>), le transfert de fichiers sur clé USB ainsi que l'impression et la numérisation de documents à l'aide d'autres périphériques USB. Les appareils les plus avancés permettent de naviguer sur internet, gérant les fichiers multimédia, les emails, la bureautique, et intégrant parfois même une sortie vidéo. Les prix vont de 1 000 USD pour les modèles les plus basiques à 7 000 USD (et plus) pour les plus avancés<sup>65</sup>.

Ils sont relativement répandus chez les élèves déficients visuels malgré leur coût élevé, chez qui d'ailleurs ils représentent l'outil de prédilection pour la prise de note en cours de ce que nous avons

<sup>63</sup> Voir <http://webaim.org/projects/screenreadersurvey5/>

<sup>64</sup> Personal Digital Assistant : l'ancêtre du smartphone.

<sup>65</sup> Voir : <http://www.perkinsproducts.org/store/en/38-notetakers>

pu observer pendant l'étude en situation. En effet, sans cet outil les élèves doivent « poinçonner » le braille manuellement pour la prise de note, ou alors requérir l'aide d'un auxiliaire de vie scolaire. Les quelques travaux de recherches impliquant des blocs-notes braille sont d'ailleurs centrés autour de l'enseignement et de l'insertion et l'utilisation des technologies dans ce cadre [Tsiakali and Argyropoulos 2015]. Ces dispositifs sont mis en avant comme permettant de répondre au déclin de l'apprentissage du braille [Wiazowski 2014], notamment en facilitant l'accès en braille aux contenus proposés par les enseignants qui peuvent les transférer directement sur le bloc-notes par bluetooth ou via une clé USB. D'autres études appuient l'intérêt pour les étudiants non-voyants d'utiliser des blocs-notes braille : ils permettraient une plus grande implication dans les activités de lecture et d'écriture [Bickford and Falco 2012] et n'auraient aucun impact négatif sur la qualité de l'apprentissage du braille comparativement à l'utilisation de braille papier [Kamei-Hannan and Lawson 2012].

## 5.4. Format et lecteurs DAISY

DAISY (Digital Accessible Information System) est un standard technique visant à rendre les documents écrits accessibles aux déficients visuels et à tout autre type de déficience relative à la lecture visuelle. DAISY vise à standardiser la production et la diffusion de livres audio (ou DTBook pour Digital Talking Book), ou plus généralement de documents oralisables ; à noter que même si le mode de sortie utilisé est l'audio avec la parole, certains prototypes de recherche intègrent une sortie braille [Minatani 2012]. Le standard DAISY est basé techniquement sur des fichiers XML, et spécifie comment les fichiers doivent être construits et annotés ; les fichiers sont ensuite soit lus par synthèse vocale, soit simplement joués sur la base de fichiers audio pré-générés. Un livre-audio DAISY diffère d'un livre-audio « classique » (constitué d'enregistrements de texte lus par synthèse vocale ou par un lecteur humain) par plusieurs aspects : possibilité de régler le débit de lecture, navigation entre les titres/paragraphes/lignes, insertion de signets dans le livre audio, synchronisation entre le visuel du document et le texte lu et enfin la possibilité de rechercher du texte dans le livre. Ces fonctions sont disponibles à travers un lecteur DAISY, qui peut être un lecteur logiciel utilisable sur ordinateur ou tablette comme le lecteur AMIS [Demeglio et al. 2002], ou un lecteur physique lisant les livres audio DAISY depuis un CD ou une clé USB. Les lecteurs DAISY permettent également l'oralisation de tableaux, d'expressions Mathématiques et de contenus graphiques lorsqu'ils sont décrits verbalement [Motti and Burger 2009].

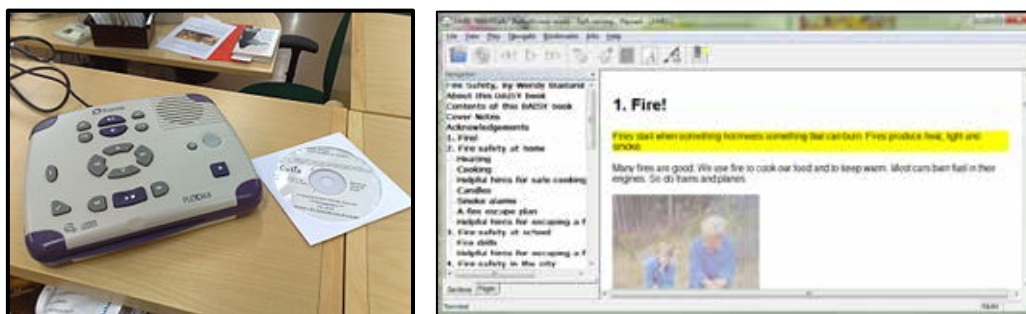


Figure 16 - Lecteur de livres audio DAISY physique (à gauche), et logiciel avec le lecteur AMIS (à droite)

Le standard DAISY est largement répandu dans les bibliothèques numériques de par le monde, notamment en France avec la BNFA (*Bibliothèque Numérique Francophone Accessible*)<sup>66</sup>, ainsi que dans d'autres recueils numériques de contenus comme au Danemark avec Biblus [Christensen and Stevns 2012] ou encore en Corée du sud [Hann et al. 2013]. Le standard DAISY et les lecteurs logiciels semblent jouir d'une opinion positive auprès des élèves déficients visuels [Spooner 2014] qui apprécient leur simplicité d'usage, particulièrement lorsqu'un lecteur logiciel est installé sur leur bloc-notes braille personnel.

Enfin, on peut mentionner l'existence du Kurzweil 3000<sup>67</sup> qui est un environnement logiciel reprenant les mêmes bases fonctionnelles que les lecteurs DAISY pour l'accès aux documents. Cette application diffère cependant des lecteurs DAISY logiciels par le fait qu'elle n'utilise pas de format particulier de fichier et gère les fichiers Web et documents textuels, permet l'ajout de notes manuscrites ou vocales et l'édition de documents, inclut un dictionnaire et un thesaurus, et gère également la numérisation et l'impression de documents [Steinberg and Murphy 2012]. Ainsi le Kurzweil 3000 est un environnement de travail destiné à un cadre éducatif pour les élèves déficients.

## 5.5. Conclusion

---

Dans cette partie nous avons vu les différentes technologies permettant aux non-voyants d'accéder aux contenus numériques. En résumé, la transmission de l'information se fait principalement par synthèse vocale, mais aussi par lecture braille à travers les plages braille dynamique ; l'expressivité des modes audio et tactile sont assez peu exploités dans les technologies actuelles, bien que de nombreux travaux de recherche existent montrant le potentiel de restitution de ces modes. Les lecteurs d'écrans font office d'interface entre les périphériques informatiques basés sur des interfaces graphiques et les utilisateurs non-voyants ; ils gèrent l'accès aux contenus mais aussi au système d'exploitation et aux applications installées. D'autres technologies plus spécialisées existent comme les navigateurs audio permettant d'accéder aux contenus Web et les lecteurs DAISY permettant la lecture de livres audio. Enfin, les blocs-notes brailles sont des ordinateurs portables conçus avec une interface non-visuelle native, surtout utilisés dans un cadre scolaire, et représentent une solution « tout en un » matérielle et logicielle pour l'accès aux contenus numériques (pour les plus récents).

Les lecteurs d'écran, que ce soit sur ordinateur, sur dispositif mobile ou sur bloc-notes braille, représentent l'intermédiaire principal utilisé par les déficients visuels pour accéder aux contenus numériques. Il en existe un certain nombre mais seuls quelques-uns sont plébiscités par les utilisateurs ; de plus, tous les lecteurs d'écrans présentent un mode de fonctionnement et des commandes très similaires. Ils fonctionnent par utilisation de raccourcis clavier associés à des commandes de lecture pour gérer la synthèse vocale, la navigation et les différentes fonctions d'accessibilité. La quantité de commandes disponibles et le potentiel de configuration de ces logiciels sont assez extensifs (ex. plus de 300 raccourcis clavier pour JAWS), ce qui montre qu'ils ont adopté un point de vue « exhaustif » pour l'accessibilité, à savoir qu'ils essaient de rendre disponibles un maximum d'informations et de fonctions pour assurer un premier niveau d'accessibilité (*disponibilité* de l'information, voir 2). Les recherches semblent d'ailleurs centrées autour de l'accessibilité des

---

<sup>66</sup> [www.bnfa.fr](http://www.bnfa.fr)

<sup>67</sup> [www.kurzweiledu.com](http://www.kurzweiledu.com)



contenus (constituant certes un prérequis important) et non pas de l'utilisabilité des technologies d'assistance, avec une approche basée sur les problèmes rencontrés avec ces technologies ; cela montre qu'une étape reste à franchir pour tendre vers une plus grande utilisabilité de l'accès aux contenus numériques, à travers l'amélioration des technologies d'assistance. Plusieurs auteurs évoquent d'ailleurs la nécessité de passer d'une approche exhaustive à une approche *holistique* [Giraud et al. 2011; Kelly et al. 2007] pour la conception des technologies et interactions dans le cadre de l'accessibilité, à savoir une approche globale reprenant les principes de la conception centrée utilisateur, tenant en compte à la fois des contenus, des utilisateurs et du contexte d'utilisation.

Enfin, nous verrons dans la section suivante que les technologies d'assistance pour l'accès aux contenus ne prennent en compte qu'une partie réduite de la sémantique de la MFM des documents, en se basant sur principalement sur les langages et formats de fichiers utilisés pour structurer et diffuser les documents numériques.

## 6. Vers la restitution des architectures textuelles pour l'accessibilité des contenus numériques

---

---

Nous avons vu dans le chapitre précédent que la mise en forme matérielle (MFM) des textes constitue l'ensemble des propriétés typographiques, dispositionnelles et lexico-syntaxiques des textes qui sont issues d'une volonté de l'auteur de structurer son discours et d'informer le lecteur. En clair, la mise en forme des textes inclut un ensemble de marques signifiantes (les marques de MFM) qui font partie intégrante de la sémantique du texte. Cette « sémantique visuelle » permet de définir des architectures textuelles, c'est-à-dire des objets textuels de certains types, et des relations entre ces objets typés (ex. un objet est une illustration / un label d'un autre objet) ; de plus, la MFM possède un ensemble de propriétés *fonctionnelles* (voir Chapitre I pour plus de détails). On comprend dès lors pourquoi prendre en compte ces marques de MFM dans le cadre d'un accès non-visuel aux documents est essentiel, ne serait-ce qu'à cause du fait que la MFM fait partie du signifiant<sup>68</sup> global des textes et qu'elle rentre donc en compte pour l'interprétation de leur sens, du signifié.

Nous verrons dans cette partie comment les technologies de lecture actuelles prennent en compte la mise en forme des textes, et quels travaux de recherche ont été conduits dans ce domaine. Enfin, nous terminerons par quelques exemples d'architectures textuelles dont la restitution serait bénéfique pour un accès non-visuel aux textes.

### 6.1. Prise en compte de la MFM par les lecteurs d'écran

---

La question posée ici est de savoir dans quelle mesure les technologies de lecture, permettant un accès non-visuel aux contenus, restituent les architectures textuelles exprimées par la MFM des textes ainsi que les fonctions de lecture correspondantes. Nous nous plaçons ici dans le cas de l'accès à un contenu annoté « au mieux », c'est-à-dire implémentant l'ensemble des normes d'accessibilité décrites précédemment (voir 4.3).

Les lecteurs d'écran sont le principal outil utilisé par les déficients visuels pour accéder aux contenus ; d'autres technologies plus ciblées existent toutefois (voir partie précédente) mais reposent sur les mêmes principes de fonctionnement. L'accès aux contenus par les lecteurs d'écran n'est pas basé sur la MFM, au sens où les pages ne sont pas segmentées sur la base des marques de mise en forme significatives, mais sur la base des principaux objets définis par l'encodage des fichiers et le langage utilisé, les principaux étant : les titres, les frames, les liens hypertextes, les tableaux, les listes (énumérations), les paragraphes et les emphases. Certains de ces objets sont segmentables sur la base de leur traits de MFM mais d'autres ne correspondent pas toujours à des objets textuels, cela étant lié au fait qu'un élément défini, par exemple, comme un titre dans le code source du document, n'aura pas forcément les traits visuels d'un titre et/ou ne sera pas utilisé en tant que titre dans le document.

---

<sup>68</sup> Au sens linguistique du terme, c'est-à-dire le support physique du sens (du signifié) qui est une représentation mentale des concepts dénotés et connotés par le signifiant

Les lecteurs d'écran prennent donc en compte certains objets textuels représentés par des objets du langage dans lequel est produit le contenu (ex. les titres, en HTML <h>), et le font de deux manières : à travers des fonctions de navigation, et à travers leur signalement pendant la lecture du contenu. Les fonctions de navigation permettent de lister les objets de la page par type et naviguer séquentiellement d'un objet à l'autre. Le signalement des objets dans la lecture se fait (par défaut) en lisant par synthèse vocale le type de l'objet rencontré et parfois certaines de ses propriétés (ex. « titre de niveau 1 » ou « lien hypertexte renvoyant à <un emplacement> dans la page courante»), ou alors par changement de voix (ex. lire les liens hypertextes avec une voix dédiée) ou ajout de sons non-verbaux ; la signalisation des objets est configurable. Ainsi, la lecture est limitée à quelques objets textuels et uniquement s'ils sont correctement annotés dans le contenu ; par exemple, si sur une page Web, les définitions sont en orange, cette information ne peut pas être restituée (autres exemples dans la section 6.3). De la même façon, les relations entre objets du texte, exprimées visuellement, ne sont pas prises en compte : un encadré de texte peut représenter une illustration du contenu en vis-à-vis.

Les choses sont cependant en train d'évoluer avec les nouvelles normes d'accessibilité ARIA, qui permettent une annotation sémantique du contenu, et notamment des « zones » des pages Web et des relations de composition (imbrication visuelle de zones de contenu) et de labellisation (un élément textuel constituant le label d'un autre élément) entre objets. Les *zones* ARIA permettent d'introduire un typage des différentes régions de la page Web (contenu principal / périphérique, navigation, etc.) et en cela prennent en compte la MFM des documents, car c'est elle qui permet de définir ces zones ; les utilisateurs peuvent ensuite naviguer d'une zone à l'autre. La relation de composition est souvent rendue explicite par les marques de MFM car elle réfère à l'imbrication visuelle d'éléments. La relation de « labellisation » est également issue de marques de MFM, par exemple le fait qu'un texte soit placé à proximité d'une image ou d'une zone de texte plus grande, éventuellement avec une typographie particulière, fait que l'on peut de manière *probable* inférer que celle-ci labellise l'élément à proximité.



Figure 17- Exemple de relation de composition (gauche) et de relation de labellisation (droite)

Depuis l'avènement des normes ARIA, les lecteurs d'écran commencent cependant à prendre en compte les relations de composition et de labellisation (exemples de la Figure 17), en proposant des fonctions permettant de naviguer entre objets en fonction de ces relations [Freedom Scientific 2013]. Une autre relation récemment intégrée dans ARIA est l'ordre de lecture des éléments, qui peut être défini pour éviter les confusions pendant l'accès via un lecteur d'écran. Ainsi, seules les relations de MFM prises en compte par les lecteurs d'écran sont celles implémentées par les normes ARIA et les relations spatiales définies par les tableaux (navigation entre colonnes et lignes des tableaux).

Enfin, les lecteurs d'écran permettent d'accéder à certaines propriétés de mise en forme « brutes » du texte en cours de lecture : sa couleur et sa police.

Les autres technologies de lecture comme les navigateurs audio et les lecteurs DAISY utilisent le même mode de fonctionnement que les lecteurs d'écran, à savoir une navigation et une restitution

basées sur certains types d'objets définis par les formats des documents ; ces technologies sont cependant plus limitées que les lecteurs d'écran en termes de prise en compte de la MFM car ils ne prennent pas en compte les normes ARIA.

## **6.2. Exemples empiriques d'apports de la MFM à l'accessibilité des contenus**

---

On retrouve dans la littérature de nombreux travaux traitant directement ou indirectement du rôle de la MFM pour l'accès aux documents ; nous en citerons quelques-uns afin d'illustrer l'importance de la MFM comme objet d'étude dans ce cadre. Nous séparerons les travaux qui traitent implicitement de la MFM, portant sur certaines propriétés visuelles des documents, et les travaux qui en traitent explicitement basé sur le cadre théorique de la MFM posé par le Modèle d'Architectures Textuelles (Chapitre I).

La MFM permet au lecteur (par l'utilisation conventionnée de marques de MFM) d'inférer l'ordre de lecture des blocs de texte ; [Ferilli et al. 2014] utilisent une méthode semi-automatique basé sur la disposition des blocs de texte et la présence de séparateurs visuels pour inférer l'ordre de lecture sur des pages de journaux. La MFM donne également des informations sur les regroupements et le type de ces regroupements : [Baumgartner et al. 2010] tentent de rendre accessibles ces informations qu'ils appellent de « *gestalt* » en proposant des chemins de navigation entre les différents regroupements identifiés sur une page Web à l'aide d'indices visuels et textuels. [Murphy, Kuber, McAllister, et al. 2007] ont montré à travers une série d'interviews l'utilité de la spatialisation de l'information pour les non-voyants, c'est-à-dire la localisation des informations sur la page, qui est importante pour construire une représentation de celle-ci et pour collaborer avec des personnes voyantes [Jovina and Van Oostendorp 2006] ; ils ont ici développé un prototype audio-haptique permettant de localiser les informations sur une page web pendant sa lecture [Murphy, Kuber, Strain, et al. 2007]. [Pontelli et al. 2002] se sont eux intéressés à la reformulation de tableaux sous forme de graphes, cette reformulation étant basée sur la proximité spatiale des cellules, les titres de colonnes et de lignes mais aussi des propriétés visuelles (ex. les cellules avec un fond gris) afin que ces graphes soient représentatifs de la *structure sémantique* des objets d'origine. Enfin, il est intéressant de faire le parallèle avec l'accès non-visuel aux expressions mathématiques, qui elles présentent une organisation explicite à travers des indices syntaxiques (signes opératoires, parenthèses, etc.) là où l'organisation et les informations de MFM des textes sont plus implicites. De nombreux travaux ont montré la complexité d'un accès non-visuel à ces objets [Archambault 2009], et l'utilité de restituer dans ce cadre les informations relatives aux regroupements définis par la syntaxe mathématique [Bates and Fitzpatrick 2010].



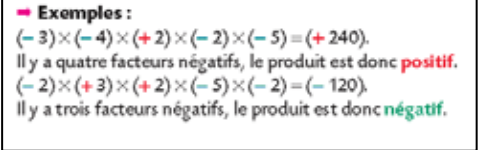
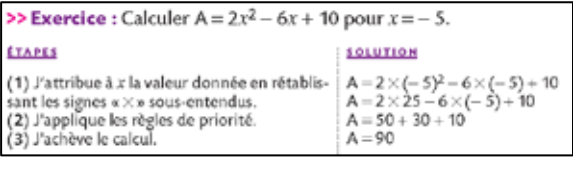
Nous avons vu dans le chapitre précédent le rôle prépondérant que joue la MFM pour la compréhension et la représentation du contenu d'un document. Le Modèle d'Architectures Textuelles offre un cadre théorique permettant de formaliser les marques de mise en forme pour leur restitution. En utilisant ce modèle, [Maurel 2004] propose dans sa thèse 3 stratégies différentes pour restituer à l'oral les marques de MFM en utilisant des descriptions verbales et des indices prosodiques. Ces travaux ce sont concentrés sur l'étude des énumérations qui à eux seuls représentent des objets riches de différentes propriétés visuelles intriquées avec d'autres relations du discours [Luc 2000]. Plus récemment, le même auteur a étudié la restitution vibrotactile de la structure visuelle de documents [Maurel et al. 2012], c'est-à-dire l'organisation spatiale des

différentes zones de contenu. Ces différentes « zones » délimitées visuellement peuvent d'ailleurs (d'après le MAT) présenter certaines relations entre elles et appartenir à certaines catégories (ex. zone de navigation) ; [Bou-Issa 2010] a lui étudié la représentation de ces informations sur des pages tactiles en relief, utilisant 3 niveaux de représentation incluant chacun des informations jugées plus ou moins importantes d'après des experts.

Cette section montre la diversité des propriétés et informations liés à la MFM, qui ne sont pour la plupart que très sommairement prises en compte par les technologies de lecture. Ces informations revêtent d'une importance particulière pour les lecteurs voyants, c'est pourquoi l'étude de la MFM et de sa restitution non-visuelle représente une étape importante vers l'amélioration de l'accessibilité des documents numériques pour les déficients visuels.

### 6.3. De l'intérêt de la restitution de la sémantique de la MFM : exemples d'architectures textuelles non-restituées

En plus de tous les travaux de recherche cités dans la section précédente, il est intéressant d'illustrer la diversité des architectures textuelles et la complexité de leur accès non-visuel à travers quelques exemples. On présentera donc dans cette section des architectures textuelles qui ne sont pas prises en compte par les technologies d'assistance actuelles, afin d'illustrer les gains potentiels de la restitution de cette sémantique pour le moment ignorée.

Image incluant l'architecture textuelle	Description
 <p>The image shows a news article layout with multiple headlines and sub-headlines. The headlines are: 'Syria Government air strikes on market place kill at least 82 people', 'Tianjin blasts Sodium cyanide on site may have been 70 times allowed amount', 'They're worth a lot, and they know it' Amazon boss on Top Gear revival.</p>	<p><u>Titres et sous titres</u> : ici sur le site du <i>Guardian</i> la page d'accueil comprend des renvois vers différents articles. Chaque renvoi comprend un titre et un sous-titre qui se complètent pour intituler les articles. Or, la relation « sous-titrer » n'existe pas pour les technologies de lecture. Restituer ces deux objets comme deux titres pourrait d'ailleurs induire en erreur.</p>
 <p>The image shows a programming definition for a table. The text is: 'Définition: Un tableau est une structure de données contenant un groupe d'éléments tous du même type. Les [] spécifiant qu'il s'agit d'un tableau peuvent être placés avant ou après le nom du tableau.' Below this is code: '// T1 est un tableau à une dimension de type int T1 []' and '// T2 est un tableau à une dimension de type char char[] T2'. Below the code is: 'Un tableau peut être initialisé: int T1x (3) = {1, 2, 3, 4}; char[] T2 = {'a', 'b', 'c'};</p>	<p><u>Typage visuel</u> : ici sur un extrait d'un cours de programmation, les commentaires sont en vert et les types de variables en gras. Les technologies de lecture actuelle ne prennent pas en compte ce typage visuel, et sont restreintes à certains types d'objets HTML.</p>
 <p>The image shows mathematical examples: 'Exemples : (-3) x (-4) x (+2) x (-2) x (-5) = (+240). Il y a quatre facteurs négatifs, le produit est donc positif. (-2) x (+3) x (+2) x (-5) x (-2) = (-120). Il y a trois facteurs négatifs, le produit est donc négatif.'</p>	<p><u>Typage visuel et correspondance entre objets</u> : on observe sur cet extrait d'un manuel de mathématiques (numérisé) une correspondance entre la mise en forme des signes et des mots positifs et négatifs. Cette correspondance n'est pas restituée car le typage ne l'est pas.</p>
 <p>The image shows a math exercise: '&gt;&gt; Exercice : Calculer A = 2x² - 6x + 10 pour x = -5.' Below it are steps: 'ÉTAPE (1) J'attribue à x la valeur donnée en rétablissant les signes « x » sous-entendus. (2) J'applique les règles de priorité. (3) J'achève le calcul.' To the right is the solution: 'SOLUTION A = 2 x (-5)² - 6 x (-5) + 10 A = 2 x 25 - 6 x (-5) + 10 A = 50 + 30 + 10 A = 90'</p>	<p><u>Disposition en double colonage</u> : un phénomène relativement courant, ici extrait d'un manuel de mathématique, est une disposition en colonnes faisant correspondre les étapes d'une méthode de résolution avec les étapes de la solution. Une étape de la méthode peut ainsi être mise en relation, graphiquement, avec une ou plusieurs étapes de la solution.</p>

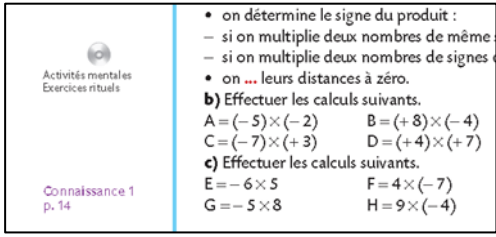
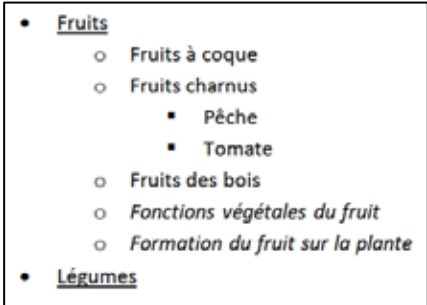
 <p>Activités mentales Exercices rituels</p> <p>Connaissance 1 p. 14</p> <ul style="list-style-type: none"> <li>• on détermine le signe du produit :             <ul style="list-style-type: none"> <li>– si on multiplie deux nombres de même signe</li> <li>– si on multiplie deux nombres de signes différents</li> </ul> </li> <li>• on ... leurs distances à zéro.</li> </ul> <p><b>b) Effectuer les calculs suivants.</b>  <math>A = (-5) \times (-2)</math>      <math>B = (+8) \times (-4)</math>  <math>C = (-7) \times (+3)</math>      <math>D = (+4) \times (+7)</math></p> <p><b>c) Effectuer les calculs suivants.</b>  <math>E = -6 \times 5</math>              <math>F = 4 \times (-7)</math>  <math>G = -5 \times 8</math>              <math>H = 9 \times (-4)</math></p>	<p><u>Disposition en marge et annotations</u> : ici la mise en vis-à-vis de blocs de texte dans la marge gauche du document permet de typer ces blocs de texte en tant que <i>renvois</i>, et la disposition du deuxième bloc (texte en violet) nous permet de voir qu'il annote la question C en particulier.</p>
 <ul style="list-style-type: none"> <li>• <b>Fruits</b> <ul style="list-style-type: none"> <li>○ Fruits à coque</li> <li>○ Fruits charnus                     <ul style="list-style-type: none"> <li>▪ Pêche</li> <li>▪ Tomate</li> </ul> </li> <li>○ Fruits des bois</li> <li>○ Fonctions végétales du fruit</li> <li>○ Formation du fruit sur la plante</li> </ul> </li> <li>• <b>Légumes</b></li> </ul>	<p><u>Structures hiérarchiques à plusieurs niveaux</u> : ce type de structure visuelle est assez courante, notamment avec les tables des matières. Or, leur restitution est assez complexe et seulement partiellement couverte par les technologies de lecture. Elle permet par exemple de savoir combien d'items il y a à chaque niveau hiérarchique, quelle est leur imbrication, et permettent de naviguer entre ces niveaux. Les lecteurs d'écrans actuels ne fournissent que des fonctions linéaires de navigation dans ces listes, alors que la navigation non-visuelle dans les structures hiérarchiques inclut de nombreuses autres fonctions importantes comme l'attestent [Smith et al. 2004].</p>

Tableau 3 - Exemples d'architectures textuelles non-restitués par les technologies de lecture

La liste précédente a valeur d'exemple, et représente un échantillon de la diversité des architectures textuelles que l'on peut retrouver dans la nature. Même les exemples qui semblent les plus atypiques peuvent être fréquents dans certains contextes ; dans le cas d'un manuel scolaire particulier<sup>69</sup> on retrouvait par exemple le 4<sup>ème</sup> exemple du tableau entre 2 et 8 fois par Chapitre du manuel. Ces exemples d'architectures textuelles représentent une infime partie de la quasi-infinité de phénomènes existants, d'autant que la liberté d'utilisation de la MFM par les auteurs fait qu'ils sont libres, dans une certaine mesure, de créer leurs propres architectures textuelles, et sont surtout libres d'utiliser les marques de MFM qu'ils souhaitent pour les exprimer. Comparativement aux quelques architectures textuelles prises en compte par les technologies d'assistance, il semble ainsi important de poser un cadre théorique pour la formalisation de la MFM afin de rendre accessibles un maximum de cette sémantique pour le moment ignorée.

<sup>69</sup> Manuel « TRIANGLE 4<sup>ème</sup> » (2009) des éditions HATIER dont une copie nous a généreusement été fournie à des fins de recherche par ces éditions.

## **7. Étude en situation : accès et accessibilité de l'information pédagogique pour les élèves déficients visuels**

---

---

Dans le cadre d'une démarche ergonomique de conception, nous avons réalisé une étude exploratoire [Sorin, Lemarié, Dupuis, et al. 2015] auprès d'élèves déficients visuels dans un contexte d'éducation inclusive. L'éducation inclusive ou *éducation en inclusion* désigne, d'après la Déclaration de Salamanque<sup>70</sup> [UNESCO 1994], un « *consensus par lequel les enfants avec des besoins éducatifs particuliers devraient être inclus dans les mêmes institutions scolaires que la majorité des enfants* ». Depuis l'introduction du concept, de nombreux efforts en termes politiques, de ressources et de structures mises en place [UNESCO 2008; Chirac 2005] ont été réalisés afin d'inclure en classe ordinaire les élèves présentant une déficience quelle qu'elle soit. L'objectif de notre étude en situation était, dans le cas d'élèves déficients visuels, d'évaluer l'accessibilité des documents numériques et papier mis à leur disposition pour apprendre, et plus particulièrement l'accessibilité de la sémantique de la mise en forme de ces documents et l'impact éventuel de leur inaccessibilité. Le choix de ce cas d'étude est lié à l'importance de l'accessibilité des contenus pour l'apprentissage dans un contexte scolaire ; de plus, la mise en forme joue un rôle particulièrement important pour la compréhension des contenus didactiques et donc dans le processus d'apprentissage (voir Chapitre I).

Cette étude visait également, de par sa nature exploratoire, à dresser un portrait de la situation des apprenants déficients visuels inclus en classe ordinaire, du contexte de l'apprentissage, des outils à leur disposition, des acteurs impliqués et surtout des différents problèmes et contraintes qu'ils rencontrent dans leur accès à l'information pédagogique de manière général (documents, discours oral, technologies d'assistance, etc.). Ce type d'approche d'étude exploratoire d'une situation est répandu en ergonomie [Vicente 1999] et appliqué à la conception de systèmes informatiques [Fidel and Pejtersen 2004]. Les résultats de cette étude pourront ainsi servir de guides à la conception de technologies d'assistance pour les élèves déficients visuels scolarisés en inclusion, mais aussi potentiellement contribuer à faire évoluer les pratiques des acteurs du domaine. L'objectif était également pour nous, en tant que concepteurs et chercheurs, de développer une connaissance et une compréhension des situations rencontrées par les utilisateurs concernés par les applications de nos travaux.

Nous avons réalisé pour cette étude 13 entretiens semi-directifs avec les différents acteurs de la scolarité des élèves déficients visuels, 3 observations de cours (4h30 au total) impliquant les élèves déficients à l'aide d'une grille d'analyse, et enfin une analyse de 29 pages de documents de cours transcrits en braille.

### **7.1. Contexte général : institutions, acteurs et moyens**

---

Les élèves déficients visuels étudiant dans un contexte inclusif sont la plupart du temps suivis par des institutions spécialisées qui jouent plusieurs rôles notamment celui de soutien scolaire et de suivi général du bien-être et de l'autonomie des jeunes élèves déficients. Dans certains cas particuliers, ces institutions prennent en charge l'intégralité de l'éducation des élèves. Ici, les élèves partageaient leur temps entre leurs établissements d'éducation publics et *l'Institut des Jeunes Aveugles de*

---

<sup>70</sup> Conférence mondiale sur l'éducation et les besoins éducatifs spéciaux : accès et qualité.

*Toulouse*<sup>71</sup> (IJA). Les acteurs de l'institut sont en contact constant avec les établissements publics de leurs élèves pour leur fournir des documents braille, effectuer des ajustements pédagogiques avec les enseignants de l'établissement et résoudre des problèmes pratiques. Les instituts spécialisés comme l'IJA sont également en contact avec les parents d'élèves et les auxiliaires de vie scolaire (AVS).

On retrouve au sein de l'IJA deux types d'acteurs majeurs : (1) les éducateurs spécialisés et (2) les transcrip-teurs braille. (1) Les éducateurs spécialisés font le lien entre les élèves, les professeurs des établissements publics et les transcrip-teurs ; chaque élève est suivi par un et un seul éducateur à l'institut. Leurs rôles sont très divers et ils se chargent aussi bien des cours de soutien scolaires, de fournir de l'aide aux élèves pour utiliser les technologies d'assistance à leur disposition, ou encore des problèmes plus pratiques. Ils se chargent également d'ajuster le volume et la nature des contenus à disposition des élèves, et doivent pour cela connaître parfaitement les capacités des élèves qu'ils prennent en charge et effectuer les ajustements nécessaires auprès des professeurs des établissements publics et des transcrip-teurs braille. (2) Ces derniers produisent des documents braille et autres documents en relief à partir des recommandations des éducateurs spécialisés.

Au sein des établissements publics, on retrouve (1) les professeurs et (2) les auxiliaires de vie scolaire. Les professeurs doivent fournir, en amont de leur diffusion aux élèves, les documents qu'ils vont utiliser en classe, afin de laisser le temps aux transcrip-teurs d'en produire des versions braille. En classe ils doivent adapter leur discours et leurs pratiques aux besoins des élèves déficients visuels, malgré le fait qu'ils ne reçoivent la plupart du temps aucune formation particulière pour ce faire, ce qui nous a été rapporté dans notre étude et ce que corrobore également la littérature [Pieper et al. 2003]. (2) Les auxiliaires de vie scolaire, sur le plan de l'éducation, prennent des notes manuscrites pendant les cours (qui doivent ensuite être relues à l'élève par une aide extérieure), et permettent de répéter les instructions du professeur lorsque c'est nécessaire. À noter que leur présence en cours n'est pas systématique et dépend de leur disponibilité et des capacités des élèves suivis.

En ce qui concerne les élèves déficients visuels, le volume de documents braille qui leur était fourni est très variable, et dépend des méthodes et pratiques de leurs professeurs, des contraintes de disponibilités des transcrip-teurs braille, et de l'existence de transcriptions antérieures des documents demandés. Les élèves peuvent dès lors obtenir des versions braille complètes de leurs manuels scolaires utilisés en cours, ou bien seulement certaines transcriptions d'extraits de leurs manuels ou du cours du professeur. La transcription de documents particuliers se fait en fonction de la demande des professeurs au cours de l'année scolaire.

Chaque élève déficient visuel ayant pris part à cette étude possédait un bloc-notes braille ; tous ont été financés par des fonds publics. Ces appareils étaient utilisés par les élèves pour prendre des notes en cours mais aussi pour fournir leurs productions à leurs professeurs, en les imprimant ou en leur transférant les fichiers correspondants sur clé USB.

## **7.2. Méthode et données d'investigation**

---

Nous avons choisi pour cette étude de nous restreindre à des élèves non-voyants ou avec une malvoyance profonde, afin d'éviter d'observer une trop grande hétérogénéité de matériels,

---

<sup>71</sup> <http://ijatoulouse.org/>



documents et stratégies d'accès à l'information. Nous avons également restreint l'étude aux cours et documents de mathématiques sur la base des retours des éducateurs spécialisés, qui rapportaient que cette discipline était celle posant le plus de problèmes aux élèves.

En ce qui concerne les données recueillies, nous avons analysé conjointement 3 types de données : les entretiens semi-directifs, les observations en situation et enfin l'analyse de documents. Le Tableau 4 présente la nature et le volume de chaque type de données récoltées.

	(1) Entretiens semi-directifs	(2) Observations en situation	(3) Analyses de documents
<b>Description</b>	13 entretiens basés sur un questionnaire <sup>72</sup> ayant durés entre 30min et 1h : - 2 élèves aveugles et 2 présentant une malvoyance profonde <sup>73</sup> (niveaux scolaires entre la 5 <sup>ème</sup> et la 1 <sup>ère</sup> ) - 3 professeurs de mathématiques en établissements publics - 2 transcripateurs braille experts <sup>74</sup> - 4 éducateurs spécialisés experts <sup>5</sup>	3 observations de cours de mathématiques dans des classes de 6 <sup>ème</sup> , 2 <sup>nde</sup> et 1 <sup>ère</sup> . Chaque classe comprenait en moyenne 15 élèves ; les cours ont duré en moyenne 1h30. Deux chercheurs suivaient les cours à deux endroits différents dans les salles, et prenaient des notes à l'aide d'une grille d'analyse <sup>75</sup> .	Corpus de cours de mathématiques et leurs contreparties transcrites en braille. Nous avons également eu accès aux notes prises par un élève pendant les cours.
<b>Données</b>	Entretiens enregistrés et retranscrits textuellement sur 59 pages de texte bruts.	3 grilles d'observations comprenant 120 événements particuliers et 4h45 d'enregistrement audio.	29 pages de cours et exercices et de mathématiques et leurs transcriptions braille. 1 fichier texte contenant les notes personnelles d'un élève déficient visuel.

**Tableau 4 – Description des types de données recueillies pour l'étude en situation**

(1) Le but des entretiens était de recueillir des données qualitatives sur le contexte d'apprentissage des élèves déficients visuels, et d'évaluer les éventuels problèmes et contraintes d'accessibilité rencontrées dans ce contexte. Les entretiens étaient semi-directifs c'est-à-dire que les questions étaient ouvertes ou semi-ouvertes. Les principaux sujets abordés durant les entretiens étaient : les problèmes rencontrés par les élèves déficients visuels et stratégies de compensation mises en place, leur utilisation des technologies d'assistance, les documents et supports de cours utilisés, les différentes aides apportées aux élèves déficients visuels pour apprendre, le processus de transcription des documents de cours, les ajustements réalisés en cours par les professeurs enseignant à des élèves déficients visuels, et enfin les interactions entre les différents acteurs de la scolarité des élèves. Les résultats de ces entretiens ont permis d'identifier des pertes et transformations d'informations impactant l'apprentissage.

(2) Les observations réalisées visaient à identifier les différentes sources d'information à disposition des élèves en cours, afin d'évaluer leur accessibilité, c'est-à-dire l'éventuelle perte d'informations ou l'accès à des informations transformées comparativement à celles auxquelles accèdent les élèves non déficients. La grille d'observation utilisée permettait aux observateurs de garder une trace de chaque événement du cours ayant produit de l'information à destination des élèves (ex. le professeur dit ou écrit quelque chose), et la redondance ou complémentarité des différents blocs

<sup>72</sup> Les questionnaires correspondant sont disponibles en annexe du manuscrit : voir Chapitre V2.1.

<sup>73</sup> D'après la définition de l'Organisation Mondiale de la Santé [WHO - World Health Organization 2015].

<sup>74</sup> Ayant au moins 10 ans de pratique dans leur domaine [Ericsson et al. 2006].

<sup>75</sup> La grille correspondante est disponible en annexe du manuscrit : voir Chapitre V2.2.

d'information relevés (ex. est-ce que les gestes de pointage du professeur au tableau ont été repris à l'oral ?). Les enregistrements audio permettaient de valider les grilles d'observation remplies ; l'utilisation de caméras vidéo était exclue afin de ne pas perturber le déroulement du cours.

(3) Nous avons pu étudier le processus de transcription et les transformations et/ou pertes d'informations suite à ce processus, en comparant les documents papier et leurs contreparties braille produites par des transpositeurs professionnels. Les documents d'origine étaient des extraits de manuels scolaires de mathématiques à destination d'élèves préparant le brevet des Collèges. Enfin, nous avons pu étayer nos résultats à l'aide des notes de prises en cours par un élève déficient visuel sur son bloc-notes braille.

### **7.3. Résultats : accès et accessibilité à l'information pédagogique**

---

Sur la base des données recueillies, nous avons pu identifier des *pertes d'informations*, c'est-à-dire des informations accessibles aux élèves voyants mais pas aux déficients visuels, et des problèmes *d'accessibilité* de l'information (pouvant impliquer des transformations de l'information initiale), ayant des conséquences potentiellement négatives sur l'apprentissage, comme une surcharge cognitive ou une vitesse de traitement diminuée. Dans cette section nous appuierons les résultats présentés à l'aide de groupes de plusieurs (2 à 4) citations extraites des entretiens semi-directifs ; on fera référence à chaque groupe de citation par la mention [CT #] dans le texte (le # sera remplacé par un numéro identifiant). Toutes les citations sont disponibles en annexe Chapitre V2.3.

Dans cette section, on développera les différents problèmes de pertes et d'accessibilité de l'information du point de vue d'un étudiant déficient visuel, selon plusieurs « sources » informationnelles : les documents braille, le discours oral du professeur et le bloc-notes braille. Nous avons étudié avec un intérêt particulier les documents braille étant donné leur importance pour l'apprentissage et étant donné les objectifs de nos travaux quant à la mise en forme des textes.

La Figure 18 ci-dessous montre la complexité de la situation d'apprenants déficients visuels en classe, à travers la multiplicité des différentes sources informationnelles à sa disposition, ainsi qu'à travers les transformations intermédiaires qui leur sont appliquées avant que l'élève puisse y accéder.

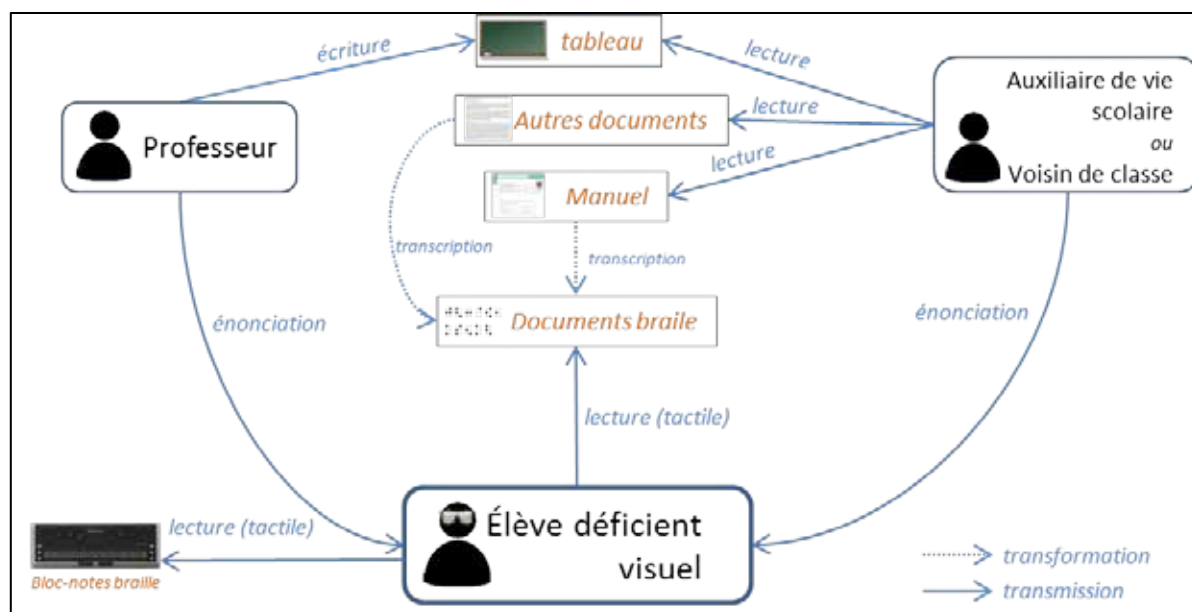


Figure 18 – Schéma résumant les sources informationnelles à disposition des élèves déficients visuels en classe

### 7.3.a. Documents Braille

L'institution spécialisée fournit aux élèves des documents braille embossés sur papier par l'intermédiaire des transcripteurs. Pour les mathématiques, la plupart des transcriptions sont des exercices et moins souvent des leçons. Nous développerons dans les sous-sections suivantes les différentes pertes d'informations et problèmes d'accessibilité relatifs aux documents braille fournis aux élèves déficients visuels. Bien que l'étude porte sur des documents mathématiques, nous sommes restreints à l'accessibilité des textes et de leur structuration, et non l'accessibilité d'objets mathématiques particuliers comme les formules ou les expressions algébriques, ces derniers sortant du cadre défini pour nos travaux.

#### 7.3.a.1. Pertes d'informations dans les documents braille

Tout d'abord, les élèves déficients visuels n'ont *jamais accès à l'intégralité de leurs manuels scolaires* [CT 1], excepté dans de rares cas où une version déjà transcrite de leurs manuels existe ; cela est lié au coût de la transcription de la totalité d'un manuel scolaire, évoqué ci-après dans cette section. En conséquence les élèves n'ont accès qu'à des extraits de leurs manuels.

En ce qui concerne les contenus transcrits, lorsque l'on compare le document d'origine (l'extrait du manuel scolaire) et sa version transcrite en braille, on peut observer que de nombreux objets textuels pourtant signifiants (au sens du MAT) ne sont pas inclus dans la version braille ; dans notre corpus ce fut le cas des renvois, de contrastes typographiques définissant les limites entre objets, des rappels de titres, et de certaines icônes (voir Figure 19). Nous avons retrouvé en moyenne 1 occurrence de contenu non-retranscrit par page (en-têtes et pieds de page exclus).

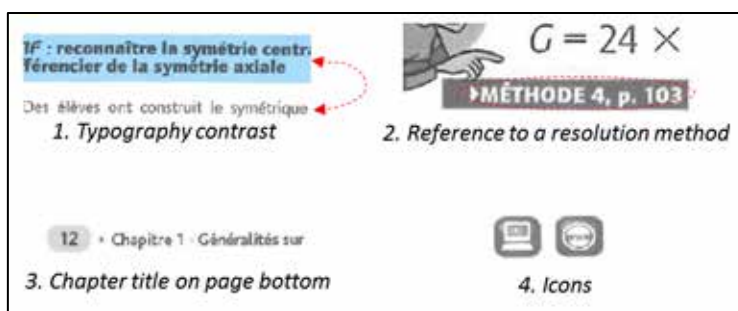


Figure 19 –Exemples d'objets textuels de manuels de mathématiques non retranscrits

De plus, l'organisation spatiale du texte est rarement retranscrite dans les versions braille ; dans le cas de notre corpus aucune disposition spatiale comme sur la Figure 20 (porteuse de sens d'après le MAT) n'était retranscrite.

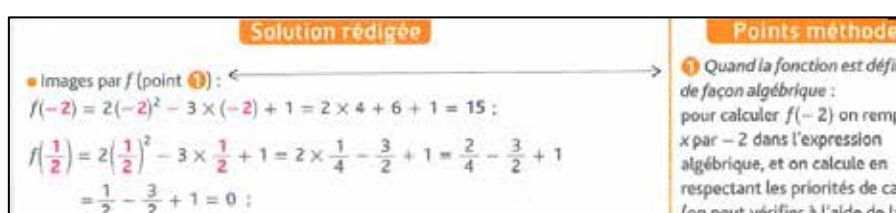


Figure 20 –Exemple de disposition spatiale de texte non retranscrite

En ce qui concerne les graphiques, figures et images, un équivalent en relief n'est produit que sur demande du professeur ; un document annexe est alors produit et référencé dans le texte en braille par un identifiant. Dans le cas de notre corpus, lorsqu'un document en relief n'était pas produit, aucune description verbale de ces objets n'était incluse en remplacement.

Plusieurs facteurs ont été évoqués dans les entretiens en ce qui concerne la non exhaustivité des transcriptions. En effet, transcrire un document en braille est un *processus long* [CT 2] et les transpositeurs sont soumis à une *charge de travail importante* [CT 3], et sont obligés de limiter le temps qu'ils peuvent allouer à chaque transcription. De plus, seule une minorité des élèves déficients visuels étant de bons brailleux [CT 11], et la vitesse de lecture braille étant de toute façon inférieure à celle d'une lecture visuelle silencieuse, ils ne peuvent pas *traiter autant de contenu* [CT 4] que les élèves voyants. Pour ces raisons, seuls les extraits des manuels considérés comme essentiels sont transcrits ; ces extraits sont épurés avant transcription, et certains des objets du texte comme les icônes, les références croisées, les phénomènes typographiques et d'organisation spatiale sont supprimés, sans en fournir d'équivalents textuels. La typographie est d'ailleurs considérée par les transpositeurs comme étant un *ornement du texte à visée esthétique* [CT 5]<sup>76</sup> ; la non retranscription des phénomènes spatiaux serait, elle, liée au fait que les brailleux n'apprécient pas les *interruptions dans leur fil de lecture* [CT 6]. Cette dernière affirmation est cependant à nuancer : un éducateur spécialisé nous a en effet rapporté que les bons brailleux utilisent les indentations et les autres dispositions des lignes braille pour *lire en diagonale* (parcourir) les textes [CT 16].

<sup>76</sup> Il faut comprendre cette affirmation dans le sens où la couleur et la disposition du texte représentent pour les transpositeurs des informations complètement superflues en soi ; cependant en tant que lecteurs experts, les transpositeurs interprètent dans une certaine mesure les architectures textuelles exprimées par ces marques, et retranscrivent une partie de leur sémantique. Toutefois, il est probable que cet *à-priori* limite leur interprétation de la MFM aux objets qui leur sont familiers, laissant ainsi de côté les autres phénomènes porteurs de sens.

### 7.3.a.2. Problèmes d'accessibilités de l'information en braille

---

En marge des différentes pertes d'informations évoquées, nous avons dégagé 3 grandes catégories de facteurs influençant l'accessibilité des documents braille pour les élèves déficients visuels : (1) les facteurs relatif au mode et aux modalités d'accès, (2) les facteurs liés au contenu et enfin (3) les facteurs liés au lecteur braille.

(1) Comme évoqué précédemment (voir partie 3), le braille impose un accès séquentiel à l'information. La construction d'une *représentation globale du contenu* est donc coûteuse cognitivement parlant [CT 7], et pourtant primordiale pour la compréhension chez les voyants d'après le modèle de [Van Dijk and Kintsch 1983]. De plus, la faible vitesse de lecture braille implique de maintenir plus longtemps le texte lu en mémoire ; ainsi, si plus de ressources cognitives sont allouées pour accéder à l'information sur son support, moins de ressources pourront être allouées à l'apprentissage (i.e. compréhension et intégration de l'information et acquisition de schémas de connaissances transférés en mémoire à long-terme [Sweller et al. 2011]). Ce dernier point sur la lenteur de l'accès à l'information en braille rejoint une problématique plus large de la temporalité de l'accès à l'information dans des modes non-visuels, qui rend toute tâche d'apprentissage plus lente qu'en visuel ; cela est illustré par un comportement observé chez un élève déficient visuel, qui en amont du cours recopiait l'énoncé des exercices qui allaient être abordés en cours, et les formulations des réponses, afin de ne plus avoir qu'à inscrire les résultats dans le fichier. Sur un plan plus pratique, les documents braille prennent « jusqu'à 10 fois plus de volume de papier<sup>77</sup> » pour le même contenu que les imprimés classiques, à cause de la surface prise par un caractère braille (elle-même faisant que les textes braille sont couramment embossés sur des feuilles au format plus grands que le classique format A4, afin d'assurer une certaine fluidité de lecture) et de l'épaisseur du papier nécessaire pour embosser les caractères. Les élèves sont donc contraints de transporter des *documents lourds* [CT 8] et doivent faire preuve d'*organisation* [CT 9] pour être capables de les utiliser efficacement et ne pas les perdre ou les mélanger.

(2) Dans notre corpus de 29 pages de manuels scolaires (71 en braille), nous avons dénombré 13 erreurs de transcription (orthographe, expressions mathématiques, numérotation, etc.), en plus des objets textuels non retranscrits (voir 7.3.a.1). Nous avons également noté que les conventions de présentation du texte braille (ex. numérotation, indentation) n'étaient pas homogènes entre transcripateurs sur des extraits différents, et parfois-même ne l'étaient pas sur un même extrait (transcrit par une seule personne). Le MAT souligne pourtant que c'est l'identité entre les marques de mise en forme du texte qui font qu'elles peuvent être interprétées correctement ; de plus, des normes de transcription existent [Ministère des Affaires Sociales and Commission Evolution du Braille Français 1999]. Par exemple l'utilisation des indentations, des retours à la ligne, de la ponctuation et de la numérotation des items d'énumérations et de pages n'étaient pas homogènes. Le fait que les documents braille de chaque élève déficient visuel soient transcrits par plusieurs personnes différentes accentue potentiellement cette hétérogénéité et donc la difficulté pour l'élève de déchiffrer le contenu. Enfin, un dernier point relatif aux contenus braille est qu'ils ne sont accessibles à priori qu'à l'élève déficient visuel suivant le cours ; les politiques actuelles veulent qu'il n'y ait au maximum qu'un élève déficient par classe. Or le fait que l'élève soit le seul à pouvoir accéder à la

---

<sup>77</sup> Ce chiffre nous a été rapporté par un éducateur spécialisé ; on peut citer à titre d'exemple les 29 pages de manuels scolaires étudiées ici représentaient 71 pages braille.

version du cours qu'il possède<sup>78</sup>, qui de plus diffère forcément (au regard du processus de transcription évoqué ci-avant) du cours des autres élèves, l'oblige à constamment évaluer la correspondance entre ses contenus braille et les références aux manuels et au cours du professeur faites en cours.

(3) Enfin, l'expertise des élèves dans le domaine de la lecture braille ainsi que leur connaissance du code braille pour les mathématiques<sup>79</sup> va grandement *contraindre leur accès à l'information* [CT 10]. La plupart des élèves déficients visuels ont apparemment des compétences limitées en lecture braille [CT 11], et leurs compétences limitent leur capacité à se représenter la structure globale des documents [CT 12].

### **7.3.b. Discours oral du professeur**

---

Lors d'un cours, le discours oral du professeur est la seule source dynamique d'information disponible pour les élèves déficients visuels. Leur discours comprend de nombreuses références au contenu du tableau et du manuel, étant donné que leur discours est justement construit autour de ces supports. Des références implicites aux contenus de ces supports (des *déictiques*) comme « ici », « là », ou « celui-ci », souvent accompagnés d'un geste correspondant, ne peuvent être interprétés par les élèves qui n'ont pas accès au contexte visuel du discours oral. Les déictiques liés au contexte visuel sont très fréquents pendant les cours de mathématiques étant donné la nature spatiale des objets du domaine (ex. « en-dessous de cette courbe », « le membre de gauche », « l'opérateur ici »). Les observations réalisées ont recensé entre 5 et 10 utilisations de déictiques (non-accessible sans contexte visuel) par heure (respectivement 5, 6 et 10 fois par heure pour chacune des 3 observations).

L'accès à l'information orale est non-seulement séquentiel, comme pour le braille, mais l'information est également fugace, non persistante, ce qui en soi rend le traitement de l'information orale coûteuse cognitivement parlant d'après le *transient information effect* [Leahy and Sweller 2011]. Afin de ne pas perdre d'information, les élèves déficients visuels doivent maintenir une attention constante vers le discours du professeur, ce qui peut être une source de fatigue mentale. Nous avons d'ailleurs observé que les élèves déficients interagissaient en moyenne 4 fois par heure avec leurs voisins de classe et 5 fois par heure avec le professeur, posant des questions ou demandant de répéter les informations énoncées. Enfin, l'accès au discours du professeur peut être perturbé par le bruit ambiant de la classe [CT 13] ; ce peut être également le cas avec le discours oral de l'auxiliaire de vie scolaire qui va rentrer en concurrence attentionnelle avec le discours du professeur. Enfin, d'après l'architecture de la mémoire de travail (voir partie 3) utilisée dans les apprentissages, la boucle phonologique des élèves déficients visuels sera sollicitée non seulement pour le traitement des différents discours oraux mais également pour le traitement du braille ; l'information n'étant accessible que par ce canal, la charge cognitive sera plus importante que si elle était répartie en plus sur le calepin visuo-spatial. Cela peut expliquer en partie la « *fatigue constante des élèves, qui doivent en permanence être attentifs et appliquer des procédures pour manger, s'habiller, retrouver leurs cours...* », rapportée par un éducateur spécialisé.

---

<sup>78</sup> Sauf cas où l'auxiliaire de vie scolaire est présente et lit le braille, ce qui est complètement hypothétique étant donné que cela ne leur est pas demandé, et qu'ils suivent des élèves avec différents types de déficience.

<sup>79</sup> Les caractères braille à 6 points ne permettant que de coder 64 caractères, les mathématiques (opérateurs, symboles, etc.) sont transcrites en braille dans un code dédié, différent de celui utilisé pour l'alphabet latin et les chiffres arabes.

### **7.3.c. Tableau noir**

---

On distingue 3 cas différents d'utilisation du tableau en cours par le professeur pour l'accessibilité du contenu qui y est inscrit. Le premier est le cas où le contenu écrit au tableau est complètement redondant avec le discours oral du professeur, par exemple lorsque le professeur lit ce qu'il est en train d'écrire (ce qui fut le cas les deux tiers du temps lors de nos observations). Dans ce cas les élèves déficients visuels ont effectivement accès à l'information, mais pas de manière redondante : ils doivent donc être plus attentifs à l'information orale étant donné qu'ils ne peuvent pas s'appuyer sur le contenu écrit au tableau pour ré-inspecter le contenu, qui doit donc être interprété et représenté en une seule fois (étant donné la fugacité de l'information orale).

Dans le second cas, le discours oral du professeur n'est que partiellement redondant avec le contenu du tableau. À noter que durant les 4h30 d'observations nous n'avons relevé que 8 occurrences d'informations pédagogiques (destinées à être comprises et apprises) complètement indépendantes du contenu du tableau, qui sont donc à priori aussi accessibles aux élèves déficients visuels qu'aux autres. Le reste du temps, le discours du professeur inclut des références au contenu du tableau à travers des déictiques et/ou des gestes. Il arrive également que certains contenus écrits au tableau ne soient pas du tout repris à l'oral : nous avons observé qu'en moyenne 4 dessins, figures, graphes ou équations par heure étaient inscrits au tableau sans contrepartie orale. Enfin, les marques de MFM utilisées par le professeur sur le texte écrit au tableau (soulignement, mise en vis-à-vis ou dispositions particulières d'éléments, encadrement de texte, utilisation de couleurs, etc.) n'étaient que rarement reprises à l'oral (ex. pour le soulignement, préciser que le concept souligné est important) ; nous n'avons cependant pas pu quantifier ce dernier point étant donné l'absence de photo et vidéos des cours.

### **7.3.d. Bloc-notes braille**

---

Chaque élève déficient visuel impliqué dans l'étude possédait un bloc note braille. Ces ordinateurs portables « braille » leurs permettent de prendre des notes personnelles en cours et de produire leurs exercices, devoirs et comptes rendus afin de pouvoir les communiquer à leur professeur, en les imprimant ou en les transférant sur clé USB. Cependant, dans le cas de contenus mathématiques les élèves doivent utiliser un *encodage spécial* pour les caractères mathématiques afin qu'ils soient imprimés correctement [CT 14] ; cet encodage spécial a l'inconvénient de générer des problèmes pour le transfert des fichiers textes sur clé USB qui deviennent illisibles.

Il nous a été rapporté que les blocs-notes braille tombaient régulièrement en panne et généraient de nombreuses erreurs [CT 15]. Pendant l'une des 3 observations réalisées un élève déficient visuel a d'ailleurs été confronté à l'impossibilité d'utiliser son bloc-notes braille. Évaluer la fiabilité et les différents problèmes liés aux blocs-notes braille semble un point important pour de futures études, étant donné qu'ils représentent l'outil principal utilisé par les déficients visuels pour communiquer leurs travaux à leurs professeurs et ainsi pouvoir être évalués et suivis dans leur scolarité.

## **7.4. Récapitulatif des problèmes de disponibilité et d'accessibilité de l'information**

---

Le tableau ci-dessous récapitule les différents problèmes de disponibilité et d'accessibilité de l'information pédagogique pour les élèves déficients visuels.

## 7. Étude en situation : accès et accessibilité de l'information pédagogique pour les élèves déficients visuels

	<b>Documents braille</b>	<b>Discours oral du professeur</b>	<b>Autres</b>
<b>Problèmes de disponibilité</b>	<ul style="list-style-type: none"> <li>- documents non-retranscrits ou partiellement retranscrits</li> <li>- icônes, renvois, marques de MFM typographiques et spatiales, notes en marge et notes de bas de page non retranscrites</li> </ul>	<ul style="list-style-type: none"> <li>- inaccessibilité à cause du bruit ambiant de la classe ou du discours concurrent de l'auxiliaire de vie scolaire</li> </ul>	<ul style="list-style-type: none"> <li>- gestuelle du professeur</li> <li>- contenus écrits sur le tableau non repris à l'oral (dessins, figures et équations)</li> </ul>
<b>Problèmes d'accessibilité</b>	<ul style="list-style-type: none"> <li>- faibles compétences en lecture braille et/ou en exploration tactile</li> <li>- mauvaise connaissance du code braille mathématique</li> <li>- lenteur de la lecture braille en général</li> <li>- incohérences dans les conventions de transcription braille</li> <li>- erreurs de transcription</li> <li>- les documents brailles sont des versions remaniées des contenus pédagogiques, qui diffèrent de ceux auxquels accèdent les autres élèves et le professeur</li> <li>- volume des documents</li> </ul>	<ul style="list-style-type: none"> <li>- références déictiques au contenu écrit du tableau ou du manuel</li> <li>- concurrence avec les discours oraux des camarades de classe ou de l'auxiliaire de vie scolaire</li> <li>- séquentialité et nature fugace du discours oral nécessitant une attention constante</li> </ul>	<ul style="list-style-type: none"> <li>- pannes et erreurs des blocs-notes braille</li> <li>- contenus écrits sur le tableau partiellement repris à l'oral</li> </ul>

**Tableau 5 – Résumé des problèmes de disponibilité et d'accessibilité de l'information pédagogique**

Il est important de noter que les problèmes présentés peuvent survenir simultanément pendant un cours. Par exemple, pendant les observations un élève déficient visuel a dû gérer une panne de son bloc-notes braille alors qu'il essayait de comprendre le discours oral du professeur qui incluait des déictiques. Les accumulations potentielles de ces problèmes peuvent générer des situations de surcharge cognitive et attentionnelle pour les élèves.

De manière générale le Tableau 5 montre la complexité de l'accès à l'information pour les élèves déficients visuels : multiplicité des sources d'informations, efforts cognitifs pour y accéder et enfin la transformation des informations comparativement à celles auxquelles accèdent les élèves voyants.

### 7.5. Discussion

Tout d'abord, il est important de rappeler que l'accès à l'information chez les déficients visuels est contraint par la séquentialité des modes audio et tactile et par la fugacité du premier. Ces contraintes fondamentales, comparativement au mode visuel, font que les élèves déficients vont être soumis à des situations d'apprentissage imposant des exigences attentionnelles fortes, avec des risques fréquents de surcharge cognitive. Cela rend donc les élèves déficients visuels plus enclins à la surcharge cognitive, ce qui pourrait expliquer pourquoi un éducateur spécialisé rapportait que les élèves subissent une fatigue mentale constante. En effet, plusieurs problèmes relevés peuvent amener une charge cognitive supplémentaire. Tout d'abord, le fait que l'information soit répartie sur plusieurs sources différentes (documents braille, document électroniques sur le bloc-notes braille, et discours oraux du professeur, de l'auxiliaire de vie scolaire et parfois des camarades) peut amener à des représentations mentales éparses et non reliées, induisant une charge cognitive supplémentaire pour l'intégration des informations. Cela est d'autant plus prégnant que les élèves déficients visuels n'ont pas accès à un référentiel unique et partagé auquel rattacher les informations entrantes, comme le font les élèves voyants avec le cours de leur professeur, leur manuel ou encore le tableau. Enfin, de nombreux facteurs influencent l'accessibilité de l'information et peuvent pénaliser la compréhension et augmenter la charge cognitive ; par exemple le bruit ambiant de la salle de classe,



des difficultés de l'élève en lecture braille, des contenus écrits au tableau non repris à l'oral ou encore l'emploi de déictiques et de gestes par le professeur pour appuyer son discours (voir Tableau 5 pour la liste exhaustive).

Ces problèmes d'accès à l'information et de surcharge cognitive sont illustrés par deux comportements d'élèves relevés pendant les observations en cours : l'un arrêta rapidement de suivre le cours et tombait dans l'inaction dès qu'il n'arrivait plus à suivre les informations entrantes, tandis que l'autre posait de nombreuses questions au professeur. Ces deux comportements, d'abandon et de maintien de l'attention, pourraient être symptomatiques d'une difficulté constante à intégrer et lier entre elles les informations entrantes.

En ce qui concerne les contenus pédagogiques fournis aux élèves déficients visuels, un premier constat est qu'ils ont accès à beaucoup moins de contenus que les élèves voyants. Cela empêche de potentiels créations de liens et référencements croisés entre informations dont on sait qu'ils jouent un rôle important dans l'apprentissage [Wiley et al. 2009]. De plus, les élèves ont accès à très peu de documents en relief (ex. cartes géographiques, figures) qui sont encore plus coûteux que le braille à produire ; cela limite les bénéfices potentiels de l'apprentissage à partir de plusieurs médias [Pring and Rusted 1985]. Or, un cours *multimédia* (et donc multimodes pour reprendre le terme utilisé dans ce manuscrit) bien conçu, suivant par exemple les recommandations de [Mayer 2009], sera plus efficace qu'un cours avec le même contenu véhiculé dans un seul mode. Cependant, fournir autant de contenus (braille par ex.) aux élèves déficients visuels qu'aux autres pourrait s'avérer en réalité être pénalisant pour eux. En effet les élèves pourraient ne pas être capables d'intégrer un volume de contenu plus important [CT 4] étant donné la lenteur de la lecture braille, et surtout en considérant l'absence d'un « filtrage » de l'information à l'oral et en tactile qui permet justement en lecture visuelle une sélection rapide et efficace des contenus à traiter, en se basant notamment sur leur mise en forme. Ainsi il faudrait permettre aux déficients visuels de traiter plus efficacement les contenus afin qu'ils en retirent d'autres bénéfices notamment pour l'apprentissage (en leur permettant de traiter plus de contenu). Or, la MFM joue un rôle de facilitateur de traitement du contenu textuel [Lemarié 2006a; Virbel et al. 2005] et permet aussi au lecteur de sélectionner les informations à traiter ; une restitution optimale de la MFM dans des modes non-visuels pourrait donc avoir de nombreux effets positifs. Ce dernier point est également appuyé par les différents manques dans la transcription braille de la sémantique de la MFM relevés dans notre corpus, qui sont potentiellement pénalisants pour la compréhension du contenu.

Enfin, étant donné le large spectre de problèmes de disponibilité et d'accessibilité de l'information, leurs différentes origines et interactions, il semble important d'adopter une approche globale (holistique) pour la conception de technologies d'assistance destinées à être utilisées dans un contexte d'apprentissage. Cette étude offre justement une vue d'ensemble des différentes contraintes qui devraient être prises en compte dans la conception de technologies d'assistance, que ce soient des contraintes cognitives, pratiques (ex. liées aux technologies existantes) ou encore liées aux institutions et leur fonctionnement. Étant donné la nature exploratoire de l'étude et la nature qualitative des données récoltées, des investigations futures plus systématiques seront nécessaires afin de valider les différents résultats évoqués ici. De plus, le caractère exploratoire et assez ciblé de cette étude, limite la portée des résultats : ils gagneraient en validité s'ils étaient confirmés par une étude plus systématique, inspectant un échantillon plus large et hétérogène.

## 8. Conclusion et problématique générale

---

L'accessibilité numérique peut être vue comme comportant deux niveaux : un premier niveau de disponibilité de l'information, et un second niveau d'utilisabilité des dispositifs physiques et logiciels médiateurs de cet accès. Nos travaux visent à contribuer à ces deux niveaux, étant donné qu'ils visent dans un premier temps à formaliser les architectures textuelles pour les rendre disponibles, et dans un second temps à étudier leur restitution et son impact.

Malgré les contraintes des modes non-visuels de transmission de l'information, qu'il faut prendre en compte dans la conception des interactions, nous avons vu qu'il était possible d'utiliser l'expressivité de ces modes à travers différentes modalités, comme les icônes audio ou la spatialisation du son, afin de diminuer les contraintes des modes et optimiser l'accès à l'information. Les contraintes dans l'accès à l'information numérique se situent également au niveau des contenus, dont la production relativement permissive a dû être normée afin d'en assurer un certain niveau d'accessibilité. Ces normes sont assez peu appliquées à cause des contraintes de développement qu'elles impliquent, et ne permettent pas de résoudre tous les problèmes rencontrés par les utilisateurs déficients, même si elles représentent un premier pas important. Les normes et les technologies de lecture ne prennent que très peu en compte la MFM, pourtant porteuses de sens ; les nouvelles normes d'accessibilité ARIA viennent cependant nuancer ce propos en introduisant un marquage sémantique basé (implicitement) sur les architectures textuelles. Enfin, il semble que les recherches autour de l'accessibilité numérique soient centrées sur les contenus, et laissent quelque peu de côté l'utilisabilité et le mode de fonctionnement des technologies de lecture actuelles.

Nos travaux s'inscrivent dans ce cadre comme une extension des normes d'accessibilité, en proposant une méthodologie de formalisation des architectures textuelles qui pourrait ensuite être intégrée dans les standards existants, et dont l'annotation pourrait se faire de manière semi-automatique (voir Chapitre III). Sur un autre plan, l'analyse de l'espace de conception pour la restitution des marques de MFM que nous avons menée (voir Chapitre IV) est généralisable à d'autres contextes d'interactions non-visuelles. Enfin, nos résultats plutôt prometteurs sur l'évaluation de cette restitution (voir Chapitre V) pourraient être réutilisés afin d'améliorer l'utilisabilité des technologies de lecture actuelles dans le cadre de la compréhension des contenus pour les déficients visuels. L'étude en situation réalisée a notamment montré comment faciliter la compréhension et le traitement des textes pourrait être bénéfique à des situations concrètes d'apprentissage pour les déficients visuels.

Pour terminer, au-delà du cas des utilisateurs déficients visuels, ces travaux peuvent avoir des répercussions dans de nombreux domaines comme la transcription braille ou la structuration visuelle de documents pédagogiques, ou encore pour des situations où les utilisateurs non-déficients sont dans l'incapacité d'accéder visuellement au document ; par exemple parce qu'ils doivent effectuer une tâche en parallèle qui demande leur attention visuelle, ou encore que l'ensemble des informations qui les intéressent sont réparties sur plusieurs pages différentes et éloignées dans le document, ou tout simplement quand un contenu ne tient pas sur une seule page d'écran. D'une manière générale, on retrouve cette idée dans la littérature : améliorer l'accessibilité d'un objet pour une population spécifique a très souvent des effets bénéfiques et des répercussions sur l'utilisation de ce même objet pour la plupart des utilisateurs [Persson et al. 2014].

# Chapitre III - Méthodologie de formalisation de la Mise en Forme Matérielle

Résumé : Ce chapitre présente un ensemble de méthodes visant à formaliser la sémantique véhiculée par la mise en forme matérielle (MFM) des documents, c'est-à-dire leurs architectures textuelles. Nous avons regroupé ces méthodes au sein d'un processus comprenant d'abord une tâche de segmentation des objets textuels, puis deux tâches d'annotation (i.e. leur extraction puis incorporation au document) des informations correspondant aux architectures textuelles. Les méthodes créées l'ont été dans une optique d'automatisation, et à cette fin elles ont été voulues les plus formelles possibles ; une première implémentation a été effectuée, montrant la faisabilité d'une telle approche. La notion de *convention* de MFM, qui désigne des usages partagés et codifiés des marques de MFM, est ici centrale et l'on montrera comment la modélisation d'architectures textuelles conventionnées permettra dans le futur une formalisation de plus en plus automatique. Ce chapitre présente d'abord le socle théorique de ces travaux, puis les 3 étapes du processus de formalisation, résumées dans la dernière partie qui traitera également des perspectives possibles de ces travaux.

## Sommaire du Chapitre III

<b>1. Introduction .....</b>	<b>92</b>
<b>2. Socle théorique pour la formalisation de la MFM.....</b>	<b>93</b>
2.1 Les différents niveaux de structuration des textes écrits.....	93
2.2 Analyse automatique du contenu des documents.....	95
2.3 Des bases théoriques du MAT vers une formalisation de la MFM.....	96
<b>3. Segmentation des objets textuels des documents.....</b>	<b>98</b>
3.1 Segmentation visuelle : une première approche .....	99
3.2 Le cas d'objets textuels imbriqués .....	101
3.3 Segmentation d'objets textuels sur la base d'indices lexico-syntaxiques .....	102
3.4 Exemple de segmentation .....	104
3.5 Conclusion pour nos travaux.....	105
<b>4. Définition et identification des conventions partagées de MFM .....</b>	<b>105</b>
4.1 De l'existence de conventions de MFM partagées.....	105
4.2 Modélisation des conventions de MFM partagées universellement .....	107
4.3 Exemple de formalisation des architectures textuelles d'un document sur la base des conventions partagées de MFM .....	114
4.4 Conclusion .....	115
<b>5. Modélisation des conventions locales de MFM .....</b>	<b>115</b>
5.1 Méthode d'identification des conventions locales .....	116
5.2 Exemple d'annotation des conventions locales .....	118
5.3 Limites de l'approche .....	118
<b>6. Vue d'ensemble du processus d'annotation et opérationnalisation technique .....</b>	<b>119</b>
6.1 Schéma global .....	119

6.2 Implémentation informatique du processus de formalisation.....	120
<b>7. Vers une formalisation automatique de la MFM par apprentissage automatique.....</b>	<b>123</b>
<b>8. Conclusion.....</b>	<b>126</b>

## 1. Introduction

---

La Mise en Forme Matérielle (MFM) des textes correspond, d'après le Modèle d'Architectures Textuelles (MAT [Virbel 1985; Virbel 1989], voir Chapitre I), à un ensemble de marques typographiques, dispositionnelles ou lexico-syntaxiques incorporées au texte, exprimant les intentions *architecturantes* de l'auteur quant au contenu de son texte. Ces marques permettent de distinguer au sein des textes des *objets textuels*; l'ensemble des objets textuels et leurs relations formant des *architectures textuelles*. La mise en forme des textes n'est donc pas un simple embellissement mais bien un vecteur de sémantique permettant d'organiser et structurer le contenu des textes, et possède en plus une dimension fonctionnelle pour la lecture visuelle (qui sera abordée dans le chapitre suivant). L'étude de l'existant atteste de l'importance de la MFM en lecture visuelle, notamment pour la compréhension du contenu, mais aussi de la prise en compte très superficielle de la MFM par les technologies d'assistance actuelles proposant un accès non-visuel aux textes. Afin de pouvoir fournir des équivalents non-visuels à la sémantique des architectures textuelles, pertinents pour l'accès aux documents écrits par les déficients visuels, il était nécessaire en amont de pouvoir les formaliser. Le terme *formalisation* fait référence au fait de rendre formelles, explicites, les informations véhiculées par les marques de MFM (i.e. les architectures textuelles) qui relèvent pendant la lecture d'une interprétation (subjective) du lecteur.

Ce chapitre constitue une proposition de méthodologie, permettant de formaliser la mise en forme matérielle des documents afin de pouvoir ensuite mettre en place des stratégies de restitution des architectures textuelles ainsi extraites, puis enfin d'évaluer cette restitution. Il est donc question ici de rendre les architectures textuelles *disponibles*, explicites. Le MAT postule que la MFM d'un texte est équivalente, en termes informatifs, au métatexte explicitant les architectures textuelles correspondantes aux marques de MFM du texte ; la MFM peut donc être explicitée sous forme discursive, par un ensemble de métaphrases. Cependant, au-delà de ce *postulat*, le *processus* de production du métatexte, et donc d'interprétation des marques de MFM explicitant les architectures textuelles sous-jacentes, n'est pas développé par le MAT. De plus, aucun des travaux s'appuyant sur le MAT ne s'est penché sur cet aspect de manière formelle. Ainsi nous tenterons dans ce chapitre de répondre aux questions suivantes, en définissant un processus de formalisation : quelles informations véhicule la MFM ? Comment les extraire de façon automatique ? Comment les représenter de la manière la plus formelle possible ?

La MFM peut être étudiée de différents points de vue comme nous le verrons dans la partie 2, mais à ce jour aucune méthode ne permettait de formaliser la MFM dans le cadre théorique du modèle d'architecture textuelle (MAT). L'objectif était à la fois de produire un ensemble cohérent réunissant certaines techniques et travaux existant, et de faire en sorte que cette méthodologie soit la plus automatisable possible afin de rendre la MFM la plus accessible possible dans le futur. Nous avons proposé une implémentation de cette méthodologie, rapportée dans la partie 6 de ce chapitre, où le processus de formalisation est réalisé de manière automatique à l'aide de techniques de *machine learning* [Fauconnier et al. 2014]. Le terme *annotation* sera fréquemment employé dans ce chapitre et désigne ici le fait d'incorporer au fichier du document d'origine le métatexte explicitant sa MFM ; en d'autres termes l'annotation représente l'ajout de données au code source du document afin de qualifier ses architectures textuelles (le processus de formalisation est donc un processus d'annotation).

La segmentation constitue la première étape du processus de formalisation de la MFM, et consiste à identifier les différentes unités contrastées visuellement dans un document écrit : les objets textuels. Les deux étapes suivantes décrites dans ce chapitre vont permettre l'extraction des informations de MFM relatives aux objets identifiés. Ces étapes reposent sur l'identification de ce qu'on l'on appellera ici des *conventions* de MFM ; ce mot revêt une importance particulière puisqu'il est au cœur de l'utilisation de la MFM dans le discours écrit. Dans le Chapitre I nous avons vu que les objets textuels étaient définis de manière relative, par contraste avec le reste du texte, et par identification (ou identité) des objets entre eux. Cela relève bien d'un usage conventionné de la MFM, c'est-à-dire respectant des conventions qui peuvent être locales au document (ex. un auteur peut mettre certains mots clés en vert) ou alors partagées de manière quasi-universelle (ex. les titres). Si l'auteur ne fait pas un usage conventionné des marques MFM alors elle perd tout son sens ; par exemple s'il emphase des concepts importants parfois en vert, parfois en les soulignant et d'autres fois encore en italique, il est très probable que sans une légende explicitant l'utilisation de ces marques de mises en forme, le lecteur ne puisse les interpréter. Il est probable que la majeure partie du message de l'auteur véhiculé par le contenu primaire puisse tout de même être interprété, mais pas la signification de la MFM puisque l'usage des marques est incohérent. Ainsi les étapes suivantes du processus d'annotation consistent à identifier les *conventions partagées* (dont une liste des plus courantes a été établie) et les *conventions locales* aux documents.

Pour illustrer les différentes étapes de la formalisation nous utiliserons un extrait d'un manuel de mathématiques, les manuels scolaires présentant une mise en forme particulièrement riche.

## 2. Socle théorique pour la formalisation de la MFM

---

---

Cette partie présente uniquement l'existant en ce qui concerne l'extraction des architectures textuelles, pour introduire la méthodologie adoptée dans le cadre du Modèle d'Architecture Textuelle (MAT) ; le MAT et la théorie de la genèse des marques de MFM ont été présentés dans le Chapitre I.

### 2.1. Les différents niveaux de structuration des textes écrits

---

La problématique de formalisation de la MFM rejoint les domaines du traitement automatique des langues et de l'ingénierie du document, qui visent notamment à identifier, classer et comparer les différents éléments structurels que peuvent inclure les documents textuels. À noter que dans ce cadre (et plus généralement dans ce chapitre) nous ne considérerons que le texte des documents, laissant de côté les autres objets qui peuvent y apparaître (ex. images, figures). La littérature distingue donc 3 niveaux de structuration qui peuvent être analysés [Power et al. 2003; Bateman 2008]:

- § La structure physique/visuelle : elle correspond à la forme visuelle dans laquelle le document apparaît, sans prendre en compte la sémantique du discours ou des marques de MFM. On considère à ce niveau d'analyse le texte comme étant une image, sur laquelle on distingue plusieurs zones aux propriétés communes (typographie, disposition). Cette analyse vise donc à extraire les valeurs des propriétés du texte (couleurs, polices, coordonnées des zones, etc.) et à identifier différentes zones aux propriétés communes (regroupements, au sens des lois

du regroupement visuel du *Gestalt*, voir 3.1), généralement dans un but de numérisation du document ou dans le but ultérieur d'analyser d'autres structures comme celles présentées ci-après.

- § La structure discursive : elle organise le message de l'auteur (écrit en langage naturel) en un ensemble de propositions (ou plutôt unités de discours) reliées entre elles par des relations rhétoriques [Mann and Thompson 1987] (ex. une relation de condition entre une proposition conditionnelle et une proposition résultante). De nombreuses approches étudiant la structure discursive existent, celle-ci étant centrale dans le traitement automatique des langues et en linguistique [Lascarides and Asher 1991; Charolles 1997].
- § La structure logique : de nombreuses définitions existent [Liang et al. 1996; Power et al. 2003] mais s'accordent toutes sur l'existence d'unités logiques reposant sur la structure visuelle du document ; en clair des unités de texte peuvent être regroupées ou scindées puis typées, et participent alors à l'élaboration de la structure logique du document. Le lien fort avec la *structure visuelle* est illustré par le fait que les termes structure visuelle et structure logique sont souvent utilisés de manière interchangeable dans la littérature, et parfois tous deux fusionnés sous l'appellation *layout* [Antonacopoulos et al. 2011]. Les principaux types attribuables aux unités logiques sont : le titre, le paragraphe, la section, l'énumération et l'item d'énumération, ainsi que d'autres types plus spécifiques comme les tableaux ou notes de bas de page ; on remarquera que ces types d'objets logiques correspondant aux vocabulaire d'objets utilisables dans des langages comme LATEX ou HTML [Power et al. 2003]. Cette liste n'est cependant pas fixe et peut être étendue en fonction du type de document étudié et des objectifs de l'analyse. Ainsi l'identification des unités logiques permet, sur la base des niveaux hiérarchiques des titres, de définir un arbre hiérarchique des constituants du document [Fauconnier et al. 2014].

Ces 3 niveaux sont bien sûr liés et se recoupent. Comme évoqué précédemment, la structure logique et la structure visuelle peuvent être vues comme étant confondues si l'on considère que chaque unité visuelle peut être typée par un label logique ; cet ensemble logico-visuel sert les fins du *discours* pour compléter sa structuration rhétorique et la mettre en valeur. Par exemple une énumération peut se trouver être sous la forme d'un paragraphe comprenant des indices lexicaux comme « premièrement, deuxièmement, ensuite... », ou bien être sous une forme visuellement structurée qui marquera l'ordre des items à l'aide d'indices comme les puces et les indentations. [Luc 2000] a d'ailleurs proposé un modèle unifiant structures logico-visuelles (dans le cadre du MAT) et structures rhétoriques. Le MAT souligne ainsi le lien entre discours et structures logico-visuelles en montrant que ces dernières sont issues d'un processus de transformation de métaphrases architecturales.

On pourrait qualifier la théorie du MAT comme agissant au niveau de la structure logique des documents, en donnant un cadre pour un typage et une modélisation fine des unités de texte « visuelles », et de leur mise en relation. Là où la structure logique telle qu'elle a été présentée et telle qu'elle est étudiée dans certains domaines ne considère que certains types d'objets logico-visuels comme les titres ou les énumérations, et ce dans leur participation à une organisation hiérarchique du document, le MAT permet d'étendre ce typage à n'importe quel phénomène de MFM existant et de modéliser n'importe quel type de relation entre objets, ce qui peut être vu comme la *structure logique profonde* des documents [Fauconnier et al. 2014]. Ainsi ce cadre théorique permet de passer d'une modélisation de la structure logique commune à presque tous les

documents (rejoignant le concept de *conventions partagées* qui sera défini en 4) à la modélisation fine des architectures textuelles, décrivant des objets textuels, structures et relations logiques pouvant être spécifiques à un domaine de connaissances ou même à un seul document. Plusieurs exemples d'architectures textuelles qui ne sont pas prises en compte par la structure logique décrite ici sont évoqués à la fin du Chapitre II précédent.

Cette distinction entre structures visuelle, logique, et discursive permet de mieux comprendre les travaux réalisés dans le domaine de l'analyse automatique de documents, présentés dans la section suivante.

## **2.2. Analyse automatique du contenu des documents**

---

On entend par « analyse automatique » les méthodes et algorithmes visant à extraire des informations relatives aux 3 types de structures présentées précédemment à savoir les structures visuelle, logique et discursive du document. L'objectif de ces travaux est souvent de permettre de gérer et/ou de rendre plus accessibles de grands ensembles de documents [Esposito et al. 2008], par exemple dans le cas de bibliothèques numériques. En effet, l'analyse automatique des structures visuelle et logique permet de numériser des documents initialement sous forme d'images scannées, et de les rendre accessibles en extrayant leur structure logique (ex. les titres permettent la navigation par signets). L'analyse automatique de la structure discursive quant à elle est un vaste sujet avec des applications allant de la génération de documents à la l'étude formelle du discours en linguistique [Mann 2006] ; nous l'écartérons cependant pour nous concentrer sur les structures logique et visuelle qui sont plus directement en lien avec les objectifs de nos travaux.

L'analyse automatique de la structure visuelle des documents vise donc à extraire les unités visuelles du document et leurs regroupements, à ne pas confondre avec la reconnaissance optique de caractères (*OCR – Optical Character Recognition*) dont le but est de numériser les caractères écrits du document. Ici le but d'algorithmes de reconnaissance de la structure visuelle serait plutôt d'identifier les différentes zones ou *bounding boxes* contenant du texte ou d'autres éléments, leur imbrication, ainsi que les propriétés du texte dans ces zones (police, taille, couleur, etc.). [Ramakrishnan et al. 2012]. Sans aller jusqu'à dire que cette analyse est triviale pour du contenu textuel, elle repose uniquement sur des paramètres formels comme la hauteur des lignes, l'espacement entre les mots ou entre les lignes, ou encore la taille de la police la plus répandue sur une page ; ce type d'analyse géométrique montre un taux d'erreur habituellement très faible au regard du nombre de blocs correctement extraits (voir [Ramakrishnan et al. 2012] par exemple), et repose de plus sur une littérature abondante [Liang et al. 1996; Ha et al. 1995]. L'analyse de la structure visuelle en elle-même est cependant d'un intérêt limité ; c'est pourquoi dans les travaux cités, elle précède souvent la labellisation des blocs segmentés qui deviennent alors des unités logiques.

En effet, comme le précisent [Liang et al. 1996; Cai et al. 2003] l'analyse du *layout* consiste à segmenter le document puis à classifier les segments identifiés selon un ensemble de catégories. Ces catégories peuvent être très génériques, comme dans les travaux de [Liang et al. 1996] qui identifient le contenu textuel, non-textuel et les tableaux, ou encore très spécifiques, comme c'est le cas de [Ramakrishnan et al. 2012] qui analysent des articles scientifiques, dont la structure logique est donc très codifiée, et définissent des catégories correspondant à cette structure : titre de l'article, auteurs, introduction, résultats, etc. [Power et al. 2003] précise que souvent dans ce domaine, les catégories



utilisées pour labelliser les unités visuelles des documents sont les mêmes que celles définies par les langages comme le HTML ou le Latex, étant donné que les classes d'objets qu'ils définissent sont prévus pour permettre de structurer à la fois logiquement et visuellement les documents (ex. une balise <title> permet de définir le titre du document mais aussi de lui donner un visuel particulier). Quoi qu'il en soit, l'analyse de la structure logique revêt d'une certaine complexité pour au moins deux raisons : tout d'abord, il n'existe pas de définition consensuelle des éléments formant la structure logique (ce qui semble justifié si l'on considère, sous l'angle du MAT, qu'il existe une quasi-infinité de possibilités pour la génération d'architectures textuelles et donc de structurations logiques des documents) ; de plus, les conventions de structuration logique des documents varient en fonction du contexte de publication du document, de son lectorat, etc.

La méthodologie d'annotation de la MFM présentée ci-après permet, dans une certaine mesure, de répondre à ces deux questions.

### **2.3. Des bases théoriques du MAT vers une formalisation de la MFM**

---

Le MAT stipule que les marques de mise en forme significantes (MFM) sont l'expression d'une volonté de l'auteur d'informer le lecteur, et peuvent être formalisées sous forme discursive par des métaphrases décrivant ces intentions (voir Chapitre I pour plus de détails). Par exemple, « *L'auteur <développe> <un paragraphe PO>* » est une métaphore qui pourrait être réduite visuellement sous la forme de retours à la ligne ou d'une disposition particulière du bloc de texte correspondant ; inversement les marques de MFM d'un paragraphe pourraient être développées discursivement par la métaphore citée. Les métaphrases étant des équivalents informationnels aux phénomènes de MFM du document, on peut donc formaliser la MFM d'un document par un ensemble de métaphrases. À noter que les métaphrases ciblent des *unités textuelles* (UT) du texte, c'est-à-dire des segments de texte du document d'origine, mais qui sont considérés dépourvus de toute mise en forme et donc réduits à leur expression linguistique, du texte « brut » dont on ne considère que le discours. Bien que chaque UT constitue une représentation minimale, indécomposable, l'ensemble des UT du métatexte peuvent se recouper et être imbriquées (ex. une emphase dans un paragraphe), il ne s'agit pas d'une segmentation du document au sens strict du terme qui impliquerait de diviser l'ensemble du texte en segments contigus. Ainsi, les métaphrases font référence à des unités textuelles d'au moins 3 façons différentes : en attribuant une propriété à une unité textuelle (ex. numérotée), en attribuant un statut à une UT (ex. paragraphe, titre) ou en mettant en relation plusieurs unités textuelles (ex. UT0 annote UT1 ; UT1 est composée d'UT0 et UT2). Dans la thèse de [Bou-Issa 2010], qui se base sur le MAT pour produire des pages tactiles représentant le contenu et les architectures textuelles de pages Web, on retrouve également ces 3 catégories d'informations traduites sur des pages tactiles : les types d'objets des pages web (ex. menu, groupement, titre, texte), les propriétés de ces objets (ex. couleur, caractère dynamique ou fixe) ainsi que les relations entre les objets (ex. composition : relation père-fils, est un zoom de l'objet, a la même mise en forme que).

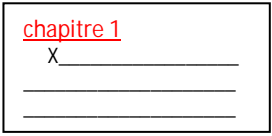
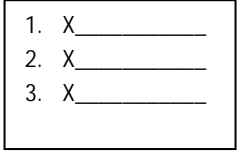
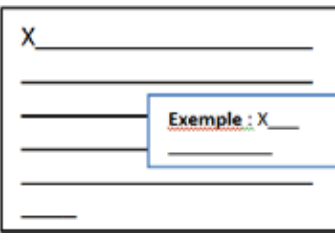
Phénomène visuel de MFM	Métaphrases correspondantes
	<p><u>Statut :</u> <i>L'auteur distingue un titre &lt;t0&gt;</i></p> <p><u>Propriété :</u> <i>L'auteur confère le niveau de chapitre à &lt;t0&gt;</i></p>
	<p><u>Statut :</u> <i>L'auteur distingue une énumération &lt;e0&gt;</i></p> <p><u>Propriété :</u> <i>L'auteur utilise le système arabe pour énumérer &lt;item1&gt;, &lt;item2&gt; et &lt;item3&gt;</i></p>
	<p><u>Statuts :</u> <i>L'auteur distingue un paragraphe&lt;p0&gt;</i> <i>L'auteur distingue un paragraphe&lt;p1&gt;</i> <i>L'auteur confère le statut d'exemple au paragraphe &lt;p1&gt;</i></p> <p><u>Relation :</u> <i>L'auteur illustre &lt;p0&gt; par &lt;p1&gt;</i> <i>L'auteur compose son texte de &lt;p0&gt; et &lt;p1&gt;</i></p>

Tableau 6- Exemples de métaphrases

Le MAT ne donne pas une liste de métaphrases ou de statuts arrêtée, mais spécifie simplement la forme syntaxique que les métaphrases doivent prendre et les contraintes du métatexte (ensemble des métaphrases d'un document). Le modèle est donc ouvert dans le sens où de nouvelles métaphrases peuvent être définies librement ; la seule contrainte pour définir une métaphrase est de respecter le schéma syntaxique d'origine, et que l'ensemble des métaphrases du métatexte respectent les lois de précedence et d'obligation définies par le modèle. Ces dernières stipulent que certaines métaphrases nécessitent la présence préalable d'autres métaphrases (précedence), ou impliquent la présence d'autres métaphrases à leur suite (obligation).

Un processus de formalisation des architectures textuelles a été décrit par [Maurel 2004] (p.95) ; les deux étapes correspondantes sont la *segmentation des objets textuels* et *l'extraction des paramètres formels* des objets textuels. Cependant ces travaux ne précisent pas comment opérationnaliser le processus, recommandant simplement d'utiliser une « base de règles éditoriale s'appuyant sur des indices syntaxiques, pragmatiques, stylistiques et structurels ». De plus, l'extraction des paramètres formels des objets textuels ne concerne ici qu'un objet particulier, à savoir les énumérations. Ce processus ne permet pas de modéliser les architectures textuelles d'un document (objets textuels et relations), mais simplement d'identifier différentes composantes des énumérations et d'en extraire certains paramètres (ex. type d'énumération : énumération parallèle).

En ce qui concerne la formalisation de la MFM, on peut également citer le modèle SARA (Signal Available Relevant Accessible information) de [Lemarié et al. 2008]. Bien que son objectif premier soit de prédire les effets d'objets textuels (de signaux) sur le lecteur, SARA inclue également un modèle descriptif des architectures textuelles. Ce modèle décrit les objets textuels en termes de *fonctions informatives* qui sont les informations que donne l'objet au lecteur, et en termes de *portée* des fonctions informatives des objets textuels (portée locale ou globale). Le modèle ne comprend que 6 fonctions informatives, dont 2 sont de l'ordre de relations de discours ; leur nombre restreint pose problème pour la description d'architectures textuelles variées. De plus les fonctions

informatives ne distinguent pas différentes composantes des architectures textuelles que l'on souhaite distinguer pour ensuite les restituer (statuts des objets textuels, relations entre objets textuels). Enfin ces travaux ne fournissent pas de méthode de formalisation d'un texte en termes de fonctions informatives.

D'autres travaux notables sont ceux de [Luc 2000], qui propose une liste de métaphrases et des classes d'objets textuels pour décrire les architectures textuelles mais, ces travaux s'inscrivant dans le domaine de la génération de texte, ils n'incluent pas de méthode d'extraction de ces informations.

Ainsi, bien que plusieurs travaux utilisent le MAT pour décrire certaines architectures textuelles, aucun ne fournit une méthode de formalisation dans le cadre d'un modèle unifié, permettant la description de toute architecture textuelle et de ses composantes. Sur les bases théoriques évoquées précédemment, nous pouvons approcher la formalisation de la MFM à travers deux étapes que nous reprenons des travaux de [Maurel 2004] : l'identification des phénomènes visuels (objets textuels) issus d'intentions auteuriales, puis sur la base du contexte du discours et surtout des marques de MFM, l'extraction des métaphrases correspondantes, explicitant les architectures textuelles. Cette seconde « grande » étape, la plus complexe, sera divisée en deux afin de la rendre la plus automatisable et formelle possible ; nous proposons donc une méthodologie en 3 temps, qui seront présentés successivement dans ce chapitre :

- § Segmentation visuelle du document (partie 3) : l'objectif est d'isoler les différents objets (textuels) impliqués dans des architectures textuelles, présentant donc des marques de MFM ; ils constitueront l'unité de modélisation de la MFM.
- § Identification des instances de conventions partagées (partie 4) : nous verrons ce qui justifie de l'existence de telles conventions d'architectures textuelles et nous en définirons un modèle, décrivant les objets textuels les plus répandus (ex. titres et les paragraphes), leur sémantique, ainsi que les propriétés typo-dispositionnelles et lexico-syntaxiques permettant leur identification. Cette étape consiste donc à identifier dans le texte les architectures textuelles relevant d'un usage conventionné des marques de MFM à un niveau inter-documents (relevant de conventions partagées). L'intérêt de cette étape du processus est de pouvoir maximiser l'automatisation du processus de formalisation.
- § Identification des conventions locales et de ses instances (partie 5) : cette étape permet d'identifier manuellement les architectures textuelles spécifiques au document étudié. L'intérêt de cette étape est son adaptabilité.

### **3. Segmentation des objets textuels des documents**

---

Cette partie propose une méthode pour segmenter le document en objets textuels. Le MAT postule que la réalisation des métaphrases architecturales donne lieu à des marques de MFM pouvant se présenter sous forme de : contraste *typographique* (police, graisse), contraste *dispositionnel* (espaces, indentation, positionnement) ou encore à travers la présence de *marqueurs lexico-syntaxiques*. Ainsi, la segmentation des objets textuels peut se faire selon deux axes : les propriétés typo-dispositionnelles et/ou les indices lexico-syntaxiques. Nous verrons comment le premier axe rejoint une littérature existante fournie, et comment le second axe pourrait contribuer à une formalisation de la MFM plus poussée.

### 3.1. Segmentation visuelle : une première approche

Nous avons vu que la segmentation visuelle des documents a déjà été vastement étudiée. Cette partie vise donc à recontextualiser ces travaux dans le cadre du MAT.

Dans le premier chapitre, nous avons vu que l'utilisation des marques de MFM dans un texte se fait suivant un principe de contraste. Le terme contraste est utilisé dans le vocabulaire commun pour qualifier une *différence*, par exemple une différence de luminosité dans une image. Dans le cas d'un texte le MAT définit le contraste comme une différence (relative donc) de typographie ou de disposition d'un segment de texte par rapport au texte environnant, permettant alors d'identifier ce segment comme étant une unité visuelle distincte. La segmentation de la structure visuelle, quel que soit le domaine qui l'aborde, repose toujours sur ce principe de contraste pour différencier les segments textuels. En effet les algorithmes correspondants utilisent des traits visuels pour repérer les ruptures dans le texte, comme le changement de police, implémentant ainsi les lois du regroupement visuel tels qu'énoncés par l'école du *Gestalt* [Wertheimer and Riezler 1944]. Ces lois décrivent le fonctionnement de la perception visuelle humaine, et plus précisément de la perception des formes à travers la capacité du système visuel à effectuer des regroupements de « points » ; ces lois du regroupement visuel (et donc de la dissociation) offrent ainsi un angle pertinent pour décrire le fonctionnement des algorithmes de segmentation visuelle des textes, en prenant ici le caractère écrit comme unité visuelle remplaçant le point.

- § Loi de proximité : *nous regroupons d'abord les points les plus proches entre eux*. Ainsi dans le cas d'un texte, des critères tels que l'espacement interlignes et inter-mots permettent une première passe de segmentation [Ramakrishnan et al. 2012]. On rappelle que, quelle que soit la loi énoncée, ses critères sont applicables de manière relative au contexte dans lequel elles sont appliquées (ce qui rejoint la définition initiale du contraste visuel). Ainsi pour extraire des objets textuels d'après la loi de proximité, on pourra comparer l'espacement interlignes de chaque occurrence à l'espace interlignes moyen<sup>80</sup> du document, et de même avec l'espace inter-mots.
- § Loi de similitude : *si la proximité ne permet pas de regrouper les points, nous regroupons les points les plus similaires entre eux*. Ce qui peut être traduit dans le cadre d'un texte par une similitude typographique (de police, taille, couleur, grasse, etc.) entre les caractères. Par exemple pour cet item de l'énumération des lois du Gestalt, on peut distinguer l'amorce qui est soulignée, l'annonce de la loi et le reste de l'item en italique.

À noter que pour la segmentation des objets textuels, ces lois peuvent être étendues à l'arrière-plan du texte qui représente parfois un indice pertinent supplémentaire. Par exemple, les séparations formées par un tableau (loi de proximité permettant de dissocier les cases à travers les lignes du tableau) ou un arrière-plan de couleur (loi de proximité et de similitude) peuvent être suffisantes pour dissocier des objets textuels, là où sa disposition et la typographie ne le permettraient pas.

<sup>80</sup> Ou plus exactement, pour reprendre les critères utilisés pour la segmentation visuelle, de l'espacement le plus représenté dans le document, ce qui correspond en termes statistiques à la classe modale d'espacement (le mode).

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer vitae ligula eu tortor tempus lobortis. Nullam pulvinar est vitae condimentum venenatis. Sed ac nisi varius, luctus lectus vel, sollicitudin arcu.	Duis quis molestie massa, at tempor purus. Donec pulvinar dictum ante, quis tempus odio faucibus sit amet. Sed iaculis commodo est at feugiat. Vestibulum imperdiet pharetra ullamcorper.
--	--

**Tableau 7 – Illustration de segmentation visuelle basée sur l’arrière-plan**

Ce type de phénomène fait partie de l’ensemble des *marques de MFM* applicables à un texte. On les classe dans la même catégorie que les marques typo-dispositionnelles de MFM ; on pourrait d’ailleurs considérer qu’ils relèvent de la typographie d’un texte, suivant la définition que l’on donne à la typographie. Quoi qu’il en soit, on considèrera que l’expression *marques typo-dispositionnelles de MFM* englobe ce type de marque de MFM.

Trois autres lois du regroupement visuel sont à l’œuvre, mais semblent moins pertinentes dans le cadre de la segmentation visuelle des textes :

- § Loi de continuité (ou clôture) : *des points rapprochés tendent à représenter des formes lorsqu’ils sont perçus ; nous les percevons d’abord dans une continuité, comme des prolongements les uns par rapport aux autres.* Cette loi s’applique à l’inférence de formes du monde réel (représentées de manière iconique) ; dans le cas du texte, elle peut justifier de la perception de *rectangles englobant* les blocs de texte, car les conventions d’écritures font que la juxtaposition des lignes proches peut former un rectangle. Nous ne considérerons pas les formes possiblement émergentes de l’agencement des caractères du texte, phénomène que nous considérerons comme marginal dans la mise en forme des textes (voir exemple ci-dessous).



**Figure 21 – Illustration de la loi de clôture du regroupement visuel (Gestalt)**

- § Loi du destin commun : *des parties en mouvement ayant la même trajectoire sont perçues comme faisant partie de la même forme.* On peut illustrer cette loi avec l’exemple des démonstrations de voltige aériennes, où plusieurs avions volant de manière synchronisée peuvent être perçus comme faisant partie d’un seul et même objet.
- § Loi de familiarité : *on perçoit les formes les plus familières les plus significatives.* Dans l’amas de tâches ci-dessous apparaît bien souvent un animal, pourtant invisible à la plupart des algorithmes de reconnaissance de formes :



Figure 22 – Illustration de la loi de familiarité du regroupement visuel (Gestalt)

La loi de proximité s'applique donc aux caractères proches et concerne les phénomènes de disposition du texte à savoir l'indentation, le colonage, l'interlignage, etc. La loi de similitude quant à elle s'applique aux propriétés typographiques des caractères ; c'est ce qui nous permet de regrouper visuellement **une emphase** dans une phrase. Ces lois permettent aussi d'étendre la définition de la MFM au-delà des caractères du texte, en incluant par exemple des changements d'arrière-plan du texte pour segmenter les objets textuels. On considèrera dans la suite du manuscrit que les marques typographiques du texte incluent également ce type de phénomènes de contraste visuel, couvrant les caractères et leur contexte visuel sur la page.

En conclusion, seront considérés comme objets textuels auxquels on attribuera des intentions auteuriales lors du processus d'annotation : toute unité textuelle présentant du contraste visuel comme définit par les lois de proximité et de similitude du Gestalt. Ces segments seront donc considérés comme porteurs d'informations sur les architectures textuelles, qui seront extraites dans les phases ultérieures du processus d'annotation. Cependant, cela n'implique pas que tous les objets textuels soient exprimés par des marques typo-dispositionnelles, certains pouvant l'être uniquement par des marques lexico-syntaxiques (voir section 3.3).

### 3.2. Le cas d'objets textuels imbriqués

Un constat empirique important est que la structuration visuelle des textes n'est pas simplement linéaire mais peut se faire avec plusieurs niveaux d'imbrication des objets textuels. Un objet textuel peut ainsi être contrasté par rapport à un autre, mais, au sein de chacun, il peut y avoir encore d'autres moyens de contraste définissant plusieurs autres objets imbriqués. Par exemple ci-dessous, le regroupement du haut est contrasté par rapport au reste du texte par sa couleur de fond, et au sein de ce regroupement le mot « propriété » est contrasté par sa couleur et sa police :

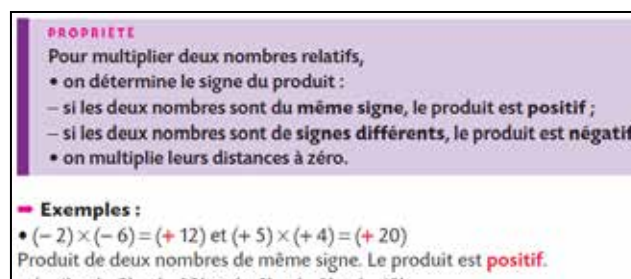


Figure 23 – Extrait d'un manuel de mathématiques : illustration de segments visuels imbriqués

Ces phénomènes d'imbrication visuelle nous ramènent à la notion de composition du MAT introduite par [Pascual 1991] dans sa thèse et reprise par [Luc 2000]. Ils définissent le verbe performatif «

composer » du métatexte comme permettant de « *déterminer comment les objets textuels sont organisés à l'intérieur d'un autre objet textuel, le plus grand objet textuel étant le texte lui-même* ». Cette notion de composition telle qu'elle est définie mélange deux niveaux de l'organisation des textes qu'il est important de distinguer dans le processus d'annotation. En effet, elle comprend à la fois les *regroupements visuels* que l'on peut physiquement percevoir au niveau d'une page du texte par des jeux de contraste basés sur les lois du regroupement visuel (exemple du rectangle violet sur Figure 23), et les *regroupements logiques* qui peuvent être réalisés à travers des convention de MFM, comme par exemple la création de *sections* regroupant le contenu imputable à un titre donné, ou encore la hiérarchie entre items d'une énumération (exemple des deux sous-items du premier item de l'énumération sur la Figure 23) . Pour l'exemple des sections, on remarque en effet que la plupart des réalisations correspondantes dans les textes (y compris dans ce manuscrit de thèse) ne comportent pas d'indices typo-dispositionnels permettant de regrouper les éléments d'une section entre eux : la loi de proximité et de similitude ne s'appliquent pas ici. Cela est particulièrement notable lorsqu'une section est étalée sur plusieurs pages. Ainsi la composition d'une section se fait très souvent de manière « logique » sur la base de certaines règles conventionnées (décrites dans la partie 4).

À ce stade de l'annotation, on s'intéresse uniquement à l'identification des objets textuels et notamment des objets imbriqués visuellement. Les étapes suivantes du processus permettront de définir les différents statuts et relations logiques (dont celle de composition) attribuables aux objets textuels. La notion de *composition* intervient donc lors des étapes ultérieures de l'annotation, lorsqu'on donne du sens à ces objets textuels.

### **3.3. Segmentation d'objets textuels sur la base d'indices lexico-syntaxiques**

---

Les marques lexico-syntaxiques sont des reliquats de la réduction visuelle des métaphrases ; par exemple, la réalisation du métatexte d'une énumération peut se faire sans puce ni retour à la ligne, mais de manière entièrement discursive avec des indices lexicaux (premièrement, enfin, etc.) et/ou des indices syntaxiques (points-virgules entre les items) ; voir Tableau 8 pour exemple. Ainsi la réduction visuelle des métaphrases peut se faire soit en laissant uniquement des traces typo-dispositionnelles, soit en laissant des traces typo-dispositionnelles et lexico-syntaxiques, soit en laissant uniquement des indices lexico-syntaxiques comme c'est le cas pour le précédent exemple de l'énumération. Dans les deux premiers cas, la segmentation des objets textuels peut être réalisée sur la base de leurs marques typo-dispositionnelles. En revanche, dans le dernier cas, les métaphrases sont reformulées sous forme discursive sans présenter aucun indice typo-dispositionnel, et la segmentation de phénomènes de MFM relève alors du traitement du discours.

<i>Métatexte</i>	<i>Réalisation par des marques et lexico-syntaxiques</i>	<i>Réalisation par des marques typo-dispositionnelles et lexico-syntaxiques</i>	<i>Réalisation par des marques typo-dispositionnelles</i>
L'auteur distingue une énumération <enum0> L'auteur organise <enum0> en deux items <item1> et <item2>	Tout d'abord [...] ; ensuite [...]	<ul style="list-style-type: none"> <li>• Tout d'abord [...]</li> <li>• Ensuite [...]</li> </ul>	<ul style="list-style-type: none"> <li>• [...]</li> <li>• [...]</li> </ul>

**Tableau 8 - Exemples de réalisations d'un métatexte par différentes marques de MFM**

Ce dernier point questionne les limites à poser pour la segmentation des objets textuels. Rappelons qu'à partir du moment où un texte est inscrit sur un support physique, il est mis en forme au sens du MAT, c'est-à-dire que cette inscription même et la forme dans laquelle elle est réalisée résulteraient d'intentions auteuriales. Si l'on suit cette définition, l'analyse complète de la MFM d'un document devrait aussi prendre en compte le contraste inter et intra-phrastique, ce qui reviendrait à étudier les différences et similitudes entre tous les indices lexicaux et syntaxiques. Bien que cet aspect soit intéressant et fasse partie intégrante de la MFM, les applications de tels travaux visent plutôt l'analyse des langues et du discours et non la restitution d'informations non-disponibles aux déficients visuels dans leur accès aux documents, ce qui nous intéresse plus particulièrement dans nos travaux.

En effet, si la MFM est réalisée uniquement de manière discursive dans le texte, on peut considérer qu'elle est déjà disponible et donc accessible jusqu'à un certain point si l'on se place dans le cas de l'accès aux documents par synthèse vocale ou en braille, modalités permettant toutes deux de restituer les informations du discours. Cependant, il peut s'avérer pertinent dans certains cas de segmenter des objets textuels discursifs particuliers, afin d'explicitier les informations qu'ils véhiculent. Par exemple, l'utilisation de guillemets peut dénoter d'un dialogue entre personnages, du report d'une phrase prononcée à l'oral, ou encore de l'utilisation d'un mot dans un contexte inhabituel ou de manière informelle (ex. tu te « maganes »<sup>81</sup> la santé pour rien à travailler autant, comme on dit au Québec). Pendant une lecture par synthèse vocale, ces informations dénotées par l'usage des guillemets ne sont pas accessibles, sauf peut-être si la prosodie de la voix est modifiée auquel cas la restitution n'est pas totalement explicite et dépend de la capacité de l'auditeur à l'interpréter. On peut également imaginer des réalisations lexico-syntaxiques plus complexes et atypiques, par exemple pour décrire un algorithme : étape 1 => étape 2 ; condition ; si oui => étape3A, si non => étape 3B. Ce type d'objet textuel gagnerait à être identifié puis reformulé pour un accès non-visuel.

Ainsi certains objets textuels discursifs peuvent gagner à être identifiés afin de servir un but d'accessibilité de l'information. L'utilité d'une telle démarche dépend à la fois des phénomènes discursifs présents dans le document et du contexte d'accès. Par exemple si le texte est lu par une plage braille les guillemets seront identifiés via le caractère braille correspondant. De plus, certains phénomènes discursifs peuvent être plus ou moins explicités par le discours environnant, par exemple si l'on utilise des parenthèses pour citer un exemple, on peut le faire en rajoutant le terme « exemple » après la parenthèse ouvrante (ou « ex. »), ce qui permet de rendre le phénomène plus explicite à l'oral où les parenthèses peuvent ne pas être directement mentionnées par la synthèse

<sup>81</sup> Maganer est un verbe du langage familier encore en usage au Québec dont le sens est dérivé du verbe « estropier », son sens d'origine.



vocale. Ainsi l'identification systématique d'objets textuels marqués uniquement par des indices lexico-syntaxiques n'est pas pertinente dans le cadre de nos travaux. Elle peut cependant l'être dans le cas de certains objets et pour des contextes d'utilisation particuliers, ce qui implique que cette identification se fasse manuellement. Nous pensons qu'un critère pertinent pour savoir si un objet discursif doit être identifié est son degré d'ambiguïté lors de son accès (à l'oral ou en braille), ainsi que l'impact de son explicitation pour la réalisation de la tâche impliquant le document (ex. une tâche de compréhension). Ce double critère sera d'ailleurs utilisé de manière plus générale pour étudier les méthodes de restitution des architectures textuelles.

Lorsque cela s'avère être pertinent d'après les critères évoqués, la segmentation de certains objets textuels discursifs peut se faire à l'aide de règles simples et donc de manière automatique, et ce grâce au principe d'identité des marques de MFM. En effet ce dernier stipule que, pour un même texte, deux réalisations d'un même objet textuel (issues des mêmes métaphrases) doivent présenter les mêmes marques de MFM afin de respecter le principe d'identité. Ainsi en identifiant les schémas syntaxiques et/ou les listes de mots du lexique utilisés pour exprimer les objets jugés intéressants, ces derniers peuvent être identifiés de manière automatique.

### 3.4. Exemple de segmentation

Ci-dessous le résultat d'une segmentation visuelle d'une page de manuel scolaire, où les segments identifiés sont encadrés en vert.

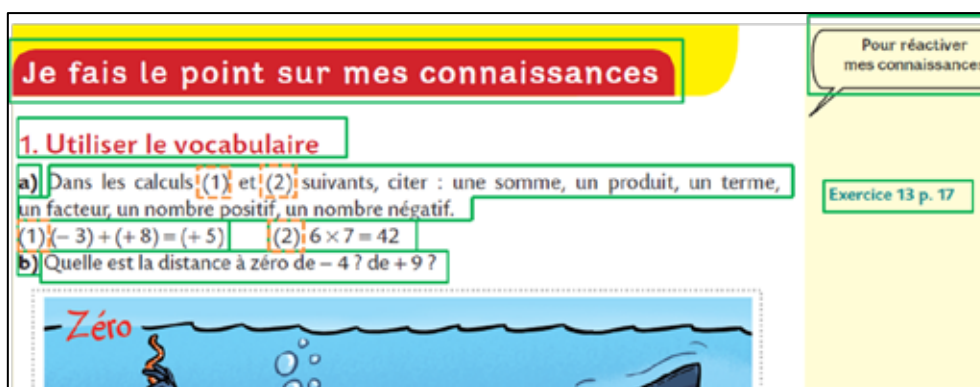


Figure 24 – Exemple de segmentation des unités visuelles d'une page d'un manuel de Mathématiques

Les différences de typographie (couleur, police, grasse) et de disposition (espacement, positionnement en marge) permettent donc de segmenter le document. L'utilisation de parenthèses permet ici de relever un objet textuel discursif (en pointillés orange sur la Figure 24) qu'il pourrait être intéressant de formaliser, étant donné que les parenthèses permettent ici de donner le statut de *label* aux chiffres 1 et 2. On aurait pu également segmenter l'énumération discursive de la première question, dont l'amorce est séparée des items par « : » ; cela illustre le fait que la segmentation d'objets textuels peut se faire uniquement sur la base d'indices lexico-syntaxiques. Il est donc important de mettre la segmentation des objets textuels en perspective avec l'utilisation que l'on en fait, afin de pouvoir se départir de la complexité d'une analyse exhaustive et systématique du discours du texte.

### **3.5. Conclusion pour nos travaux**

---

Cette phase de segmentation visuelle (automatisable) permet de définir objets textuels auxquels on attribuera des intentions auteuriales lors du processus d'annotation, à savoir toute unité textuelle présentant un contraste visuel ; ce qui se traduit par l'application des lois du regroupement visuel du Gestalt. Des objets marqués uniquement par des indices lexico-syntaxiques peuvent aussi être identifiés dans les cas où cela s'avère pertinent.

Une fois ces unités de MFM repérées, les prochaines étapes vont permettre l'extraction des informations correspondant aux architectures textuelles, impliquant les différents objets segmentés.

## **4. Définition et identification des conventions partagées de MFM**

---

---

Cette partie postule l'existence de conventions partagées de MFM, postulat qui sera argumenté dans la section suivante, c'est-à-dire l'existence d'un *code* universel (mais aux réalisations variables) d'utilisations des marques de MFM. Ces conventions permettent dans le cadre de l'annotation des documents de fournir une première interprétation automatisable des architectures textuelles sur la base des objets textuels de MM identifiés dans la phase précédente du processus.

### **4.1. De l'existence de conventions de MFM partagées**

---

Les travaux sur le MAT attestent de l'usage conventionné des marques de MFM et des architectures textuelles qu'elles expriment, au niveau local d'un document ou au niveau d'un genre de documents ou d'un domaine de connaissances [Lemarié 2006a; Pascual 1996; Virbel 1989]. Cependant ces travaux ne recensent pas ce que l'on pourrait appeler les conventions « universelles » de MFM. Or, l'existence de conventions de MFM partagées universellement se justifie à travers plusieurs constats empiriques. Premièrement, on retrouve certains objets de manière systématique quel que soit le document textuel étudié, indépendamment de la langue et du domaine dans lequel il s'inscrit, à travers cependant un ensemble de réalisations différentes (aux marques de MFM différentes). Les titres et le découpage du texte en paragraphes en sont les exemples les plus parlant ; que ce soit dans les journaux, les manuels scolaires, les articles scientifiques, les romans ou sur le web, les documents sont structurés quasi-systématiquement à l'aide de titres. Ce constat est également vérifié pour des documents plus anciens comme des manuscrits religieux, et également pour des langues utilisant un sens de lecture et un alphabet différent de l'alphabet latin, comme l'Arabe.



Figure 25 – Manuscrit Arabe des mille-et-une nuits datant du XIV<sup>ème</sup> siècle (image libre de droits)

Dans une moindre mesure, d'autres objets textuels sont également très répandus comme les énumérations ou les emphases sur des groupes de mots.

L'existence de telles conventions se constate aussi à travers la littéracie des lecteurs : présenter un document inconnu à un lecteur « expert », ne contenant qu'un seul titre et une seule énumération, et il saura très probablement interpréter ces phénomènes sans qu'il ait besoin d'inférer leurs significations à partir de plusieurs occurrences dans le texte (principe d'identité). Cela montre bien l'existence d'un code d'utilisation de la MFM, qui est acquis plus ou moins implicitement par les lecteurs au cours du développement de leurs compétences de lecture.

Enfin, un dernier argument en faveur de l'existence de conventions partagées de MFM est que l'on retrouve de manière récurrente dans différents langages de structuration de document et outils d'édition de texte les mêmes structures visuelles pour organiser les documents, qui seront présentés dans la section suivante. On peut citer les langages HTML et LATEX, et des outils d'édition de texte comme MS-Word ou Open-Office, qui incluent tous par défaut la possibilité d'insérer des titres, énumérations, emphases et autres objets des conventions partagées.

Pour les mêmes arguments que ceux justifiant de l'existence de conventions universelles, on constate empiriquement que d'autres niveaux de conventions de MFM existent, qui auraient émergé d'une culture de la langue écrite développée dans différents contextes. Ainsi la diversité des contenus et des utilisations des documents a pu faire naître, avec le temps, des conventions par genre de document (ex. les références entre crochets pour les articles scientifiques) ou encore par domaine de connaissances (ex. les couleurs utilisées dans un code informatique écrit). Par exemple, [Richard 1999](p. 27) pointe le fait que la familiarité avec un domaine (ici des textes de type « consigne ») permet au lecteurs de mieux comprendre les textes, et notamment leur MFM : « *La compréhension des textes de consignes d'action ne pose guère de difficultés à ceux qui sont familiers du domaine, alors qu'ils sont souvent incompréhensibles pour ceux qui ne le sont pas. C'est vrai pour les différents types de textes, que ce soient des recettes de cuisine, des textes administratifs, des modes d'emploi* ». Ainsi un auteur produisant un document d'un certain genre dans un certain domaine pourra réutiliser les conventions existantes, s'il en a connaissance, afin que son message soit le plus aisément compréhensible par le plus grand nombre, pour répondre à des utilisations particulières du document ou pour véhiculer des architectures textuelles particulières. Enfin, d'après la définition de la MFM et en respectant les principes de contraste et d'identité, un auteur a toujours la possibilité de définir ses propres conventions de MFM dont la portée est, sauf coïncidence, locale au document produit.

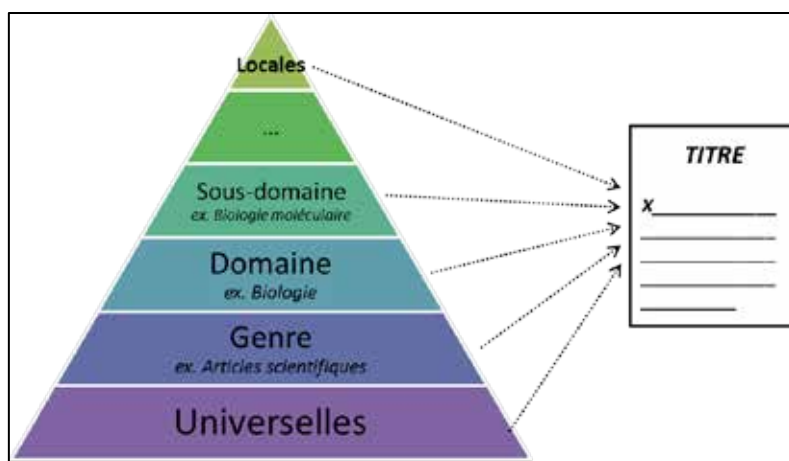


Figure 26 – Pyramide des couches de conventions de MFM

Dans le cas du processus d’annotation de la MFM que nous décrivons dans ce chapitre, nous ne considérerons que les conventions universelles dans la première phase de l’annotation, qui se veut automatisable, puis le reste des conventions de MFM utilisées dans les documents comme des conventions *locales* au document, qui seront annotées manuellement dans la seconde phase. Évidemment certaines conventions considérées ici comme *locales* peuvent être partagées au niveau du domaine ou du genre correspondant au document. Toutefois, cette distinction nous paraît utile pour tendre vers une annotation automatique de la MFM dans le futur. Ce que nous proposons ici est un premier niveau d’automatisation qui consiste à repérer les conventions universelles dans les documents écrits, puis de manuellement modéliser le reste des conventions du document (considérées comme locales). Le repérage automatique de conventions partagées par certains genres de documents ou domaines de connaissances impliquerait de réaliser des études longitudinales sur des corpus de documents de différents genres et domaines, ce qui actuellement est à l’état de perspective pour de futurs travaux.

## 4.2. Modélisation des conventions de MFM partagées universellement

On suppose donc l’existence de classes d’architectures textuelles (d’objets textuels et de relations entre objets) conventionnées, et dont les réalisations dans les textes sont des instanciations locales. Cela est rendu possible par l’iconicité des objets textuels, c’est-à-dire leur propension à représenter un représentant, au sens de la *sémiotique*, ce que le MAT reconnaît comme étant le principe *d’identité*, avec ici une portée inter-documents.

Nous présenterons d’abord les classes d’objets (ou statuts pour reprendre le terme utilisé pour décrire les métaphrases) relatives aux conventions de MFM dites « universelles », puis ensuite les relations entre objets. Le repérage de ces classes et relations constitue une première étape dans la formalisation des architectures textuelles.

### 4.2.a. Classes d’objets textuels

Sur la base d’une étude d’un corpus de manuels scolaires<sup>82</sup>, mais aussi à l’aide des balises utilisées dans les documents textuels produits avec les langages HTML et LaTeX, nous avons défini des classes

<sup>82</sup> Le corpus était constitué de 2 manuels de mathématiques complets de 3<sup>ème</sup> et de 4<sup>ème</sup> d’environ 300 pages chacun, et des extraits d’autres manuels niveau 1<sup>ère</sup> représentant environ 30 pages.

d'objets textuels constituant une partie des conventions de MFM les plus largement partagées. Le concept de classe est issu de la programmation orientée-objet, et désigne des descriptions abstraites de catégories d'objets, qui peuvent être instanciées avec certaines valeurs particulières. Ici chaque classe désigne donc une description abstraite d'un statut d'objet textuel, dont les objets textuels retrouvés dans les documents appartenant à cette classe en sont des instances ayant des propriétés de réalisations particulières. Par exemple pour la classe *titre*, tous les objets textuels de classe titre dans les documents auront des propriétés communes en termes d'architectures textuelles, mais n'auront pas tous les mêmes marques de MFM associées à leur réalisation.

Chaque classe (ou statut d'objet textuel) est repérable par un ensemble d'indices typo-dispositionnels et lexico-syntaxiques qui sont énumérés dans le Tableau 9 ci-après. Ces listes d'indices rendent l'identification des classes automatisable, ce qui ne veut pas dire que ces indices suffisent pour confirmer avec certitude la présence des objets. En effet un auteur peut très bien occulter les conventions de MFM partagées pour définir ses propres conventions, ou alors reprendre les conventions partagées de manière modifiée ; par exemple un objet peut avoir les traits d'un titre sans pour autant en jouer le rôle dans le contexte du document. Ainsi la deuxième phase d'annotation des *conventions locales*, qui est réalisée manuellement, permettra d'infirmier ou confirmer les architectures textuelles extraites d'après les conventions partagées définies.

Étant donné qu'il existe une quasi-infinité de réalisations (de formes dans le texte) d'un même objet textuel, la liste des marques de MFM associées aux classes d'objet ne constituent pas des listes d'indices discriminants qui, lorsqu'ils sont tous présents, permettent de classifier l'objet. Certaines marques sont, en effet, nécessaires à la réalisation de l'objet (symbole Ū dans le tableau), mais les autres n'apparaissent que pour certaines réalisations des objets correspondants, et ne permettent donc pas d'identifier à elles seules les classes des objets. Mis à part les marques systématiquement présentes (nécessaires) pour identifier certaines classes d'objet, on propose ici un modèle « probabiliste » dans le sens où plus le nombre d'indices correspondant à une classe est grand, plus la probabilité d'attribuer cette classe à l'objet est grande. Enfin, il faut préciser que le Tableau 9 ne recense que les objets de MFM ayant une réalisation au moins partiellement visuelle, laissant de côté les phénomènes purement discursifs pour les raisons évoquées précédemment (voir 3.3).

Classe d'objet de MFM	Marques de MFM associées	Balises HTML correspondantes	Commandes LaTeX correspondantes
<u>Regroupement</u>	Tout contraste visuel (principes de similitude et de proximité) permettant de dissocier une zone d'une autre et de l'arrière-plan. <i>C'est la classe « mère » de toutes les classes présentées ici, dans le sens où toutes les autres classes d'objets sont aussi des regroupements. Tous les segments identifiés après la première phase de l'annotation sont d'ailleurs d'office des regroupements, étant donné que la segmentation s'effectue sur ce principe.</i>	-	-
Page	ŪRegroupement de forme rectangulaire englobant le texte	<html>	\pagebreak
Bloc de texte	ŪRegroupement contenant uniquement du texte non contrasté dispositionnellement en son sein	Plusieurs balises utilisables. Ex. : <span>	Plusieurs commandes utilisables.
Paragraphe	ŪEst un bloc de texte ŪDébute sur une nouvelle ligne - espacement avant et/ou après - police du texte la plus représentée sur la	<par>	\par

#### 4. Définition et identification des conventions partagées de MFM

	page (on l'appellera PM <sup>83</sup> ) - alinéa sur la première ligne		
Emphase	Ü regroupement de texte présentant un contraste typographique au sein d'un bloc de texte englobant	<strong>, <em>, <cite>	\emph ; \textbf;\texti, \begin {quotation}
Marqueur d'ordre	- positionné en début de ligne - formé de 1 à 2 caractères alphabétiques / chiffres / symboles - suivi d'un caractère de séparation : . / ; / - / )	textuel ou <ol><li>	textuel ou \itemize
Titre	Ü Est un bloc de texte - Espacement avant et/ou après - positionnement centré ou décalé par rapport à la marge - taille de police plus grande que la PM - couleur différente de la PM - présence d'un marqueur d'ordre	<h(X)>, <title>	\maketitle ; \chapter ; \section ; \subsection ; \subsubsection ...
Item d'énumération	Ü Est un bloc de texte - présence d'un marqueur d'ordre - présence d'une puce, symbole ou autre objet graphique en début de ligne - indentation - présence d'autres items précédents ou suivants	<li>	\itemize
Énumération	Ü Présence d'au moins 2 items - bloc de texte précédent les items terminé par " : / , / ; " ou comprenant certains indices lexicaux : « suivant / ci-après » - présence de marqueurs d'ordre - marques dispositionnelles identiques sur une séquence de plusieurs objets textuels	<ol> et <ul>	\begin{itemize} et \begin{enumerate}
Objet graphique	Ü Regroupement non-textuel - forme simple et iconique - une seule couleur	<img>	\begin{figure}
Image	Ü Regroupement non-textuel - formes avec des détails - formes du monde réel - de forme rectangulaire	<img>	\begin{figure}
Cellule de tableau	Ü Forme rectangulaire - délimité par des bordures (traits verticaux et horizontaux) Ü Présence d'au moins 1 autre cellule avec une bordure commune - contenant un bloc de texte	<tr> <td> <th>	"&" et \hline
Tableau	Ü Présence d'au moins 2 cellules avec une bordure commune - forme rectangulaire - croisement de lignes représentant les bordures	<table>	\begin{tabular}

**Tableau 9 - Classes d'objets relatives aux conventions partagées de MFM**

Cette liste de classe est non-exhaustive et a vocation à être étendue ; la même remarque s'applique également aux indices visuels et lexico-syntaxiques (marques de MFM) permettant d'identifier ces objets dans les textes. Cependant nous pouvons émettre l'hypothèse que tout objet textuel peut être

<sup>83</sup> PM pour Police Modale, c'est-à-dire la police de texte (incluant la famille de la police, sa taille, etc.) la plus représentée sur la page.

décrit de façon minimale par les classes définies ci-avant. Concernant les indices typo-dispositionnels, comme évoqué précédemment la réduction de ces métaphrases étant sujette à variations, il n'est pas possible de définir précisément les indices dénotant de chaque classe de manière exacte. Malgré cela, la classification des objets textuels peut être réalisée de la manière suivante : lorsque la présence de certaines marques de MFM caractéristiques de la classe est vérifiée, plus l'objet présente de marques caractéristiques de la classe listées dans le tableau, plus il est probable qu'il appartienne effectivement à cette classe. De plus, la définition de certains indices se fait de manière relative (ex. par rapport à la police plus représentée sur la page), ce qui les rend plus adaptables.

Les balises et commandes correspondant aux classes définies soulignent le fait que ces types d'objets sont fréquemment retrouvés dans la nature. Cependant, la correspondance entre classes d'objets de MFM et balises des langages cités n'est pas systématique (voir Chapitre II pour plus de détails sur le fonctionnement des langages de markup), les auteurs de documents n'utilisant pas toujours les balises de manière sémantique, préférant contrôler directement l'apparence visuelle du document. Ce dernier point montre la nécessité de pouvoir annoter automatiquement les documents, même ceux issus de langages comme HTML ou LaTeX.

Enfin, il est important de préciser que nous avons choisi de lier entre elles les classes de notre modèle des conventions partagées de MFM par *un paradigme d'héritage*, tel qu'il est défini dans la programmation orientée objet en informatique, où chaque classe d'objet peut présenter une ou plusieurs classes filles. Les classes filles héritent alors des propriétés des classes mères, et chaque objet instanciant une classe fille *est* également un objet de la classe mère. Dans notre cas, chacune des classes définies *est un* regroupement, une *classe fille* de la classe regroupement ; cela impliquant que chaque classe fille hérite des propriétés de la classe mère (regroupement). Dans notre cas ces propriétés sont les métaphrases explicitant l'acte textuel correspondant à la classe d'objet regroupement (voir Tableau 10 ci-après). Certaines autres classes héritent également de la classe « bloc de texte ». L'intérêt de cette hiérarchie d'héritage est double : permettre la création de nouvelles classes par exemple lors de l'annotation selon des conventions locales du document, et permettre de systématiser la formalisation de la MFM dans un cadre unifié.

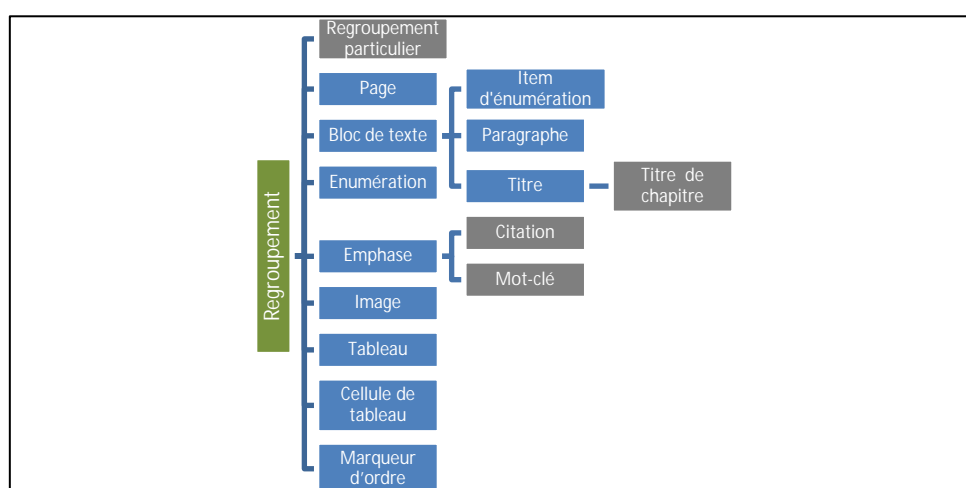


Figure 27 - Hiérarchie d'héritage des classes de conventions partagées de MFM

Les classes définies permettent de donner des statuts aux objets textuels. Cependant, au-delà d'un simple label, certains statuts dénotent de *propriétés* particulières ; le terme *propriété* désigne de

manière générique toute métaphore ciblant un objet textuel, qui ne représente ni l'application d'un statut ni d'une relation. On cherche donc à *dérivée* les statuts des objets textuels en un ensemble de propriétés, pour tenter d'explicitier les intentions sous-jacentes aux statuts « conventionnés » des objets textuels. D'un point de vue théorique, les métaphrases énoncent des actes de langage et à ce titre, comme pour tout acte de langage, [Austin 1975] considère qu'ils peuvent être soit directs soit dérivés de l'acte primitif. Ainsi chaque métaphore peut être dérivée en une ou plusieurs autres métaphrases explicitant les intentions liées à l'acte (la métaphore) initial ; par exemple « *donner le statut de Titre* » peut impliquer « *intituler une partie* » et également « *distinguer un bloc de titre (du reste du texte)* ». La dérivation des statuts permet donc d'explicitier les propriétés attribuables aux classes définies ; elle se fait également à travers le paradigme d'héritage : les classes d'objets filles héritent des propriétés des classes mères. La classe regroupement étant la classe racine (au même titre que la classe « objet » en programmation orientée-objet), elle transmet les métaphrases explicitant ses propriétés à toutes les classes de notre modèle. La classe regroupement (et par héritage tous les autres objets du modèle) possèdent la propriété suivante :

**Démarquer le segment** : un regroupement est un segment démarqué visuellement, ce qui en soi véhicule trois informations qui sont : le début du segment, la fin du segment, et la troisième presque implicite est le « typage » du contenu du segment entre le début et la fin. On retrouve cette métaphore dans le modèle de [Lemarié et al. 2008] décrivant les effets des signaux sur le lecteur, comme étant la fonction informative permettant de démarquer les frontières entre objets.

Ci-dessous les métaphrases dérivées des classes d'objets de MFM définies. À noter que nous faisons ici le choix d'un vocabulaire arrêté même si chaque métaphore appartient à un ensemble de métaphrases équivalentes syntaxiquement et sémantiquement ; de plus, les métaphrases rapportées ci-dessous le sont sous une forme plus succincte que leur forme d'origine générale : « *L'auteur <verbe performatif> <compléments/liens de dépendance> <identifiant du ou des objets textuels>* ».

<b>Classes d'objet de MFM</b>	<b>Métaphrases dérivées</b>
Regroupement	démarquer le début et la fin du segment démarquer le segment des autres segments <i>La notion de démarcation a été reprise de [Lemarié et al. 2008] où elle désigne le fait de rendre perceptible des frontières des objets textuels.</i>
Page	<regroupement>
Bloc de texte	<regroupement>
Marqueur d'ordre	<regroupement> numéroter le segment subséquent définir un niveau hiérarchique pour le segment subséquent <i>(conventionnellement une relation père-fils peut être définie par la succession de plusieurs marqueurs d'ordre, par exemple : a) père a)1. Fils 1, a)2. Fils 2, etc.)</i>
Paragraphe	<regroupement> marquer une rupture dans le discours par rapport au paragraphe précédent
Emphase	<regroupement> marquer une différence avec le texte du bloc auquel il appartient
Titre	<regroupement> posséder un niveau hiérarchique <i>(défini explicitement par le marqueur d'ordre associé au titre et/ou par la différence de taille et/ou de saillance des polices utilisées pour les titres du</i>



	<i>document</i> marquer l'importance du contenu du segment intituler les regroupements subséquents jusqu'au prochain titre de même niveau
Énumération	<regroupement> énumérer un ensemble d'items ordonner un ensemble d'items hiérarchiser un ensemble d'items
Item	<regroupement> développer l'amorce de l'énumération
Image	<regroupement>
Tableau	<regroupement> organise un ensemble de cellules en ligne et colonnes définit des relations entre cellules en fonction de leur ligne/colonne
Cellule de tableau	<regroupement> appartient à une ligne et une colonne donnée

**Tableau 10 - Métaphrases dérivées des classes d'objets de MFM**

En conclusion, chaque objet textuel identifié à l'issue de la segmentation du document se verra attribuer un statut parmi ceux définis dans les conventions partagées, ce qui implique qu'on pourra le formaliser à l'aide des métaphrases associées à son statut.

#### 4.2.b. Relations conventionnées entre objets textuels

De la même façon que certaines conventions partagées de MFM permettent de définir des classes d'objets attribuables aux segments, certaines relations entre segments de MFM sont également « conventionnées » et partagées très largement. Nous avons relevé 7 relations que l'on retrouve dans presque tous les documents écrits : composition (agrégation) de segments, intitulation de segments, définition de l'ordre et de la hiérarchie entre segments, labellisation de segments, numérotation, et enfin annotation d'un segment par un autre. Rappelons qu'une relation au sens du MAT est formalisée par une phrase constituée d'un verbe performatif exprimant l'action d'un ou plusieurs segments de MFM sur un ou plusieurs autres segments.

Relation	Description	Métaphrases dérivées
Composer	La <i>composition</i> fait référence au fait que plusieurs objets peuvent être réunis pour composer un objet englobant. Cette composition est, par convention, réalisée de deux manières : - à travers des indices typo-dispositionnels permettant de définir une imbrication des segments (voir exemple en 3.2). - avec les titres qui définissent par convention des regroupements, composés de tous les segments subséquents au titre jusqu'au prochain titre de même niveau (ex. chapitre, section). Nous utiliserons le terme <i>section</i> pour désigner les regroupements logiques engendrés par la présence de titres.	<i>L'auteur compose &lt;section 2.B&gt; par &lt;titre_t0&gt;, &lt;paragraphe_p0&gt; et &lt;tableau_t0&gt;</i>
Intituler	La présence d'un objet de la classe titre implique, conventionnellement, qu'il va intituler le regroupement composé subséquent (voir « composition » ci-avant), ou selon la convention appliquée un ensemble de segments marqués visuellement (ex. un titre d'une colonne de tableau ; cela rejoint des conventions locales particulières qui sont prises en compte	<i>L'auteur intitule &lt;section 2.B&gt; par &lt;titre 2.B&gt;</i>

#### 4. Définition et identification des conventions partagées de MFM

	dans l'annotation à l'étape suivante d'extraction des conventions locales).	
Ordonner	Le sens de lecture de la langue fournit par convention un premier ordonnancement des segments de MFM (de haut en bas et de gauche à droite pour les langues d'origine Latine). Parfois cet ordre peut être défini explicitement à l'aide de marqueurs (ex. « a. » ou « II. »), dits marqueurs d'ordre.	<i>L'auteur organise &lt;section 2.B&gt; par &lt;paragraphe_p0&gt;, &lt;tableau_t0&gt;</i>
Hiérarchiser	Des relations hiérarchiques « père-fils » sont définies à travers plusieurs conventions : - les niveaux des titres permettent de les hiérarchiser entre eux : un titre de niveau N est le père de tous les titres de niveaux supérieurs à N jusqu'au prochain titre de niveau N d'après leur ordre d'apparition dans le document. - les marques de MFM des items des énumérations permettent de les hiérarchiser entre eux ; par exemple la distance par rapport à la marge gauche et/ou le niveau hiérarchique explicite définit par des marqueurs d'ordre permettent de définir une hiérarchie. - d'une manière plus implicite, un objet composé de plusieurs autres objets (marqués par une imbrication visuelle par exemple) est le père de ces objets.	<i>L'auteur divise &lt;section 2 &gt; en &lt;section2A&gt; et &lt;section2B&gt;</i>
Labelliser	Certains objets textuels peuvent être labellisés par d'autres, c'est-à-dire qu'on peut y faire référence à travers le label qui leur est associé. On retrouve cette convention pour : - les sections et autres objets composés définis par les titres, qui sont labellisés par les titres qui les intitulent (ex. dans la section « mécanique des fluides»...). - les marqueurs d'ordre numérotant des objets peuvent labelliser ces derniers (ex. question a) de l'exercice). - définis de manière plus ou moins explicite, par exemple pour les légendes des figures et tableau (ex. tableau 3, figure 7), ou encore dans le texte (ex. « deux propositions (A) et (B) seront présentées... »).	<i>L'auteur labellise &lt;titre section&gt; par le label &lt;label 2.B&gt; (ce label serait d'ailleurs « 2.B »)</i>
Numéroter	Les marqueurs d'ordre permettent de numéroter les objets textuels subséquents, ce qui permet par convention de : - leur donner un label, comme décrit sur la ligne précédente (ex. on peut faire référence à un titre soit par son contenu soit par le marqueur d'ordre qui lui est associé comme « partie III. »). - expliciter l'ordre des objets de même type numérotés - expliciter les relations hiérarchiques entre objets de même type numérotés ; comme évoqué précédemment l'utilisation de plusieurs marqueurs d'ordre successifs permet de définir une hiérarchie entre objets : a) Père a)1. Fils 1 a)2. Fils 2	<i>L'auteur numérote &lt;titre section&gt; par le marqueur d'ordre &lt;marqueur 2.B&gt;</i>
Associer	La relation d'association est une relation <i>abstraite</i> dans le sens où elle dénote uniquement du fait que deux objets sont en relation, sans en préciser la nature. Elle peut être utilisée pour décrire certains usages conventionnés de MFM : - deux objets textuels en vis-à-vis (c'est-à-dire disposés sur le même plan horizontal) sont associés. C'est le cas de l'encadré rouge à gauche de cette ligne, des notes en marge, des renvois,	<i>L'auteur associe &lt;encadré_ECO&gt; et &lt;ligne_L0&gt;</i>

Pour plus de détails voir...

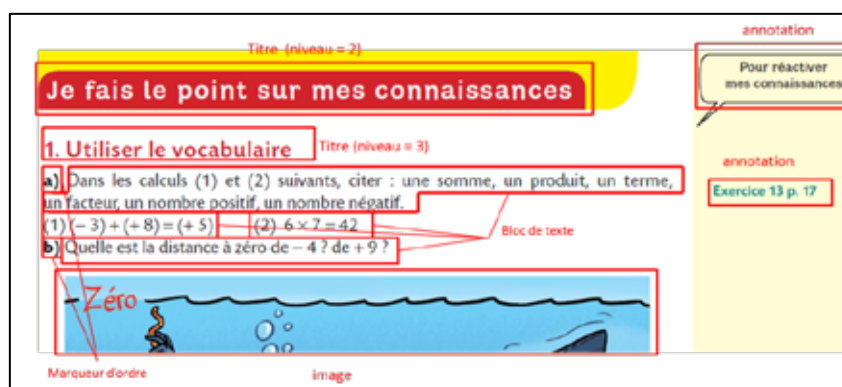
	<p>des cellules d'un tableau, etc.</p> <ul style="list-style-type: none"> <li>- deux objets proches (relativement à la distance modale<sup>84</sup> entre les segments de la page) disposés sur le même axe médian vertical sont associés (ex. légende d'un tableau ou d'une figure, cellules d'un tableau).</li> <li>- deux objets reliés par un objet graphique, comme une flèche ou un trait sont associés (ex. commentaires Word et segment commenté).</li> <li>- on pourrait aussi citer les notes de bas de page, qui associent un mot à un objet en bas de page par un marqueur matérialisant le lien.</li> </ul> <p>La liste n'est pas exhaustive mais permet de fournir une première annotation des relations particulières entre segments de MFM.</p>	
--	---	--

**Tableau 11 - Relations entre segments de MFM d'après les conventions partagées**

Les relations listées montrent que de nombreux liens et de nombreuses propriétés sont définies de manière implicite par les l'utilisation conventionnée de la MFM dans les textes. La liste de relations précédente n'est pas exhaustive et vise à permettre une base d'annotation automatisable. Il est important de préciser que ces relations peuvent se combiner pour un même segment ; par exemple un titre peut être numéroté, intituler la section subséquente, éventuellement être associé à une note en marge, etc.

### 4.3. Exemple de formalisation des architectures textuelles d'un document sur la base des conventions partagées de MFM

Ci-dessous un exemple d'identification des classes d'objets de MFM appartenant aux conventions partagées. Cette identification a été réalisée sur la base des marques de MFM c'est-à-dire les indices typographiques, dispositionnels et lexico-syntaxiques.



**Figure 28 - Exemple d'annotation des statuts de MFM d'après les conventions partagées**

De plus, les relations conventionnées définies dans la section précédente permettent de préciser la segmentation et de relier certains objets entre eux. Par exemple la présence de titres implique la création de *sections* qu'ils intitulent.

<sup>84</sup> Le mode est la valeur la plus représentée dans une série statistique.

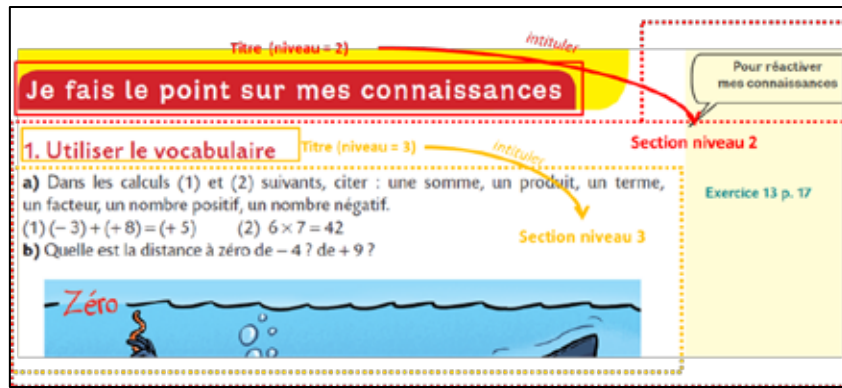


Figure 29 – Exemple d'annotation de la relation *intituler*

D'autres relations seraient ici produites : les deux blocs de texte sur la colonne de droite seraient *associés* à la section de niveau 3, les marqueurs d'ordres « a » et « b » *numérotent* et *labellisent* les blocs de texte subséquents, le marqueur d'ordre « 1 » fait de même avec le titre « Utiliser le Vocabulaire ».

#### 4.4. Conclusion

En conclusion, nous avons défini un ensemble de classes d'objets textuels, et de relations entre objets, relatifs à des conventions d'utilisation de la MFM partagées. Nous avons montré qu'il était possible de repérer ces conventions d'architectures textuelles à travers les marques de MFM qu'elles présentent, et qu'une fois identifiées leur sémantique pouvait être formalisée par un ensemble de métaphrases.

Sur la base de ces conventions et de leurs paramètres formels il est ainsi possible de produire de manière automatique une première formalisation de la MFM des textes, et d'extraire ainsi une partie de la sémantique des architectures textuelles correspondantes. Ce processus a été implémenté et testé à l'aide d'algorithmes d'apprentissage automatique ; les résultats sont reportés dans la partie 7. Cependant, cette phase de l'annotation présente deux limites qui sont liées à la diversité immense des phénomènes de MFM dans les textes : la méthode décrite dans cette partie est forcément sujette à erreur puisqu'elle est basée sur une liste d'indices et de conventions arrêtée, et pour les mêmes raisons, elle ne permet qu'une description relativement superficielle des informations véhiculées par la MFM, en laissant de côté toutes les architectures textuelles particulières qu'un auteur peut mettre en place. Ainsi, si l'on souhaite « interpréter » la MFM des textes le plus précisément et exhaustivement possible, et aussi pour corriger d'éventuelles erreurs de la première méthode, une deuxième phase d'annotation manuelle est nécessaire. Elle sera décrite dans la partie suivante et consiste à extraire les *conventions locales* de MFM.

## 5. Modélisation des conventions locales de MFM

Nous avons montré qu'il existait des conventions de MFM partagées de manière quasi-universelle ; nous les avons explicitées dans la partie précédente sous l'angle du MAT, c'est-à-dire en termes de statuts (classes) des objets textuels et de relations entre ces objets. La MFM d'un document donné suit *probablement* pour partie les conventions universelles de MFM, mais une grande partie des architectures textuelles véhiculées par la MFM du document ne peuvent être capturées par ces

dernières. En effet, en marge de conventions universelles de mise en forme, chaque genre de document, domaine de connaissances et ultimement chaque document peut présenter des conventions d'utilisation de la MFM qui lui sont propres, pouvant être en contradiction avec les conventions universelles ou au contraire en être des extensions, des cas particuliers. De plus les marques de MFM peuvent s'avérer être insuffisantes dans certains cas pour interpréter un objet textuel, qui ne prend son sens que lorsqu'il est considéré parallèlement à la sémantique du discours marqué. Ainsi, pour faire face à cette complexité nous proposons dans notre méthodologie de faire intervenir un ou plusieurs annotateurs humains, dont les rôles seront les suivants : 1) confirmer ou infirmer les statuts et relations « à priori » attribués aux objets textuels et 2) d'étendre l'annotation en précisant et/ou rajoutant de nouveaux statuts et relations aux objets.

Les conventions locales représentent donc, en théorie, la *sémantique réelle* de la MFM du texte, ou tout du moins ce qui s'en rapprocherait le plus ; autrement dit, il s'agit de retrouver le plus précisément possible les intentions auteuriales initiales qui ont donné lieu à la mise en forme du texte telle qu'elle apparaît, qui peut être issue de conventions du domaine, du genre de document, ou simplement d'intentions originales de l'auteur.

### **5.1. Méthode d'identification des conventions locales**

---

La méthode proposée ici permet à un annotateur d'étendre les informations déjà incorporées au document à l'aide des classes d'objets et relations relatives aux conventions partagées. Cet enrichissement vise à identifier les conventions locales de MFM, à savoir la sémantique des architectures textuelles.

Tout d'abord, il est important que l'annotateur possède des connaissances assez développées du domaine et du genre dans lesquels le document s'inscrit, mais aussi qu'il soit un lecteur expert de manière générale (on pourra prendre comme critère le seuil des 10 années de pratique, qui constitue un seuil empirique [Ericsson et al. 2006]). Ainsi l'annotateur expert devra enrichir le document, déjà annoté d'après les conventions partagées définies précédemment. L'annotation se déroulera en 3 étapes :

1. Recensement des objets textuels récurrents (principe d'identité): l'objectif est de recenser tous les phénomènes de MFM identiques en termes de marques typo-dispositionnelles et lexico-syntaxiques, ou à défaut ayant des traits communs ; on considérera ensuite un par un les phénomènes *uniques*. On obtient ainsi un ensemble de catégories de phénomènes qui comprennent chacun un certain nombre d'instances identiques ou très similaires. L'objectif de ce recensement est de faciliter le travail d'annotation ; à noter que cette étape peut être réalisée de manière automatique par comparaison des objets déjà identifiés.
2. Enrichissement de l'annotation : on commence par considérer un représentant d'objet par catégorie issue de l'étape précédente. Pour chaque représentant, on va ajuster et enrichir l'annotation déjà réalisée par application des conventions partagées, et ce en 3 étapes :
  - a. Modification de la segmentation : ajout, division ou suppression de segments en fonction des besoins du point « c ». À noter que la segmentation n'est, à ce stade de l'annotation, pas contrainte ; par exemple deux segments peuvent se recouper. On s'intéressera notamment aux objets textuels discursifs, marqués uniquement par des

indices lexico-syntaxiques, qui ont pu ne pas être identifiés par l'application des conventions partagées.

- b. Confirmation / infirmation des, statuts des objets textuels, attribuées par application des conventions partagées. En cas d'infirmation, l'annotateur pourra soit attribuer un autre statut des conventions partagées ou créer un statut dédié à l'objet examiné. Cette étape est réalisée conjointement avec le point « c ».
  - c. Ajout / suppression / modifications des propriétés et relations relatives à l'objet textuel, avec le formalisme des métaphrases. Les indices utilisés pour cette étape sont décrits ci-après. L'annotateur pourra dans un premier temps formuler en langage naturel ce qu'il pense être caractéristique du phénomène étudié ; dans un second temps ce développement est formalisé soit sous forme de propriété(s), s'il porte sur le segment en lui-même, soit sous forme de relation(s) s'il porte sur le lien entre le segment étudié et un autre. L'annotateur peut s'appuyer sur les propriétés et relations déjà définies dans le modèle des conventions partagées. Ces informations sont ensuite incorporées au document.
3. Application des nouvelles annotations à toutes les instances des catégories définies en 1 : on généralise l'annotation à toutes les instances de chaque catégorie recensée à l'issue de la première étape. Cette généralisation doit être menée conjointement avec une vérification que l'annotation s'applique bien à l'instance, et dans le cas contraire retourner à l'étape 2 pour annoter spécifiquement cette instance.
  4. Recensement des nouvelles classes d'objets textuels : cette étape d'annotation peut faire émerger de nouvelles classes d'objets, lorsque ces derniers présentent les mêmes propriétés annotées. Ainsi l'on peut recenser de nouvelles classes d'objets textuels en recherchant les ensembles identiques de propriétés, afin de les incorporer au modèle défini pour les conventions partagées. La généralisation d'un tel processus pourrait permettre de modéliser les conventions de MFM à plusieurs niveaux (voir 7 pour plus de détails).

Pour réaliser les étapes précédentes visant à enrichir l'annotation (point 2 précédent), plusieurs indices seront utilisés par l'annotateur conjointement avec ses connaissances du domaine abordé par le document, et ses connaissances des phénomènes de mise en forme :

- § L'éventuelle définition explicite des conventions de mise en forme de l'éditeur ou de l'auteur. Parfois, à la manière d'une légende sur une carte géographique, on retrouve au début ou en fin d'ouvrage des précisions sur l'utilisation de certains phénomènes de mise en forme. C'est d'ailleurs le cas pour ce manuscrit de thèse, où l'on donnait des *indices de lecture* en préambule (voir Préambule et indices de lecture). Une perspective est que ce type de définition explicite des conventions de MFM puisse être réutilisé de manière automatique dans le processus de formalisation.
- § L'annotation automatique des propriétés et relations fournie par le modèle des conventions partagées de MFM. Ces propriétés et relations donnent une première interprétation du ou des rôles joués par les objets textuels.
- § La sémantique du discours du texte renferme des indices pertinents pour extraire les informations véhiculées par la mise en forme. Nous avons vu précédemment que discours et MFM étaient fortement liés ; fond et forme se supportent mutuellement. Ainsi la sémantique du discours de l'objet textuel étudié, ainsi que la sémantique du discours environnant l'objet sur la page, permettent d'inférer les propriétés et relations de ce dernier.

§ Enfin, l'annotateur peut comparer les différentes instances des segments ayant une mise en forme identique (classés dans la même catégorie au point 1) à celle du segment étudié. Ces différents segments étant utilisés dans des contextes différents mais avec la même mise en forme, l'annotateur peut ainsi inférer par comparaison le rôle de ces segments dans le message de l'auteur.

Enfin, une perspective pour cette phase du processus d'annotation est de renforcer sa validité, en faisant appliquer la méthode décrite dans cette section à plusieurs annotateurs, en parallèle sur le même corpus. En effet il existe certaines méthodes utilisées dans le domaine du traitement automatique des langues, permettant de calculer des accords inter-annotateurs [Scott 1955] et de choisir une annotation finale (un *alignement*) à partir d'annotations divergentes [Mathet et al. 2011].

## 5.2. Exemple d'annotation des conventions locales

Ci-dessous un exemple des propriétés et relations attribuables manuellement aux segments à l'aide de la méthode décrite dans la section précédente.

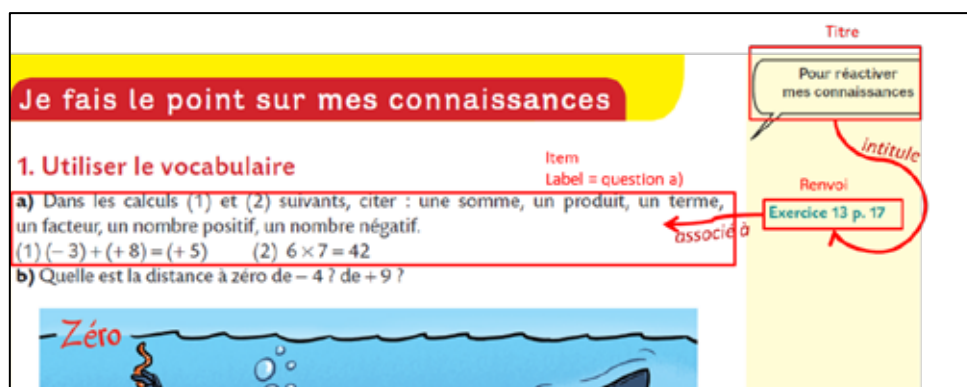


Figure 30 – Exemple d'annotation des conventions locales

Les deux segments sur la droite étaient initialement considérés comme appartenant à la classe « blocs de texte ». Il s'avère que le premier objet textuel est en réalité le titre de la colonne de droite, tandis que l'objet en dessous du premier est, d'après sa forme syntaxique et son contenu, un renvoi. D'après la position de ce dernier et d'après le contenu de l'exercice 13 p.17 on en déduit que le renvoi est associé à l'item a) de l'énumération, qui est habituellement appelé dans les manuels scolaires une question, en l'occurrence la « question A ». Son statut d'item a été déduit du fait qu'un autre marqueur d'ordre suivait le « a) ».

## 5.3. Limites de l'approche

L'annotation est à ce stade limitée par les connaissances et capacités des annotateurs à formaliser la mise en forme. De plus le processus peut être particulièrement long sans outil ad-hoc, étant donné que l'écriture de fichiers balisés doit se faire manuellement (pour plus de détails sur l'implémentation de l'annotation voir 6.2). Plusieurs étapes de cette phase d'annotation manuelle sont cependant automatisables.

Enfin, une manière de capitaliser l'annotation manuelle des conventions de MFM est de greffer les classes d'objets et leurs propriétés relevées au modèle des conventions partagées défini. Cependant

pour que ce processus soit réalisé de manière optimale, il faudrait définir différents moyens de résoudre les conflits lors de l'ajout de nouvelles classes, afin notamment d'éviter les doublons. De plus il faudrait pouvoir attribuer aux conventions locales relevées un certain degré de généralisation (ex. domaine, genre, sous-genre donné ?) afin de savoir quand est-ce que leur réutilisation est pertinente. Cela reste à l'état de perspective dans le cadre de ces travaux.

## 6. Vue d'ensemble du processus d'annotation et opérationnalisation technique

Cette partie reprend le processus de formalisation des architectures textuelles dans sa globalité, et détaille ensuite différentes réflexions sur les techniques et les technologies existantes pour opérationnaliser la formalisation pour des documents numériques.

### 6.1. Schéma global

Le schéma ci-dessous décrit l'ensemble du processus d'annotation appliqué à un document. En vert sur la colonne de gauche on retrouve les méthodes et outils utilisés pour l'annotation, et en orange sur la colonne de droite les indices de mise en forme et du discours utilisés par ces méthodes pour réaliser l'annotation.

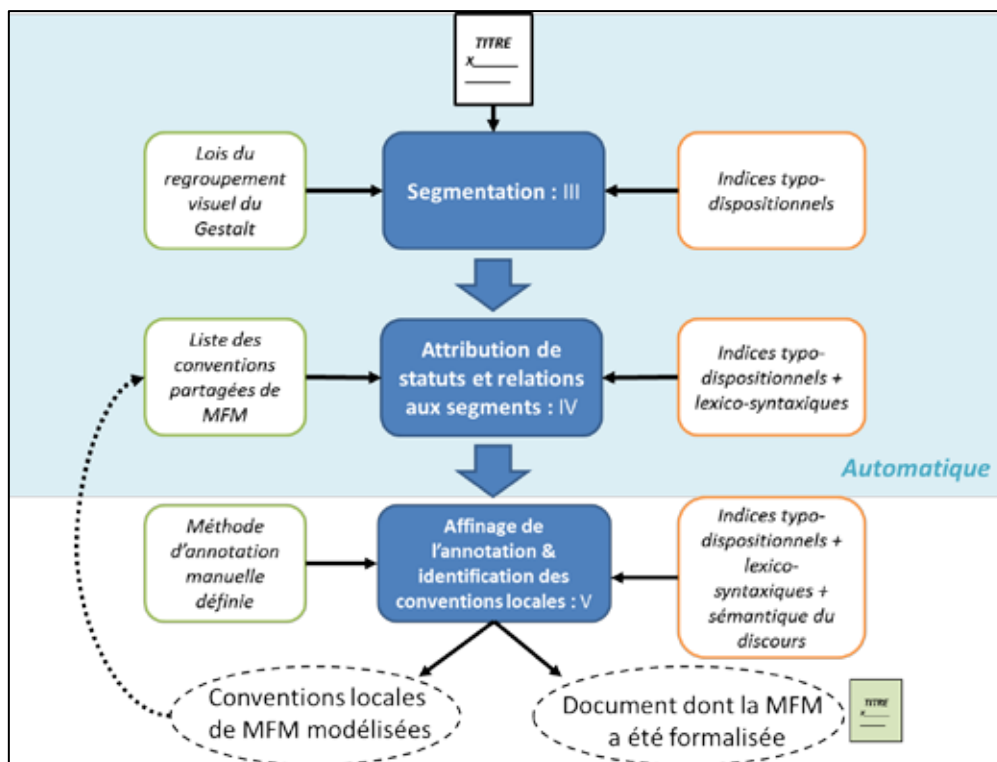


Figure 31 - Schéma global du processus d'annotation de la MFM d'un document

On retrouve les 3 étapes principales de l'annotation : segmentation, identification des statuts (classes d'objets textuels) et relations sur la base des conventions partagées de MFM, puis la complétion de l'annotation par extraction des conventions locales ; les deux premières sont réalisables automatiquement et la troisième s'effectue manuellement. Cette méthodologie vise un



compromis entre précision et complétude de l'annotation, et son caractère automatique. L'idée est également de rendre le processus de plus en plus automatique en intégrant la modélisation manuelle de conventions (locales) à la liste de conventions partagées ; cependant un modèle plus élaboré pour permettre l'agrégation des conventions doit encore être développé, point qui sera discuté dans la partie 7.

Ce processus d'annotation a été utilisé pour modifier les documents pour lesquels la restitution de la MFM a été évaluée auprès d'utilisateurs voyants et non-voyants.

## **6.2. Implémentation informatique du processus de formalisation**

---

Cette section vise à décrire les contraintes et possibilités techniques pour appliquer la méthodologie d'annotation en conditions réelles, et notamment les étapes automatisables du processus. On abordera la question du format des fichiers en entrée de l'annotation, les contraintes en termes d'informations à incorporer aux documents et quels formats de fichiers s'y prêteraient (format en sortie de l'annotation), ainsi que les développements informatiques à réaliser pour permettre l'annotation automatique et manuelle.

Les seules contraintes pour le fichier en entrée de l'annotation sont la numérisation préalable du texte (des caractères), pour pouvoir faire référence au contenu textuel dans le métatexte, et la conservation de sa forme visuelle, pour pouvoir interpréter sa MFM. La disponibilité du texte est importante pour permettre d'annoter sans ambiguïté des segments textuels et pour permettre d'y relier le métatexte ; la disponibilité de la forme visuelle est essentielle car c'est ce que l'on désire formaliser : la MFM d'origine du texte. Ainsi l'annotation peut se baser sur n'importe quel format de fichier dont le texte est mis en forme, sous réserve de pouvoir utiliser avec succès un logiciel de reconnaissance de caractères (*OCR*).

On distingue alors deux cas pour les formats de fichiers en entrée, qui vont engendrer deux approches de segmentation des phénomènes de MFM (première étape de l'annotation) : (1) soit les propriétés visuelles sont disponibles explicitement dans le document (ex. fichier HTML avec feuilles de style), (2) soit elles ne le sont pas et la mise en forme du document n'est exprimée que visuellement. (1) Dans le premier cas, un algorithme de restructuration du code doit être mis en place afin que le fichier ne relate plus que des objets textuels, et ainsi supprimer ou fusionner toutes les structures du fichier liées au contexte technique. Dans le cas d'un fichier HTML cela consisterait à le transformer en un ensemble de divisions (<div>) en conservant leurs propriétés visuelles (styles associés). Ainsi il faudrait développer un algorithme spécifique à chaque format de fichier pour obtenir un fichier segmenté en sortie. (2) Dans le cas d'un fichier « image », il faudrait développer un algorithme de reconnaissance de formes (ou réutiliser un algorithme ou des briques fonctionnelles existantes), afin de segmenter les zones de textes contrastées (objets textuels), puis d'associer ces zones aux unités textuelles qu'elles englobent à l'aide de la correspondance entre les coordonnées des zones segmentées et les coordonnées du texte sur l'image. Pour éviter le développement de multiples algorithmes afin de prendre en compte différents formats de fichier, une possibilité est de transformer les fichiers (quel que soit leur format) en images à l'aide de techniques de scan d'écran pour pouvoir utiliser un seul et même algorithme de traitement d'image pour segmenter les phénomènes de MFM.

Quel que soit le format d'origine du fichier, il faut que le format de fichier du document dont on veut formaliser la MFM puisse intégrer les types d'information suivants : le début et fin d'un objet textuel, le statut et les propriétés de chaque objet (ex. « titre » et son niveau hiérarchique), ainsi que les relations entre objets (ex. <OT1> et <OT2> composent <OT3>). Les langages de balisage semblent ainsi tout désignés pour permettre l'annotation de la MFM ; les balises permettent de délimiter le début et la fin des objets, le nom de la balise en elle-même permet d'annoter les statuts, et les attributs des balises font de même pour les propriétés et identifiants. Les relations peuvent être définies par une balise particulière dont plusieurs attributs pointeront l'identifiant des objets initiant la relation, des objets cibles de la relation, ainsi que l'identifiant de la relation (à la manière de triplets RDF<sup>85</sup>), par exemple : `<relation id="R1", cibles="OT3" origine="OT1;OT2" verbe=composer>`. Une autre possibilité pour décrire les relations dans un langage de balises serait d'incorporer dans les attributs des balises correspondant aux objets textuels (initiant les relations) les éventuelles relations et segments cibles (de la même manière que pour l'ajout de métadonnées au standard ARIA, voir Chapitre II4.3.b).

Ainsi le choix du langage XML pour lequel il sera défini un ensemble de balises et d'attributs ad-hoc nous semble le plus pertinent ; nous avons défini un vocabulaire de balises et d'attributs permettant de décrire les documents utilisés pour les évaluations rapportées au chapitre suivant (voir Chapitre V4.5). Des fichiers XML décrivant les architectures textuelles de documents pourraient tout à fait être incorporés à des standards Web (HTML étant un sous-langage XML) ou au standard EPub qui utilise d'ailleurs des fichiers XML pour les documents. Dans cette optique on pourrait même envisager d'intégrer l'annotation des architectures textuelles à un fichier HTML, non pas à l'aide de nouvelles balises mais uniquement à l'aide d'attributs de balises existantes qui n'ont pas de sémantique particulière pour la structure logique des documents, comme la balise <div> qui marque une division *abstraite* du document (marquée visuellement ou non).

L'annotation automatique des conventions partagées nécessitera deux entités distinctes : un fichier formalisant ces conventions qui constituera une *base de règles*, et un algorithme permettant l'application de cette base de règles à un fichier segmenté, afin notamment de calculer les différentes métriques utilisées pour attribuer les statuts et relations des conventions globales (ex. taille de la police modale du document). La base de règles devra spécifier les marques de MFM permettant d'identifier chaque classe d'objets textuels et chaque relation entre objets, mais aussi le contenu à incorporer au fichier du document pour son annotation. Enfin, l'annotation manuelle ne nécessite qu'un éditeur de texte et l'utilisation de la méthode définie dans la partie 5. Cependant, pour la rendre plus efficace un outil graphique d'annotation pourrait être développé ou adapté à partir d'un outil existant, comme ceux que l'on retrouve pour l'annotation de corpus de document en traitement automatique des langues (ex. la plateforme Glozz<sup>86</sup>). Cet outil pourrait notamment faciliter les phases répétitives de l'annotation manuelle en les automatisant, par exemple l'attribution d'identifiants uniques aux objets textuels segmentés ou l'application d'une propriété à un ensemble d'objets.

En résumé, l'implémentation informatique de notre méthodologie d'annotation impliquerait de réaliser les développements suivants :

---

<sup>85</sup> *Resource Description Framework* est un modèle de graphe dont l'implémentation prend la forme de triplets (sujet, prédicat, objet), qui est notamment utilisé pour associer des métadonnées à des ressources Web

<sup>86</sup> <http://www.glozz.org/>

- § Un algorithme de segmentation des phénomènes de MFM, que l'on pourrait adapter de l'existant (par exemple du logiciel LA-PDF [Ramakrishnan et al. 2012] qui permet de segmenter les zones de texte des documents PDF).
- § Une base de règles formelles pour l'annotation automatique des conventions partagées, règles qui permettraient de classifier les segments et de les mettre en relation en fonction des indices typo-dispositionnels et lexico-syntaxiques qu'ils présentent.
- § Un algorithme permettant d'appliquer cette base de règle à un document segmenté.
- § Éventuellement un outil graphique d'annotation qui permettrait d'accélérer le processus d'extraction des conventions locales, et de surcroît permettrait une mise en commun des annotations les plus efficaces afin d'étendre les conventions partagées.
- § Un standard XML (ensemble de balises et règles de construction du fichier qu'on appelle une *DTD* pour *Document Type Definition*) pour la formalisation de la MFM, afin de décrire les différentes informations évoquées précédemment.

De tels développements permettraient la génération de documents dont la MFM serait formalisée ; avec un niveau d'automatisation suffisant, l'on pourrait imaginer effectuer les différentes opérations de l'annotation sur n'importe quel document entrant, par exemple dans le cadre de l'accès à des pages Web. L'intérêt d'une telle démarche sera argumenté à travers les résultats évaluant la restitution de la MFM dans le Chapitre V. Dans le cadre de notre expérimentation, nous avons réalisé manuellement la segmentation et l'annotation de la MFM des documents utilisés, en se basant sur un standard XML ad-hoc (voir Chapitre V4.5).

Ce chapitre montre qu'il est possible d'automatiser une partie de la formalisation de la MFM des documents. Cependant le point central de cette formalisation, qui conditionne également sa propension à être automatisée, réside dans le recensement et la modélisation des conventions partagées. La modélisation gagnerait à être réalisée sur la base d'un modèle ontologique<sup>87</sup> commun, auquel pourront être rattachés : les classes d'objets de MFM et leurs relations (père-fils par exemple), leur propriétés comme celles décrites dans la partie 4.2.a, les documents et genres de documents dans lesquels on retrouve les objets, et surtout les indices visuels et lexico-syntaxiques les caractérisant. Un tel modèle permettrait d'unifier les différentes parties du modèle des conventions globales, mais permettrait surtout de diffuser le modèle et de faire en sorte que différents annotateurs puissent collaborer à son développement. Un tel modèle ontologique représente une étape importante pour que la formalisation de la MFM puisse se généraliser, et puisse être la plus précise et exhaustive possible ; son développement sera donc prochainement étudié. Dans cette optique, on présente dans les deux paragraphes ci-après deux modèles ontologiques existants, qui ont été créés dans afin d'améliorer l'accessibilité des pages Web, qui pourraient être réutilisés comme base pour le modèle ontologique décrit, ou auxquels pourraient se greffer un tel modèle.

[Fayzrakhmanov et al. 2010] ont développé un modèle unifié pour la description de pages Web dans le but de proposer des chemins de navigation à des utilisateurs non-voyants, basés sur des *modèles mentaux* de la navigation [Baumgartner et al. 2010]. Ce modèle permet à la fois de décrire les propriétés visuelles des segments, de leur attribuer des statuts décrits dans le langage HTML

---

<sup>87</sup> Une ontologie en Informatique est une structure représentant des connaissances. Une ontologie regroupe un ensemble de concepts ainsi que les relations (au sens large) entre ces concepts. (ex. un « homme » est un « individu »). En ce sens ce sont des « spécifications de conceptualisations » [Gruber 1995].

(tableau, colonne, lien, etc.), et de les lier au texte et aux propriétés des balises correspondantes. Ce modèle a pour vocation de permettre la navigation selon différents axes : sémantiques (concepts connexes), topologiques (régions proches) et spécifiques au domaine (objets HTML). Son pouvoir descriptif reste cependant limité ce qui concerne la mise en forme étant donné qu'il ne permet pas de décrire une hiérarchie de classes d'objets, ni les relations entre ces classes.

Également dans une optique d'accessibilité des contenus, [Harper and Yesilada 2007] ont développé une ontologie nommée *WAfA (Web Authoring for Accessibility)*, qui présente elle une typologie plus fine des objets retrouvés sur les pages Web organisée en *regroupements* et *objets atomiques* ; elle présente aussi une typologie des rôles des objets retrouvés sur les pages web (ex. aide à la navigation, point de repère, alerte, obstacle, etc.). Leur objectif était de pouvoir automatiquement restructurer les pages Web en fonction de leur modélisation à l'aide de *WAfA*, afin d'en générer des versions plus accessibles [Yesilada et al. 2007].

La création et le développement d'un modèle ontologique est une perspective pour de futurs travaux. La partie suivante présente des travaux réalisés en collaboration avec un autre doctorant, qui ont exploré la formalisation automatique de la MFM des textes à l'aide d'algorithmes d'apprentissage.

## **7. Vers une formalisation automatique de la MFM par apprentissage automatique**

---

---

Cette partie vise à présenter un projet réalisé en collaboration avec un autre doctorant travaillant dans le domaine du traitement automatique des langues ; ces travaux visaient à reconnaître automatiquement les objets textuels d'un ensemble de documents, et à relier ces objets entre eux par des relations de composition (voir 4.2.b). Cela correspond à une première implémentation de notre méthodologie de formalisation des architectures textuelles, en nous limitant ici à certains statuts d'objets textuels et à une relation particulière.

Cette première implémentation réalisée avec Jean-Philippe Fauconnier [Fauconnier et al. 2014], ancien doctorant au laboratoire IRIT, reprenait les deux grandes étapes de formalisation définies dans notre méthodologie, à savoir la *segmentation* des objets textuels puis l'extraction des architectures textuelles. Les différentes étapes du projet étaient : à partir d'un corpus (ensemble de documents) d'articles scientifiques dont on a formalisé manuellement la MFM, entraîner un algorithme d'apprentissage sur une partie du corpus afin qu'il puisse identifier les objets textuels à partir de leurs marques de MFM ; on produit un modèle probabiliste associant la présence de certaines marques de MFM à celle d'objets textuels particuliers. L'étape suivante était d'appliquer ce modèle à une autre partie du corpus (réservée à l'évaluation) et de mesurer les performances de reconnaissances des objets textuels. Un autre modèle était produit d'une façon similaire afin de reconnaître automatiquement la composition entre objets textuels.

La première étape a donc été la création du corpus de documents. Le corpus était divisé en 2 sous-corpus : 25 articles scientifiques de linguistique issus des actes du CMLF<sup>88</sup> 2008 (corpus *LING*), et 21

---

<sup>88</sup> CMLF : Congrès Mondial de Linguistique Française.

rapports et articles de géopolitique de l'IFRI<sup>89</sup> (corpus *GEOP*) ; tous les documents étaient initialement des fichiers PDF dont le texte n'était pas numérisé (fichiers d'images). L'intérêt de distinguer ces deux corpus est qu'ils présentaient des caractéristiques différentes en termes de mise en forme : le corpus LING était très homogène (mise en forme conventionnée pour les actes d'une conférence), alors que le corpus GEOP comprenait des documents aux mises en forme très hétérogènes (voir Chapitre V3) pour des exemples de documents pour chaque corpus). L'ensemble des regroupements visuels de ces documents ont été segmentés à l'aide de l'outil LA-PDF [Ramakrishnan et al. 2012] (voir ci-après) puis leurs statuts annotés manuellement. La Figure 32 ci-dessous montre un extrait d'un document segmenté par LA-PDF, pour lequel on a rajouté le type (*type="paragraphe"*) du regroupement (*Chunk*), c'est-à-dire le statut de l'objet textuel.

```
<Page x1="37" y1="23" x2="507" y2="539" type="paragraphe" chunkCount="4" pageNumber="1" wordCount="19">
  <Chunk x1="37" y1="23" x2="253" y2="58" type="byline" id="1">
    <Word x1="37" y1="23" x2="126" y2="56" font="Arial" style="font-size:30pt">L'Ifri</Word>
    <Word x1="38" y1="25" x2="127" y2="58" font="Arial" style="font-size:30pt">est</Word>
    <Word x1="136" y1="23" x2="224" y2="56" font="Arial" style="font-size:30pt">en</Word>
    <Word x1="138" y1="25" x2="225" y2="58" font="Arial" style="font-size:30pt">France</Word>
    <Word x1="235" y1="23" x2="251" y2="56" font="Arial" style="font-size:30pt">le</Word>
    <Word x1="236" y1="25" x2="253" y2="58" font="Arial" style="font-size:30pt">principal</Word>
  </Chunk>
```

Figure 32 - Extrait du code source d'un document du corpus GEOP

Dans ces travaux nous nous sommes limités à la reconnaissance des statuts des objets textuels, représentés par les 8 labels suivants : titre, paragraphe, item, citation, en-tête, pied de page, *byline* (informations relatives aux auteurs dans les articles scientifiques) et référence bibliographique. Ces statuts comprennent une partie de ceux définis par les *conventions partagées* de MFM modélisées dans ce chapitre, et une autre était spécifique au genre de documents du corpus (des articles scientifiques). Les objets textuels segmentés étaient, en plus de leurs statuts, annotés en fonction de leur composition : relations père-fils de subordination entre objets, et relations de coordination entre objets de même niveau.

L'étape suivante a été, en utilisant la partie du corpus de documents dédiée, de générer un modèle probabiliste pour la labellisation des regroupements ; on générerait également un autre modèle pour la construction de l'arbre hiérarchique des relations de composition entre objets textuels. Pour générer le modèle utilisé pour la labellisation, nous avons utilisé une méthode en *champs aléatoires conditionnels* (CRF pour *Conditional Random Fields* [Lafferty and McCallum 2001]), qui est une classe de méthodes non-supervisées utilisées en apprentissage automatique, permettant de modéliser la probabilité d'associer à une observation  $x$  sa probabilité d'appartenance à un label  $y$ , notée  $p(y|x)$ . Ces probabilités sont calculées en fonction de la séquence d'observation du document, et en fonction de traits visuels (marges, polices, typographie, position, surface) absolus (ex. présence de puce, de tirets) et relatifs (ex. police plus petite que la taille de la police la plus représentée du document, typographie différente de la police la plus représentée).

Familles de traits	Traits	Exemples d'informations correspondantes
Traits locaux	Marges Polices Typographie Position Ratios	Indentation à droite ou à gauche, centrage des blocs, absence d'indentation. Présence d'emphases (gras ou italique), taille de la police, etc. Présence de puces, de tirets, de numérotation, « ; » ou « , » en fin de bloc. Position verticale dans la page (haut, bas) et horizontale (droite, gauche). Ratios de la surface sur la taille de la police, de longueur sur la largeur.
Traits de séquence	Bigrammes	Label attribué à l'unité qui précède dans la séquence du document.

<sup>89</sup> IFRI : Institut Français de Relations Internationales.

	<i>Start/end</i> <i>Contraste</i>	Présence du bloc en début ou en fin de document. Rupture avec le bloc qui précède (taille, type de police, indentation, etc.).
--	--------------------------------------	---

Tableau 12 – Marques de MFM et autres indices utilisés pour la labellisation des objets textuels (repris de [Fauconnier et al. 2014])

L'ensemble de ces traits ou indices ont permis de générer un modèle probabiliste pour la labellisation des objets textuels, en entraînant le *CRF* sur le corpus de documents. Un autre modèle pour l'annotation des relations de composition a été généré de la même façon. Le principe de construction de ce dernier modèle, basé sur un algorithme *shift and reduce* [Hernandez and Grau 2005], était de parcourir la séquence des objets textuels labellisés du document en considérant chaque paire d'objets dans la séquence, et en fonction de leur label et en fonctions des traits listés dans le tableau précédent, décider de la relation la plus appropriée à attribuer à la paire.

Les résultats de l'application de ces deux modèles au corpus de test (partie du corpus dédiée à l'évaluation) sont disponibles dans le Tableau 13. On y compare pour chacune des deux tâches, de labellisation et de reconnaissance des relations de composition, les résultats obtenus par les modèles générés en termes de taux de reconnaissance, avec le taux de reconnaissance d'une approche dite naïve ; cette approche pour la première tâche consistait à considérer tous les objets textuels comme étant des paragraphes (qui représentent la majorité des objets des deux corpus), tandis que pour la seconde l'approche naïve consistait à attribuer aléatoirement les relations de composition (subordination et coordination) entre objets textuels du document.

<i>Tâche</i>	<i>Résultats sur LING</i>	<i>Résultats sur GEOP</i>	<i>Résultats sur LING + GEOP</i>
Labellisation	87.18%	82.39%	80.46%
<i>vs. labellisation naïve</i>	32.33%	44.51%	37.33%
Reconnaissance des relations de composition	96.41%	98.45%	97.23%
<i>vs. reconnaissance naïve</i>	40.21%	41.03%	39.79%

Tableau 13 - Performances pour la labellisation des objets textuels et la reconnaissance des relations de composition (repris de [Fauconnier et al. 2014])

Le tableau précédent montre des taux de reconnaissance élevés, qui constituent des résultats encourageants, mais restent de l'ordre des taux de reconnaissances atteints par [Antonacopoulos et al. 2011] sur des documents plus hétérogènes que ceux de notre corpus, mais utilisant une classification plus simple (moins de statuts différents). Dans notre cas, les performances de la tâche de labellisation ont été limitées par les variations locales de certains objets qui peuvent être uniques (ex. un titre de niveau 4, le seul du corpus), mais aussi par le faible nombre d'occurrences de certains objets (ex. les citations représentaient 1.91% des objets textuels du corpus LING+GEOP). De plus 4.26% des objets textuels du corpus ne correspondaient à aucun des labels utilisés pour attribuer leurs statuts aux objets. Même si la majorité des objets textuels ont été correctement labellisés, un faible nombre de statuts incorrects peut être suffisant pour pénaliser la compréhension du contenu dans le cadre de son accès par un déficient visuel (en supposant que les statuts des objets textuels soient restitués). Une autre remarque plus générale, rejoignant les limites de la labellisation, concerne l'hétérogénéité de la MFM, qui rend impossible une reconnaissance systématique des architectures textuelles sans aucune erreur ; cette remarque rejoint la nécessité d'inclure un humain dans le processus, comme prévu par notre méthodologie, afin de contrôler l'annotation voire de la modifier/compléter.

Les taux de reconnaissance quasi-parfaits des relations de composition entre objets textuels proviennent du caractère très formel de ces dernières ; en effet la composition des objets textuels peut être modélisée par une grammaire assez simple, reposant sur la création de sections logiques à la suite de titres, et à l'identification des objets textuels visuellement imbriqués. Cette grammaire était ici largement respectée dans les documents.

En conclusion, ces travaux montrent qu'il est possible d'automatiser la reconnaissance d'objets textuels (même particuliers) et de relations jusqu'à un certain point ; ces résultats valident notre approche semi-automatique de la formalisation de la MFM. Il pourrait être intéressant pour de futurs travaux de reprendre les traits (indices) ayant été les plus performants pour la reconnaissance des objets textuels afin de les intégrer dans notre modèle des conventions partagées de MFM.

## **8. Conclusion**

---

---

Ce chapitre constitue une première étape vers une formalisation automatique (ou la plus automatisée possible) de la MFM des documents, et pose à ces fins les bases méthodologiques nécessaires, dans le cadre théorique du Modèle d'Architecture Textuelle (MAT). Nous proposons ainsi 3 grandes étapes pour formaliser la MFM d'un document : la segmentation des objets textuels, l'annotation des architectures textuelles à travers un modèle des conventions partagées de MFM, et enfin la modélisation des conventions (architectures textuelles) locales au document et leur annotation, puis leur intégration au modèle de conventions partagées. Les deux premières étapes sont fortement automatisables ; une implémentation à l'aide de techniques d'apprentissage automatique a d'ailleurs été évaluée avec succès sur un corpus d'articles scientifiques, avec des objectifs cependant plus réduits que la formalisation « universelle » de la MFM visée par ce chapitre.

Bien que cette méthodologie ait été appliquée pour formaliser (manuellement) la MFM des documents utilisés pour nos expérimentations, une évaluation de son pouvoir descriptif devrait être menée sur un corpus étendu de documents. De plus, certains développements seraient nécessaires à cette évaluation, tout d'abord des développements techniques pour automatiser l'annotation, afin de pouvoir évaluer l'ensemble des méthodes sur un grand nombre de documents plus rapidement ; mais aussi et surtout, développer un modèle ontologique unifiant les différents concepts relatifs à la MFM : statuts, propriétés et relations, et permettant de comparer et hiérarchiser les phénomènes de MFM selon ces axes, mais aussi en fonction des réalisations visuelles associées et des genres de documents et des domaines de connaissances (relatifs aux documents) pour lesquels ces phénomènes sont présents. Ainsi sur la base d'un tel modèle, la modélisation manuelle des conventions de MFM utilisées localement dans différents documents de différents genres pourrait être agrégée, et réutilisée au sein des conventions partagées pour servir une annotation automatique.

En ce qui concerne les limites de cette méthodologie, même dans un cas de figure idéal où des outils d'annotation automatique seraient développés et où les conventions partagées de MFM seraient enrichies, une phase de modélisation manuelle ou à défaut une phase de vérification de l'annotation semble nécessaire étant donné la quasi-infinité de possibilités quant à la mise en forme des documents, mais aussi à cause de l'ambiguïté du langage naturel et des marques de MFM. De plus, d'autres aspects des phénomènes de mise en forme devraient être considérés pour l'accès aux documents (au-delà des informations de MFM), par exemple dans le cas de page web le rôle des

objets pour la navigation hypertexte ou pour le remplissage des formulaires, en somme la portée contextuelle des objets distingués par leur MFM. Cela représente un autre travail de recensement des rôles des objets, qui a déjà été mené dans le cadre du web avec *WafA* [Harper and Yesilada 2007], et qui pourrait être utilisé conjointement avec les informations de MFM pour l'étude de l'utilisation et de l'accessibilité des documents. D'autres niveaux d'annotation pourraient également être pris en compte et mis à disposition en parallèle des informations de MFM, nécessitant cependant une nouvelle segmentation du document, comme la structure du discours et les relations rhétoriques entre segments, ou encore les rôles de macro-segments pour la résolution de tâches particulières (ex. les intentions pédagogiques structurant les manuels scolaires).

Le premier objectif de la formalisation de la MFM dans nos travaux, et donc de notre méthodologie, est de rendre les informations qu'elle véhicule disponibles pour être accédées dans des modalités non-visuelles. Il est important de noter que ces informations ne seraient pas disponibles autrement, les modèles actuels de métadonnées des documents numériques ne prenant pas cette dimension en compte. Ainsi la sémantique des architectures de l'exemple présenté dans la partie 5.2, et des exemples de phénomènes de MFM particuliers donnés au Chapitre 12.4 (comme des sous-titres, ou des types d'objets textuels particuliers définis pour désigner des types de commandes différentes en programmation), peut être rendu disponible grâce à la méthodologie ici définie. Techniquement parlant, les formats actuels permettent d'incorporer les informations relatives aux architectures textuelles (voir 6.2), qui sont issues de la formalisation. De futures études devront cependant explorer les possibilités d'automatisation de la formalisation, et la mise en place d'un modèle ontologique pour mettre en commun la modélisation de différentes conventions de MFM.

D'autres applications de cette formalisation sont envisageables, comme la détection d'incohérences dans l'utilisation des marques de MFM, l'explicitation des marques pour favoriser l'apprentissage de la lecture, ou encore de manière plus générale donner un nouvel angle d'analyse des documents et de leur sémantique pour le traitement automatique des langues et l'extraction de connaissances. L'extraction de connaissances et leur incorporation à des modèles ontologiques semble déjà être un domaine pouvant bénéficier des informations relatives aux architectures textuelles des documents [Kamel and Aussenac-Gilles 2008], ces informations représentant une partie de leur sémantique.



# Chapitre IV - Restituer la Mise en Forme Matérielle à l'oral : conception et opérationnalisation

Résumé : Une fois les architectures textuelles (et donc la MFM) d'un texte formalisées, se pose la question de la *restitution* de ces informations, c'est-à-dire la manière de les représenter pour les transmettre aux utilisateurs déficients visuels pendant l'accès au document. Les caractéristiques des modes non-visuels d'accès à l'information, présentées en Chapitre II.3.2, et le volume d'informations que représente le métatexte pour un texte à la MFM riche (voir exemple de formalisation en Chapitre V1), laissent penser qu'incorporer au texte d'origine l'ensemble du métatexte sous forme discursive (pour lecture ultérieure par synthèse vocale), n'aurait probablement aucun bénéfice pour les utilisateurs déficients visuels, et serait même sûrement pénalisant. L'objectif étant que la restitution de la MFM à l'oral puisse jouer les mêmes rôles de facilitateur de traitement du contenu, et de support fonctionnel pour la lecture, que ceux joués par la MFM à l'écrit. On tentera dans ce chapitre de répondre aux questions : que restituer des informations de MFM ? et Comment (et quand) les restituer ? On explorera donc les différentes modalités utilisables pour véhiculer les informations de MFM formalisées, ainsi que différentes heuristiques et réflexions pour concevoir une restitution tendant vers un optimum théorique pour une tâche donnée. On proposera ensuite une stratégie de restitution dédiée à la tâche de compréhension de textes expositifs<sup>90</sup>. Enfin, on présentera une plateforme logicielle que nous avons développée afin de facilement implémenter et évaluer des stratégies d'accès non-visuelles aux documents textuels.

## Sommaire du Chapitre IV

<b>1. Introduction</b> .....	<b>130</b>
<b>2. Typologie des informations de MFM à restituer</b> .....	<b>131</b>
<b>3. Vers une restitution optimisée de la MFM : étude globale de l'espace de conception</b> .....	<b>133</b>
3.1 Que restituer de la MFM : filtrage des informations.....	134
3.2 Comment restituer les informations de MFM : espace de conception.....	137
3.3 Quand restituer les informations de MFM ? .....	148
3.4 Restitution de la dimension fonctionnelle de la MFM .....	150
<b>4. Proposition d'une stratégie de restitution de la MFM</b> .....	<b>151</b>
4.1 Restitution des informations locales de MFM.....	152
4.2 Restitution des informations transverses de MFM .....	152
4.3 Commentaires sur la stratégie de restitution.....	153
<b>5. READ : une plateforme expérimentale pour implémenter et évaluer des stratégies d'accès non-visuelles aux documents</b> .....	<b>155</b>
5.1 Introduction : pourquoi cette plateforme ? .....	155
5.2 Présentation générale de READ .....	156
5.3 Formats et modalités utilisées pour restituer la MFM .....	162
5.4 Perspectives et limites.....	164

---

<sup>90</sup> Texte ayant pour but d'exposer, de faire comprendre quelque chose de manière objective.

6. Conclusion..... 164

## 1. Introduction

---

---

Nous avons vu dans le Chapitre I que la mise en forme matérielle (MFM) des textes jouait un rôle majeur dans la compréhension, la représentation mentale et la mémorisation du contenu des textes, et constituait surtout un vecteur d'informations et un support fonctionnel pour les lecteurs. De plus, étant donné que les technologies d'assistance actuelles ne prennent en compte la sémantique de la mise en forme que de manière très réduite, un des objectifs de nos travaux était d'incorporer cette sémantique à l'accès non-visuel aux textes. Notre approche est, par ailleurs, supportée par des résultats montrant l'intérêt de la restitution de certaines informations de MFM à l'oral [Lorch et al. 2012; Lemarié 2006b]. Afin de généraliser la restitution<sup>91</sup> à l'ensemble de la MFM et des architectures textuelles, ce chapitre présente (dans sa première partie) une réflexion globale portant sur les modalités non-visuelles utilisables et combinables pour la restitution, ainsi que sur les méthodes pour sélectionner et incorporer au contenu les méta-informations données par la MFM. Le but est de tendre vers une restitution optimale de la MFM en termes cognitifs pour une tâche donnée, qui puisse être exploitable en tant que facilitateur de traitement du contenu, comme c'est le cas à l'écrit avec les marques de MFM.

Une fois les informations de mise en forme matérielle formalisées et incorporées aux documents, de nombreuses questions se posent : faut-il les intégrer au texte d'origine sous forme verbale, ou au contraire en proposer des équivalents non-verbaux ? Faut-il d'office tout intégrer au contenu du document, et si non, quelles interactions proposer pour accéder à ces informations ? Ces questions sont au cœur de la problématique de *restitution de la MFM*, abordée dans ce chapitre. On peut aborder cette problématique tout d'abord en observant la forme des métaphrases, et surtout le volume de métaphrases généré par le processus de formalisation de la MFM (voir en Chapitre V1 pour exemple). En effet, un premier constat naïf est que tout restituer verbalement rendrait certes ces informations disponibles, mais aurait probablement un impact très négatif sur la charge cognitive en mémoire de travail durant la lecture et, probablement, sur la compréhension du document. L'objectif de ce chapitre est de développer une réflexion sur la restitution orale de la MFM axée sur les trois composantes de l'utilisabilité : l'efficacité, c'est-à-dire que l'information restituée doit permettre au lecteur DV de réaliser la tâche (dans notre cas, comprendre un texte expositif dans le cadre de l'apprentissage de connaissances déclaratives), mais aussi l'efficience et la satisfaction : l'utilisateur doit pouvoir réaliser la tâche en fournissant un effort cognitif raisonnable et sans trop rallonger le temps de lecture/compréhension. L'enjeu que nous assignons à la restitution n'est donc pas de restituer toute l'information relative à la MFM mais de restituer uniquement ce qui est pertinent pour réaliser la tâche. Ainsi, l'approche que nous défendons dans le chapitre s'inscrit dans une approche qualifiée d'*holistique* [Phipps and Kelly 2006; Giraud 2014], à distinguer de l'approche exhaustive associée à l'accessibilité normative pointée par [Giraud 2014]. Cette stratégie a été implémentée à l'aide de la plateforme READ que nous avons développée (voir 5) ; les résultats de l'évaluation de différentes stratégies de restitution sont présentés dans le Chapitre V.

Nous avons vu dans le chapitre précédent que les deux principaux modes utilisés par les déficients visuels pour accéder aux contenus numériques sont le mode audio, avec la synthèse vocale, et le mode tactile avec le braille via les plages braille dynamiques. Nous avons choisi dans nos travaux de

---

<sup>91</sup> Le terme *restitution* fait référence au fait que notre démarche consiste à redonner accès aux informations et fonctions offertes par la MFM à l'écrit, qui sont à l'heure actuelle pour la plupart non retranscrites.

nous restreindre au mode audio, et ce pour plusieurs raisons. Tout d'abord, la compréhension du langage oral est maîtrisée de façon beaucoup plus homogène chez les DV que celle du braille (voir Chapitre II3). De plus, la rapidité de transmission de l'information est supérieure à l'oral comparativement au braille ; l'écart semble d'ailleurs se creuser avec l'expertise : la vitesse de lecture d'un bon braille peut être de 2 à 3 fois inférieure aux vitesses d'écoute pratiquées par les synthèses vocales chez des utilisateurs experts. Enfin, le coût des plages braille dynamiques ne permet pas à tous les utilisateurs non-voyants de s'en procurer, alors que le coût des synthèses vocales est réduit, certaines étant même gratuites. L'utilisation conjointe de modalités orales et tactiles (ex. vibratoire) est discutée dans la partie 3, cependant les stratégies de restitution proposées et évaluées dans nos travaux n'utilisent que des modalités orales, ce qui nous a permis de restreindre l'espace de conception des stratégies, mais aussi de nous assurer que les stratégies proposées puissent être implémentées sur n'importe quel périphérique informatique possédant une sortie audio.

La restitution orale de la MFM relève d'un certain paradoxe comme le souligne [Maurel 2006] : en effet, le développement historique de l'écrit avec l'imprimerie marque le passage d'une culture de la langue orale (avec parfois une lecture orale des textes) à une culture écrite avec la « lecture silencieuse », pour laquelle des procédés de mise en forme du texte ont vu le jour afin d'en optimiser et d'en étendre l'utilisation. Le paradoxe est donc le suivant : « il s'agit de revenir à la lecture orale d'un texte exploitant des procédés qui ont probablement émergé grâce au fait que la lecture orale n'était justement plus de rigueur ! ». Au-delà de ce constat, le mode visuel et le mode audio présentent des différences fondamentales qui rendent complexe la transposition des informations visuelles de MFM à l'oral (voir Chapitre II3.2 pour plus de détails sur les caractéristiques des modes non-visuels) [Archambault 2010]. La séquentialité de l'accès à l'information orale et sa fugacité, opposée à l'accès global et la pérennité qu'offre le mode visuel, sont sources de surcharge cognitive et contraignent fortement la restitution. Nous verrons dans la partie 3 comment les contraintes du mode audio peuvent néanmoins être minimisées.

On rappellera dans la section suivante la nature et les différents types d'informations véhiculées par la MFM, afin de poser les bases de réflexion pour la restitution qui seront développées dans la partie suivante. Sur ces bases, on présentera ensuite une proposition de stratégie de restitution dans le cadre d'une tâche de compréhension d'un texte expositif. Enfin, on terminera ce chapitre par la présentation de READ (**RE**stitution de l'**A**rchitecture des **D**ocuments), une plateforme logicielle que nous avons développée, dédiée à l'implémentation et l'évaluation de stratégies d'accès non-visuelles aux documents textuels. Cette plateforme nous a permis de réaliser les évaluations présentées dans le dernier chapitre.

## 2. Typologie des informations de MFM à restituer

---

Nous avons vu dans les chapitres précédents que les architectures textuelles d'un document pouvaient être formalisées à l'aide de métaphrases, qui sont des équivalents discursifs aux marques typo-dispositionnelles et lexico-syntaxiques de MFM [Virbel 1989]. Chaque métaphore décrit une *intention* de l'auteur d'informer le lecteur quant à l'organisation du texte, intention exprimée dans le texte à travers une ou plusieurs marques de MFM, exprimant ainsi des *architectures textuelles*, c'est-à-dire des objets textuels qui sont en relation. On rappelle que les objets textuels sont définis comme

des segments de texte rendus perceptibles, discernables du reste du texte, par un jeu de contraste de leurs propriétés typographiques, dispositionnelles ou lexico-syntaxiques<sup>92</sup>. Les métaphrases peuvent donc contenir des informations de différentes natures et portant sur un ou plusieurs objets textuels ; au vu de cette diversité, il est ainsi important de classer les métaphrases afin d'en proposer une restitution pertinente.

Nous avons choisi de séparer dans notre classification des métaphrases celles donnant des informations **locales** à un objet textuel, de celles donnant des informations **transverses** à plusieurs objets textuels. Ce choix a été guidé par la séquentialité du mode audio et plus généralement des modes non-visuels d'accès à l'information. En effet, cette séquentialité transforme un texte écrit et sa MFM en une séquence linéaire d'informations qui suit l'ordre de lecture visuelle, rendant plus difficile et coûteux (du point de vue de l'auditeur en termes cognitifs) de *faire référence* à des informations éparpillées dans la séquence qu'à des informations contiguës. Or il est nécessaire de pouvoir *faire référence* aux objets textuels afin de restituer les *méta*-informations de MFM, qui de par leur nature *portent sur* certains objets. On reprend donc ci-après les différents types d'informations donnés par les métaphrases, déjà évoqués dans le chapitre précédent, en les classant en fonction de leur caractère *local* ou *transverse*.

Concernant les informations *locales* données par les métaphrases, nous proposons de distinguer :

- § La démarcation d'un objet, qui est la propriété fondamentale de tout regroupement visuel marqué par un contraste. Elle implique que l'on distingue le début et la fin de l'objet du reste du contenu du texte.
- § Le statut d'un objet, qui rejoint la notion de démarcation puisqu'elle dénote du fait que l'objet possède un type particulier, qu'il appartienne à une classe d'objets donnée, et marque donc un potentiel changement de type ou tout du moins souligne le changement d'objet et donc sa démarcation du reste du texte. On peut citer comme statuts d'objets textuels, dont les plus répandues ont été recensées au Chapitre III dans un modèle des conventions partagées de MFM, les classes « titre », « énumération » ou encore « emphase ». Ces statuts permettent donc de typer les objets ; chaque type peut engendrer certaines relations (voir la liste des informations transverses ci-après) ou encore certaines propriétés pour l'objet typé.
- § Les propriétés d'un objet, qui désignent de manière générique toute information locale qui lui est attribuable, autre que son statut et ou sa démarcation. On peut par exemple citer les propriétés engendrées par des relations entre objets (ex. la relation « labelliser » va donner une propriété de label à un objet) ; d'autres peuvent être issues de la présence de marqueurs d'ordre (ex. objet numéroté) ; d'autres encore peuvent être dérivées des statuts attribués aux objets (ex. un titre aura une propriété de *niveau hiérarchique*). Toute information locale à un objet textuel peut ainsi être qualifiée de propriété ; nous avons cependant traité séparément deux propriétés des objets qui sont la démarcation et les statuts des objets, à cause de leur universalité. La restitution des autres propriétés locales aux objets devra être étudiée au cas par cas.

Enfin, concernant les informations *transverses* données par les métaphrases, portant sur plusieurs objets textuels distincts, on distingue :

---

<sup>92</sup> Les objets textuels peuvent être imbriqués : un objet textuel peut être discernable et en son sein d'autres sous-objets discernables (ex. un paragraphe contenant une emphase).

- § Les relations qui lient au moins 2 objets textuels différents à travers un verbe performatif comme le verbe « intituler ». Étant donné que les relations ont une direction, c'est-à-dire qu'elles présentent un ou plusieurs objets d'origine et un ou plusieurs objets cibles, il faut également considérer les relations *réciproques* dans la restitution (« être intitulé par » serait la relation réciproque « d'intituler »).
- § Les informations organisationnelles données par certaines métaphrases, qui décrivent de manière transverse la structure du texte ou de certains objets textuels composés. Par exemple « L'auteur organise son texte en 3 chapitres », ou encore « L'auteur organise <énumération0> en 3 items <it1>, <it2> et <it3> ».

Distinguer les informations locales des informations transverses permet une première approche de la restitution de la MFM. En effet, les informations locales peuvent être retranscrites au fil de la lecture (non-visuelle), en fonction de l'objet abordé ; ce qui ne veut pas dire qu'elles doivent l'être systématiquement, et ne présage en rien non plus de la manière appropriée de les restituer (ce qui sera discuté dans la partie suivante). En revanche, en raison de la nature séquentielle de l'accès non-visuel à l'information, la restitution des informations transverses est moins évidente. Par exemple, faut-il restituer l'ensemble des relations et leurs réciproques ? Comment faire référence aux objets textuels pointés par ces relations, comment identifier les objets ? Comment faire en sorte de ne pas contraindre la navigation dans le document, et/ou perturber la représentation du contenu ? Autant de questions qui seront abordées dans la partie suivante, dans laquelle on reprendra cette distinction entre informations locales et transverses ; cela ne représente qu'une manière de classer les informations de MFM d'un document formalisé, qui nous semble pertinente pour analyser les possibilités de restitution.

Enfin, une autre dimension de la MFM que nous ne pourrions que partiellement aborder ici étant donné son ampleur, est la dimension *fonctionnelle*. Comme évoqué dans le premier chapitre, cette dimension relève du rôle des marques de MFM pendant la lecture visuelle et dans la compréhension, et décrit donc un ensemble de *fonctions* de lecture qui émergent grâce aux propriétés du système visuel humain. Nous en avons présenté quelques-unes au Chapitre I ; nous proposerons dans la section 3.4 une méthode de restitution de certaines de ces fonctions.

### **3. Vers une restitution optimisée de la MFM : étude globale de l'espace de conception**

---

Dans cette partie, nous proposerons un ensemble de critères et de méthodes pour restituer les informations de MFM de documents formalisés (à l'aide de la méthodologie d'annotation définie dans le Chapitre précédent), de la manière la plus optimale (pertinente) possible dans le cadre de leur accès non-visuel. L'objectif est donc de restituer ces informations (les retranscrire, les rendre disponibles) initialement visuelles, de façon compatible avec les objectifs de la tâche pour laquelle le document est utilisé et avec les contraintes de l'architecture cognitive humaine. Centrer la restitution sur la tâche à réaliser avec le document nous semble être une approche pertinente pour aborder la complexité et le volume d'information et fonctions de MFM existantes ; c'est d'ailleurs une des approches utilisées en ergonomie et notamment en ergonomie des interfaces avec le *user-centered design* [Norman and Draper 1986; Abras et al. 2004].

La question de la restitution de la MFM sera traitée dans cette partie à travers 3 sous-questions qui constitueront les 3 sections suivantes : quelles informations restituer ? Comment les restituer ? Et enfin quand les restituer pendant la lecture ? Les méthodes et critères donnés en réponse à ces questions visent à définir ce que l'on appelle ici des *stratégies* de restitution qui tendent vers un optimum théorique pour une tâche et un document donné. À notre connaissance, aucun précédent n'existe dans la littérature ; la thèse de [Maurel 2004] traite cependant de l'oralisation des structures énumératives sous l'angle du MAT, mais pas sous un angle cognitif et se limite à ce type d'objets. Nous verrons cependant comment ces travaux contribuent à l'élaboration de stratégies de restitution.

Les 3 sections suivantes traitent donc de : *que* restituer de ces méta-informations, *comment* restituer les informations de MFM présentées dans la partie précédente, et enfin *quand* restituer pendant la lecture.

### **3.1. Que restituer de la MFM : filtrage des informations**

---

Les informations de MFM peuvent permettre de faciliter les inférences pendant la lecture ou de désambiguïser certains contenus, mais leur restitution peut aussi dégrader les performances et amener une charge cognitive supplémentaire importante si la restitution n'est pas réalisée de manière pertinente ; on suppose que retranscrire l'ensemble des méta-informations serait pénalisant (voir exemple de texte modélisé en Chapitre V1). Cela serait d'autant plus problématique que les modes non-visuels contraignent déjà fortement l'accès à l'information de par leurs caractéristiques. De plus, les (méta)informations de MFM annotant les segments de texte permettent d'évaluer leur pertinence pour réaliser une tâche donnée ; par exemple, les notes de bas de page peuvent s'avérer d'une importance moindre à d'autres segments pour une tâche d'exploration du contenu. En ce sens la MFM donne des éléments pour filtrer les contenus non-pertinents, ce qui permet de réduire la charge cognitive [Giraud 2014] en les éliminant de la restitution. Ainsi, cette section vise à dresser un panorama des critères de sélection des informations de MFM à restituer, et donc des critères pour choisir lesquelles supprimer, mais aussi des critères permettant de filtrer le contenu primaire du texte sur la base de la MFM.

L'objectif du « filtrage » des informations de MFM formalisées est de minimiser la charge cognitive nécessaire à leur traitement ainsi qu'à celui du contenu tout en maximisant la compréhension, et en permettant ainsi à l'utilisateur d'atteindre avec la plus grande efficacité possible les objectifs de la tâche pour laquelle le document est utilisé. Cette problématique revêt en fait d'une double contrainte qui est que la MFM permet de faciliter le traitement des contenus et dans certains cas de le désambiguïser, mais que restituer des informations redondantes et/ou non-pertinentes peut augmenter significativement la charge cognitive [Giraud 2014], d'autant que les modes non-visuels ne permettent pas de filtrer l'information entrante. Le dernier auteur cité évoque d'ailleurs le fait que les non-voyants cherchent justement à extraire ce qui est important dans le discours oral qui leur est restitué. Dans ce cadre, on propose ci-après un ensemble d'heuristiques applicables aux informations de MFM afin d'évaluer la pertinence de leur restitution. Au-delà d'un filtrage binaire (restituer/supprimer), les informations de MFM jugées moyennement pertinentes peuvent être rendues accessibles à travers des commandes utilisateurs, ou alors restituées sous une forme la moins saillante et développée possible, par exemple à l'aide d'un marqueur discursif. Ainsi

l'ensemble des heuristiques présentés ci-dessous, portant sur les métaphrases formalisant la MFM d'un document, concernent à la fois ce qu'il faudrait restituer et comment le restituer.

► Heuristique 1 : *Dans quelle mesure l'information donnée par la métaphrase est-elle accessible via le contenu discursif du texte d'origine ?*

C'est-à-dire, sans ces méta-informations, est-ce que le contenu discursif en lui-même est suffisant pour les inférer ? Par exemple, sans expliciter le statut d'item d'énumération d'un objet textuel, on se pose la question de savoir si le discours associé à l'objet, et des objets textuels précédents et/ou suivants, permettent d'inférer cette information aisément. À noter l'absence de réponse unique à cette question car elle dépend directement de la littéracie du lecteur, de son habitude et de ses capacités de représentation et d'interprétation des textes écrits. On peut cependant objectiver cette heuristique en dressant une liste d'indices du discours renvoyant à l'information donnée par la métaphrase (ex. « premièrement » pour un item).

Cet heuristique rejoint quelque peu le critère de *redondance* de l'information ; si la métaphrase donne une information redondante pour le lecteur, car déjà inférée sur la base du discours présent dans le texte, alors sa restitution peut potentiellement augmenter la charge cognitive et/ou engendrer des frustrations.

► Heuristique 2 : *Quel degré d'ambiguïté engendre l'absence de l'information donnée par la métaphrase pour l'interprétation du message ?*

L'absence de MFM peut engendrer des ambiguïtés quant à l'interprétation du message de l'auteur. Par exemple, retirer la mise en forme d'un titre peut induire l'interprétation en erreur. On évaluera donc l'importance de la ou les métaphrases pour l'interprétation du sens du message, et notamment pour l'organisation, la structuration du message qui est une des vocations premières de la MFM.

► Heuristique 3 : *Existe-t-il une métaphrase précédente du métadiscours donnant la même information, ou permettant d'inférer la même information facilement ?*

Encore une fois on retrouve le critère de *redondance* de l'information mais ici en ne considérant que le métatexte, c'est-à-dire l'ensemble des métaphrases ; cette heuristique rejoint donc le premier présenté qui porte lui sur le discours du texte d'origine. Une métaphrase totalement redondante dans le métatexte devra ainsi être supprimée. Dans le cas d'une redondance partielle on préconisera de restituer la métaphrase sous une forme moins saillante.

► Heuristique 4 : *Importance pour la réalisation de la tâche.*

Un document textuel peut être utilisé à de nombreuses fins différentes : comprendre son contenu, rechercher une information précise, explorer le contenu afin de le résumer, etc. Ces différentes tâches vont utiliser des indices différents pour leur réalisation. Par exemple, une tâche de compréhension implique principalement une construction d'une représentation globale cohérente du contenu du document et de la situation décrite [Van Dijk and Kintsch 1983]. Ainsi, l'ensemble des indices de MFM relatifs à l'organisation globale du document auront une importance accrue dans le cadre d'une tâche de compréhension. Appliquer ce critère à d'autres tâches nécessite de les avoir modélisées en termes d'indices de mise en forme utiles à leur réalisation. À noter que cette heuristique peut être étendue des métaphrases aux objets textuels sur lesquels elles portent : si la MFM nous permet de dire qu'un objet est peu utile pour la réalisation de la tâche, alors on va non-



seulement filtrer les méta-informations de MFM mais également le contenu associé à l'objet dans le texte.

► Heuristique 5 : *Complexité de la métaphore étudiée.*

Dans le cadre de la restitution non-visuelle de la MFM, on peut évaluer la complexité d'une métaphore par le nombre d'objets textuels auxquels elle fait référence ainsi que par la distance entre ces objets sur la page. Cette heuristique ne permet pas à elle seule de décider de la restitution ou la non-restitution des métaphores, mais permet de donner plus de poids aux autres heuristiques : une métaphore complexe, non-redondante et importante pour l'interprétation du message doit être restituée.

► Heuristique 6 : *Fréquence d'occurrence de la métaphore dans le métatexte.*

Cette heuristique permet de décider à la fois ce qu'il faut restituer et comment, en écho avec les modalités à disposition. C'est-à-dire que si une métaphore a une fréquence d'occurrence très élevée, on pourra choisir en fonction de son importance pour la réalisation de la tâche de ne pas la restituer du tout ou alors de le faire de la manière la moins pénalisante possible pour la compréhension du contenu et la charge cognitive, et la moins frustrante possible pour l'utilisateur.

L'ensemble de ces heuristiques peut être utilisé manuellement au cas par cas pour chaque document, ou bien pour générer des règles de restitution relatives à une tâche ou un genre de document donné, afin de tendre vers une automatisation de la restitution. Étant donné que certaines heuristiques reposent sur les caractéristiques des utilisateurs, et de manière plus générale afin de prendre en compte la diversité des populations présentant une déficience commune, il est important que la restitution puisse à terme être paramétrable pour justement s'adapter aux utilisateurs. Ce point sera discuté plus amplement dans le Chapitre V évaluant différentes stratégies de restitution.

Afin d'opérationnaliser ces heuristiques, on pourrait générer un score global d'importance représentatif de l'application de chaque heuristique à une métaphore, afin de décider en fonction du score obtenu si elle doit être restituée, et si oui comment. Concernant ce dernier point, nous distinguons 3 méthodes de restitution correspondant à 3 niveaux d'importance décroissants : la première est de restituer explicitement la métaphore dans l'une des modalités présentées dans la section précédente ; la deuxième est de restituer la métaphore d'une manière succincte qui ne va « a priori » pas engendrer de surcharge cognitive ou d'allongement trop important du temps de lecture, par exemple en rajoutant un indice prosodique au texte ; la troisième enfin est de rendre l'information accessible à travers l'utilisation de commandes clavier, et donc de laisser le choix à l'utilisateur d'y accéder ou non.

Le filtrage des informations de MFM à restituer pourrait aussi être étendu aux objets textuels sur lesquels elles portent. Par exemple, filtrer tous les objets ayant le statut de « renvoi » dans le document, lorsque ces derniers ne sont pas pertinents dans le cadre de la tâche. Filtrer l'ensemble des informations « redondantes et non-pertinentes avec les objectifs de la tâche » est en effet un moyen de diminuer la charge cognitive [Giraud 2014], et permet une simplification et un allègement du contenu pour leur accès par des déficients visuels comme avec le projet DANTE qui filtre et restructure les pages web [Yesilada et al. 2007]. Le filtrage du contenu pourrait d'ailleurs gagner à se faire sur la base d'une modélisation plus fine du contenu, au-delà des informations de MFM. Par exemple dans le cas d'un document de cours expositif, il pourrait être pertinent de modéliser les

*intentions* ou *objectifs pédagogiques* placés derrière différents macro-segments du document, en utilisant par exemple la taxonomie de Bloom [Bloom and Krathwohl 1956; Krathwohl 2002] qui comporte notamment les activités « appliquer », « se souvenir », « évaluer » et « créer ». Ainsi, on pourrait filtrer et restructurer le contenu en fonction du niveau de connaissance et des objectifs d'un apprenant déficient visuel. Ce type de caractérisation fine du contenu devra être étudié plus en profondeur dans de futures études.

### **3.2. Comment restituer les informations de MFM : espace de conception**

---

Le Modèle d'Architecture Textuelle (MAT) [Virbel 1989; Pascual 1991] postule une équivalence informationnelle entre la forme « écologique » de la MFM telle qu'on la retrouve dans les documents écrits (marques typo-dispositionnelles et/ou lexico-syntaxiques) et une forme discursive entièrement développée (les métaphrases du modèle) ; ce postulat nous a permis de proposer une méthodologie de formalisation de la MFM (Chapitre III). Les métaphrases représentent donc d'après le MAT une forme *pivot*, à partir de laquelle un continuum d'autres formes plus ou moins visuelles vs. plus ou moins discursives peuvent être construites<sup>93</sup> (voir Figure 33 ci-dessous). Cette équivalence a été étendue par [Maurel 2004], qui postule l'équivalence entre les métaphrases de la MFM et un ensemble de formes orales ayant différentes marques prosodiques. L'exemple qu'il donne (p. 101) pour illustrer ce postulat est le suivant : les phrases « est-ce que Max est venu ↓ » (marques discursives et intonation affirmative descendante) et la phrase « Max est venu ↑ » (intonation interrogative montante) sont toutes deux interprétables comme des phrases interrogatives ayant la même sémantique. L'interprétation de la première se base sur des indices lexico-syntaxiques tandis que celle de la seconde utilise l'intonation du locuteur, d'où le postulat de leur équivalence.

Nous pensons que cette équivalence peut être étendue à d'autres modalités ; ainsi, on pourrait reformuler des marques de MFM sous la forme de patterns vibratoires, d'icônes audio ou d'autres dispositifs tactiles. Ce postulat repose sur le caractère *amodal* de l'acquisition d'information : peu importe le canal perceptif et les modalités de perception d'une information, si différentes formes d'une même information sont équivalentes sémantiquement parlant, alors l'interprétation sera la même, sous l'hypothèse que le récepteur connaisse les conventions utilisées dans les différentes formes de présentation. [Archambault 2010] justifie ce postulat notamment par l'exemple de lecteurs polyglottes, qui sont souvent incapables de dire dans quelle langue était écrit un document qu'ils ont lu antérieurement ; la langue pouvant être ici vue comme une modalité de transmission de l'information écrite.

---

<sup>93</sup> Cette construction se fait par *reformulation* du contenu discursif des métaphrases sous forme d'indices typographiques, dispositionnels et/ou lexico-syntaxiques. On peut faire l'analogie avec les technologies web, où plusieurs formats de présentation d'une page voire plusieurs versions différentes du contenu (par suppression de certains blocs) peuvent être produites à partir du même code source (format « pivot »), en utilisant des feuilles de styles différentes qui vont être appliqués, par exemple, en fonction du dispositif depuis lequel la page est accédée (smartphone, PC, tablette, etc.).

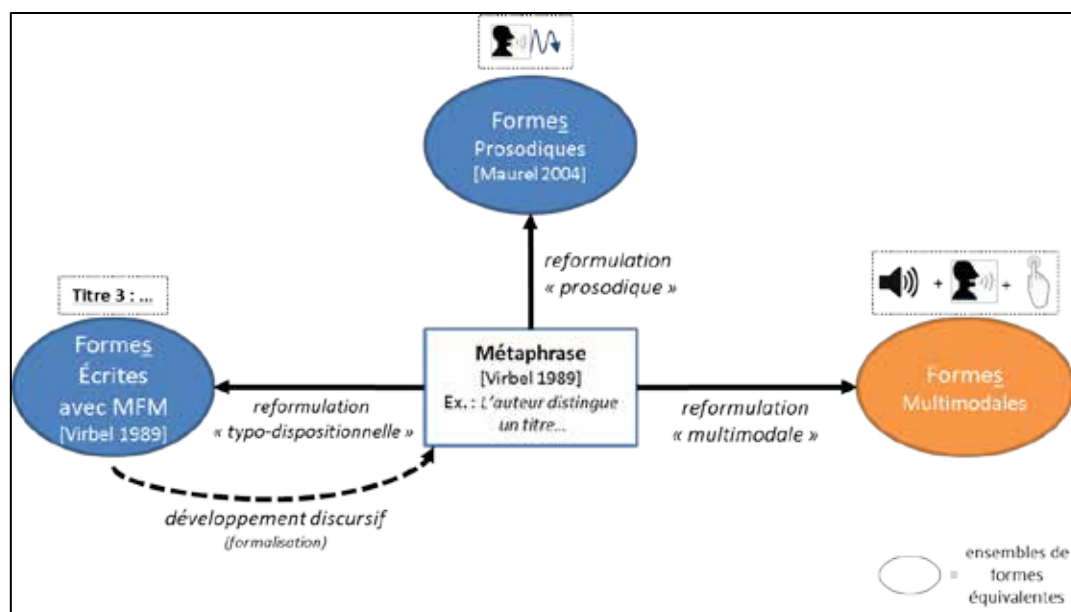


Figure 33 – Les différentes formes possible pour la reformulation des métaphrases

Cependant, si les différentes reformulations de métaphrases sont équivalentes sur le plan informatif, elles ne le sont pas nécessairement sur le plan fonctionnel. Par exemple, un texte lu par synthèse vocale n'aura pas les mêmes propriétés fonctionnelles pour le lecteur que dans sa forme écrite. Ainsi, dans le cadre de la restitution de la MFM, nous allons étudier dans cette partie les différentes propriétés des modalités utilisables pour un accès oral aux textes, ainsi que les cas d'utilisation dans lesquels ces modalités semblent pertinentes au regard des informations à restituer et du contexte de restitution.

Pour un texte lu par synthèse, les métaphrases (notre forme « pivot ») relatant des phénomènes de MFM peuvent être incorporées à la restitution à travers plusieurs modalités : verbalement en incorporant les informations de MFM au texte lu, à travers la modulation des voix de synthèse (changement de voix, prosodie, pauses), par l'insertion de sons non-verbaux (icônes audio), en spatialisant le son c'est-à-dire en simulant une source sonore dans l'espace, ou enfin par des modalités tactiles comme les pages en relief, les terminaux braille, ou encore des dispositifs vibratoires. À noter que même si l'on se restreint aux modalités orales dans les stratégies mises en place et évaluées dans nos travaux, nous présenterons tout de même succinctement les caractéristiques des autres modalités citées.

### 3.2.a. Reformulation discursive

La reformulation (ou *développement*) discursive permet initialement dans le MAT de formaliser les phénomènes de MFM sous forme de métaphrases, qui sont alors sous forme « lexico-syntaxique ». Mais la reformulation discursive permet également de produire un ensemble de *formes discursives* équivalentes à une ou plusieurs métaphrases sur le plan informatif.

[Maurel 2004] propose dans sa thèse 3 formes discursives différentes que peuvent prendre les métaphrases : une forme *descriptive* restituant les caractéristiques typo-dispositionnelles du phénomène de MFM de manière objective, et une forme dite *dénominateur* qui relate le statut et les propriétés d'un phénomène de MFM et enfin, une forme *interprétative*, qui décrit de manière plus

profonde les intentions de l'auteur. Pour reprendre l'exemple donné dans ces travaux, si l'on distingue un titre :

### RESTITUTION DE LA MFM

Qui pourrait être formalisé à l'aide des métaphrases :

*L'auteur distingue un titre <titre0>*

*L'auteur confère le niveau N à <titre0>*

Alors les 3 formes discursives descriptive, dénomminative, et interprétative de ce phénomène de MFM seraient :

- § *Forme descriptive* : « **Le texte en gros, centré, en gras et en capitales est** restitution de la MFM »
- § *Forme dénomminative* : « **Le titre de niveau N est** restitution de la MFM »
- § *Forme interprétative* : « **Je vais vous parler maintenant de la** restitution de la MFM »

Ces 3 formes sont discursives « sont en relation à travers une sémantique du « pourquoi » (relative au travail cognitif d'un lecteur) » [Maurel 2004](p.103). Ainsi ces 3 formes dénotent de 3 niveaux d'interprétation des marques de MFM du titre utilisé dans l'exemple, qui correspondent à 3 stratégies différentes de reformulation discursive, pour reprendre les termes de l'auteur. On notera que ces 3 formes ne sont pas totalement équivalentes sur le plan informatif : les formes descriptive et interprétative n'explicitent pas objectivement que l'objet est un titre avec un certain niveau hiérarchique (qui dénote de sa portée), et la forme descriptive est la seule à donner les caractéristiques physiques du titre. Ainsi, le choix de la stratégie de reformulation, qui peut être une des 3 présentées ou une autre spécifique, doit être réalisé dans la perspective de la tâche et de son contexte : la forme descriptive permet de faire le lien avec les propriétés typo-dispositionnelles (si par exemple l'on recherche un objet en fonction de ses caractéristiques typo-dispositionnelles) ; la forme dénomminative est la plus explicite mais nécessite que le récepteur sache interpréter le niveau hiérarchique d'un titre en termes d'organisation du texte ; enfin, la forme interprétative est la plus « naturelle » dans le sens où elle est totalement intégrée au discours du texte, ce qui peut faciliter la lecture, mais ne nous informe pas directement du statut de titre de l'objet, ni de son niveau hiérarchique.

Concernant les différents types d'informations de MFM à restituer, l'avantage de la forme discursive du métatexte est sa flexibilité : elle permet de décrire aussi bien le statut des objets textuels (ex. titre) que leurs propriétés (ex. niveau hiérarchique), mais aussi les informations transverses en faisant référence aux objets à travers leurs statuts et propriétés, leurs contenus, leur position dans le texte ou encore leurs propriétés graphiques (ex. cette section est intitulée « Reformulation... »). La reformulation discursive permet également de décrire les démarcations entre objets (ex. « Dans ce nouveau paragraphe... »).

Cependant, le principal inconvénient de la reformulation discursive, si cette stratégie de restitution est généralisée, est que texte et métatexte sont communiqués par une même modalité verbale. Cette indissociation a plusieurs conséquences. Tout d'abord, le temps de parcours du document est allongé. En effet, le métatexte ajouté au contenu initial devra être lu par synthèse vocale (ou en braille), modalités qui sont limitées par leur vitesse de lecture. Or, nous avons vu que le facteur

temps dans l'accessibilité des contenus numériques était un critère important, étant donné que les utilisateurs déficients souffraient d'un important allongement du temps d'interaction pour réaliser la même tâche sur les mêmes contenus que les utilisateurs non-déficients. La contrainte temporelle supplémentaire amenée par la reformulation discursive pourrait dès lors engendrer des frustrations. De plus, l'indissociation du texte et du métatexte peut augmenter la charge cognitive dans la boucle phonologique en mémoire de travail, car ces informations devront être lues, évaluées, sélectionnées et traitées, d'autant que les modes non-visuels n'autorisent pas le filtrage de l'information, ce qui est source de charge cognitive supplémentaire [Giraud 2014]. Enfin, une difficulté est la présence potentielle d'ambiguïtés quant à la portée des informations métatextuelles. L'auditeur peut ne plus savoir si une information est de nature textuelle ou métatextuelle. Pour lever les possibles ambiguïtés, il faudrait donc envisager que le métatexte indique le début d'un objet textuel mais aussi sa fin, à la façon d'une balise html. Toutefois, cette nécessité ne ferait qu'aggraver les conséquences négatives que nous venons de citer.

La reformulation discursive reste une modalité de restitution intéressante de par sa flexibilité et son caractère explicite, mais dont il faut user avec parcimonie ; ses avantages et inconvénients sont résumés dans le Tableau 14 ci-après).

Avantages	Inconvénients
<ul style="list-style-type: none"> <li>- flexible : permet de restituer tout type d'information de MFM.</li> <li>- explicite : utilisation du langage naturel.</li> <li>- plusieurs formes possibles adaptables au contexte.</li> </ul>	<ul style="list-style-type: none"> <li>- rallonge le temps de parcours du document, facteur de frustrations.</li> <li>- peut augmenter la charge cognitive.</li> <li>- peut générer des ambiguïtés.</li> </ul> <p><i>Ces inconvénients n'apparaissent pas systématiquement ; ils dépendent de la stratégie de reformulation employée, de la quantité de discours ajoutée au contenu d'origine, et de la pertinence de ce dernier relativement au contenu qu'il annote.</i></p>

**Tableau 14 – Avantages et inconvénients de la reformulation discursive**

### **3.2.b. Modulation de la ou des voix de synthèse**

Les premières composantes non-verbales modifiables d'une lecture par synthèse vocale sont les composantes de la voix. Il est ainsi possible de modifier la vitesse de la voix de synthèse, sa hauteur (*pitch*), de rajouter des pauses dans l'énonciation, d'utiliser des schémas prosodiques (intonations) particuliers, ou encore différentes voix de synthèse. Ces modifications permettent d'associer ces différents « états » de la ou des voix de synthèse à une information particulière (ex. utiliser un schéma prosodique particulier pour énoncer les titres). Le principal avantage de cette approche est de pouvoir véhiculer une information supplémentaire en *parallèle* du discours, et ainsi ne pas rallonger le temps de parcours du temps. De plus, le fait que cette information supplémentaire soit restituée dans une modalité non-verbale fait que son traitement peut être optimisé cognitivement parlant (d'après le modèle en vigueur de la mémoire de travail, cf. Chapitre II.3.2). Cependant, ce dernier point est à nuancer : un regain de performances cognitives ne peut avoir lieu que si la reconnaissance du schéma de voix utilisé a lieu, et si son interprétation (la récupération de l'information qu'il véhicule) a été apprise et intégrée. Dans le cas contraire, la non-reconnaissance du schéma peut perturber la compréhension des informations verbales qui a lieu en même temps, et la difficulté de l'interprétation (si l'apprentissage en est trop superficiel) peut nécessiter des

récupérations en mémoire à long terme des règles apprises, ce qui aura pour conséquence d'augmenter la charge cognitive.

Nous allons maintenant passer en revue chacune des possibilités de modulation des voix de synthèse ; à noter que l'ensemble des modulations du discours oral correspond dans le langage courant à ce que l'on appelle la *prosodie*. La prosodie (i.e. l'ensemble des modulations de la voix) représente à priori une solution « naturelle » pour la restitution de la mise en forme ; il est en effet pointé dans la littérature que la prosodie joue à l'oral un rôle de facilitateur de traitement de l'information en lui donnant une structure à travers les modulations de la voix, de manière similaire à la MFM à l'écrit qui structure visuellement le contenu [Coirier et al. 1996; Nespoulous et al. 2002]. Afin de mieux distinguer les différentes modulations des voix utilisées dans des interfaces non-visuelles, nous utiliserons ici le terme *prosodie* pour désigner uniquement les *schémas intonatifs* utilisés à l'oral. Nous détaillerons en suivant d'autres modulations particulières comme la hauteur de la voix ou l'utilisation de pauses dans le discours oral.

Les schémas prosodique sont des variations intonatives de la voix sur l'énonciation d'un ou plusieurs mots ; par exemple, une intonation ascendante, souvent utilisée dans des phrases interrogatives (ex. « Où est Charlie ↑ ? »). Cependant, la définition de schémas prosodiques particuliers qu'on associerait à des informations de MFM se heurte à l'utilisation de la prosodie dans le langage naturel. En effet, l'interprétation de ces nouveaux schémas pourrait rentrer en concurrence avec leur interprétation codifiée par le langage naturel ; par exemple, une intonation montante pourrait être confondue avec une question dans le discours oral. D'ailleurs, [Maurel 2004] exploite des indices du discours pour utiliser différents schémas prosodiques (rhème, thèmes, et topiques). De plus, la lecture orale d'un texte mis en forme n'associe apparemment pas (en tous cas de manière systématique) des schémas prosodiques particuliers aux phénomènes de MFM [Maurel 2004]. Ce dernier auteur souligne notamment la grande variabilité des oralisations des titres et énumérations dans des corpus oraux. Enfin, nous questionnons la viabilité de l'utilisation de la prosodie lorsque les débits de lecture des synthèses vocale sont fortement augmentés relativement au débit moyen du langage oral, comme c'est le cas pour certains utilisateurs déficients visuels [Asakawa and Takagi 2003] . Ainsi l'utilisation de la prosodie (et plus précisément de schémas intonatifs) pour véhiculer des informations de MFM semble délicate, en marge de la ponctuation des phrases. Ils ont cependant été utilisés avec succès par [Lemarié 2006b] pour marquer les fins d'item dans une lecture par synthèse vocale de menus de restaurants, améliorant la compréhension et la mémorisation par rapport à une forme orale dépourvue de prosodie ; à noter que dans ce cas précis les schémas prosodiques étaient suivis de pauses dans l'énonciation, ce qui fait qu'on ne peut attribuer les résultats à l'une ou l'autre modulation.

En marge des phénomènes de MFM, l'étude de la lecture orale de textes par [Ouden et al. 2002] a montré que la hauteur de la voix et les pauses effectuées par les locuteurs variaient en fonction du niveau hiérarchique du contenu lu, ainsi que d'autres paramètres discursifs comme la nucléarité des segments au regard de la théorie des structures rhétoriques. Ainsi, on pourrait faire varier la hauteur de la voix (fréquence fondamentale du spectre de la voix) en l'associant à différentes informations de MFM, comme par exemple le niveau hiérarchique des segments lus (information indirectement donnée par la MFM). On sait qu'en moyenne un humain est capable de discriminer et reconnaître 6 classes de hauteurs différentes de sons [Miller 1994], ce qui peut s'appliquer dans une certaine mesure aux voix en considérant que leurs spectres fréquentiels sont probablement plus étendus et

moins symétriques que ceux des sons utilisés dans le cadre expérimental rapporté précédemment. Cependant, dans le cadre de la lecture de texte, les modulations extrêmes de la hauteur de la voix utilisée peuvent altérer la compréhension du discours ; on subodore que cela peut également affecter l'aisance d'écoute, la satisfaction des auditeurs. De manière générale il est important de ne pas pénaliser la compréhension en restituant la MFM, restitution qui vise justement dans notre cas l'objectif inverse.

Rajouter des pauses dans l'énonciation du texte peut également être un vecteur d'information, dont [Lorch et al. 2012] ont montré qu'il était pertinent pour marquer la présence de titres dans un texte oralisé par synthèse vocale (sans autre modulation de la voix). Ces auteurs ont montré que l'insertion de pauses après les titres augmentait significativement l'identification des thèmes correspondant par rapport à une version orale du texte sans pause, et ce chez des auditeurs non déficients. Cependant, il semble difficile de discriminer précisément différentes durées de pauses de l'ordre de ce qui semble acceptable dans une oralisation de texte<sup>94</sup> ; des pauses trop longues entre la lecture des objets pénaliseraient vraisemblablement la durée totale de lecture et l'aisance d'écoute. De plus les pauses ne peuvent intervenir qu'avant ou après la lecture d'un objet textuel. Elles semblent toutefois intéressantes pour appuyer la démarcation entre objets, mais ne peuvent que difficilement être associées à des informations plus complexes étant donné la difficulté d'une discrimination fine de leurs durées. L'augmentation du débit des synthèses peut également être problématique pour la discrimination de différentes durées de pause : même si elles resteront proportionnelles avec les changements de débit, leurs durées d'autant plus réduites pourraient les rendre encore plus difficilement discriminables.

Modifier la vitesse d'énonciation de la voix de synthèse représente une possibilité de modulation ; plusieurs arguments nous poussent cependant à l'écarter de la restitution. Tout d'abord, définir plusieurs vitesses différentes de lecture pour un même texte (en fonction de l'information à restituer) risque de pénaliser la compréhension pour des vitesses élevées ; cela peut aussi engendrer des frustrations si celle-ci est trop lente, ou bien simplement à causes de changements trop fréquents (ce qui enlève en partie le peu de contrôle du flux audio que les utilisateurs ont). Cela rend également la définition d'une vitesse globale de lecture plus difficile (les utilisateurs souhaitant accélérer la lecture seraient limités par la nécessité de comprendre les vitesses accélérées). Enfin, on peut supposer que la discrimination et la reconnaissance fines de plusieurs vitesses d'énonciation est une tâche ardue, car peu répandue (pour ne pas dire inexistante) dans les situations de discours oral. Nous préférons donc l'utilisation d'autres indices pour la restitution. À noter que la modification du volume de la voix souffre des mêmes inconvénients que les changements de vitesse.

Enfin, une dernière possibilité de modulation est l'utilisation de plusieurs voix de synthèse, en associant chaque voix à une information ou un type d'information particulier. Cette solution semble plus naturelle que la modification de la hauteur des voix, car basée sur notre capacité naturelle à discriminer les voix de nos semblables. En effet, [Kreiman and Papcun 1991] ont montré que reconnaître une voix masculine parmi 10 se faisait avec plus de 80% de réussite, lorsque ces voix ont été entendues pour la première fois une semaine avant les tests. Ainsi, en choisissant des voix

---

<sup>94</sup> [Fraise 1984] suggère que l'estimation du temps chez l'homme se distingue en 3 composantes fonctionnelles : l'instantané (moins de 100ms), le présent (entre 100ms et 5secondes) et le présent étendu (au-delà de 5 secondes) qui implique la mémoire à long-terme. Ces résultats appuient les limites des capacités humaines d'estimation des (courtes) durées.

masculines et féminines aux timbres assez différents, on peut sans doute, après apprentissage, associer aisément une dizaine de voix différentes à différentes informations (peut-être plus en utilisant des voix « familières »). Enfin, l'utilisation de voix peut se faire indifféremment à plusieurs débits de lecture, ce qui est un argument pour son implémentation pour des utilisateurs déficients visuels.

Concernant les types d'informations de MFM à restituer, les modulations de la voix de synthèse permettent de véhiculer des informations locales de statut et de propriété (ex. les titres sont lus par une voix particulière parmi plusieurs différentes, et le niveau du titre module la hauteur de la voix). La démarcation entre objets peut être restituée à travers les pauses dans l'énonciation, mais avec très peu de modulations (durées de pauses) différentes. Les informations transverses ne peuvent pas être véhiculées par ce type d'indice étant donné qu'il agit en parallèle du discours oral.

Type de modulation de la voix	Avantages	Inconvénients
Schémas prosodiques	<ul style="list-style-type: none"> <li>- modalité « naturelle » retrouvée dans le discours oral.</li> <li>- possibilité de définir de nombreux schémas différents.</li> <li>- modalité parallèle au discours oral.</li> </ul>	<ul style="list-style-type: none"> <li>- définition limitée par les schémas existants conventionnellement utilisés dans le discours oral (ex. intonation interrogative).</li> <li>- modifiés par l'augmentation du débit d'énonciation de la voix.</li> <li>- nécessite un apprentissage.</li> </ul>
Hauteur de la voix	<ul style="list-style-type: none"> <li>- propriétés inchangées par une modification du débit de la voix.</li> <li>- modalité parallèle au discours oral.</li> </ul>	<ul style="list-style-type: none"> <li>- limité à la définition de 6 classes maximum.</li> <li>- peut perturber la compréhension.</li> <li>- peut pénaliser l'aisance d'écoute</li> <li>- nécessite un apprentissage.</li> </ul>
Pauses dans l'énonciation	<ul style="list-style-type: none"> <li>- permet de signaler "naturellement" les démarcations entre objets.</li> </ul>	<ul style="list-style-type: none"> <li>- mauvaise discrimination : définition de différentes classes très limitée.</li> </ul>
Changement de vitesse ou de volume de la voix	<ul style="list-style-type: none"> <li>- modalité parallèle au discours oral.</li> </ul>	<ul style="list-style-type: none"> <li>- mauvaise discrimination : définition de différentes classes très limitée.</li> <li>- peut perturber la compréhension.</li> <li>- peut pénaliser l'aisance d'écoute.</li> <li>- peut engendrer des frustrations en réduisant le contrôle du flux audio par l'auditeur.</li> <li>- nécessite un apprentissage.</li> </ul>
Utilisation de plusieurs voix différentes	<ul style="list-style-type: none"> <li>- discrimination : possibilité d'utiliser jusqu'à une dizaine de voix différentes.</li> <li>- propriétés inchangées par une modification du débit de la voix.</li> <li>- modalité parallèle au discours oral.</li> </ul>	<ul style="list-style-type: none"> <li>- nécessite un apprentissage.</li> </ul>

Tableau 15 – Avantages et inconvénients des différentes modulations des voix de synthèse pour la restitution

### 3.2.c. Modulation du son : icônes audio et son spatialisé

Les *icônes audio* (*earcons*) représentent un moyen simple de véhiculer de l'information à travers des sons courts. L'exemple le plus répandu est celui des sons Windows associés par exemple à un message d'erreur. Nous avons vu précédemment que l'on distinguait au moins 2 paradigmes de conception pour ces sons : le premier étant d'utiliser des métaphores sonores pour que la signification des sons n'ait pas besoin d'être apprise (ex. son de froissement de papier pour la



suppression d'un fichier), tandis que le second paradigme consiste à associer de manière abstraite des sons à des informations, en clair à définir un vocabulaire sonore. Le premier paradigme nous semble difficilement applicable dans le cadre de la restitution des informations de MFM ; en tous cas les métaphores sonores mises en place pourraient n'être pas assez *affordantes* et nécessiter tout de même un apprentissage au vu de la spécificité et de la complexité des informations de MFM à restituer. [Blattner et al. 1989] précise que ce type d'icônes audio présentent les mêmes inconvénients que les représentations picturales. Nous nous concentrerons donc sur l'utilisation d'icônes audio « abstraites ».

Les icônes audio ont l'avantage de présenter succinctement une information à l'oral, et celles évaluées dans la littérature durent généralement entre 1 et 1.5 secondes [Brewster et al. 1993], mais elles peuvent être beaucoup plus courtes (de l'ordre de 20ms) [Bates and Fitzpatrick 2010] si leur fréquence d'occurrences est élevée. L'augmentation de la vitesse de lecture (débit de la synthèse vocale) ne devrait toutefois pas diminuer la durée des icônes audio en dessous de ce seuil de 20ms ; cela reste toutefois à vérifier empiriquement. Concernant la reconnaissance des icônes audio, [Brewster et al. 1993] ont montré qu'après 3 répétitions de chaque icône, les sujets étaient capables de les reconnaître avec une taux de réussite de plus de 80%. Les icônes audio utilisées dans ces expérimentations n'étaient pas toutes entièrement différentes : le timbre des icônes définissait des catégories ou familles d'items, et le rythme des sons utilisés pour chaque icône définissait des sous-catégories. Ainsi, en jouant sur la similarité des différentes composantes des icônes audio, on peut aisément en définir différents types et ainsi faciliter leur reconnaissance. Ce dernier auteur suggère d'utiliser en priorité le timbre et le rythme pour différencier les icônes, et d'utiliser la hauteur, le registre ou l'intensité pour affiner la classification. Enfin, à noter que la musicalité d'icônes audio abstraites peut, à elle seule, véhiculer de l'information sans apprentissage (ex. accord mineur pour une information connotée négativement). Un autre de leurs avantages tient à leur caractère non-verbal, permettant un traitement cognitif séparé du contenu et donc optimisé dans une certaine mesure ; [Klatzky et al. 2006] a d'ailleurs montré que les sons non-verbaux étaient plus efficaces que le verbal pour véhiculer une information dans une situation où la charge cognitive est importante. Cependant, l'acceptabilité de l'utilisation d'icônes audio pendant la lecture d'un document reste inconnue ; une heuristique déjà appliquée est de ne les utiliser que pour les informations avec les fréquences d'occurrence les plus élevées [Hussain et al. 2012]. Enfin, en ce qui concerne la temporalité d'utilisation des icônes audio, [Brewster et al. 1995] ont montré que le taux de reconnaissance de deux icônes audio présentées en simultané était aussi bon que lorsque les deux mêmes icônes audio étaient présentées de manière séquentielle, sous réserve que les icônes respectent certains critères de conception. La superposition de 2 icônes audio permet de ne pas rallonger l'écoute pour l'auditeur, ce qui à l'échelle du temps d'écoute d'un document de plusieurs pages a toutefois un impact faible. En effet, dans les travaux de Brewster les icônes audio utilisées duraient en moyenne 1 seconde ; dans le cas où les icônes sont utilisées en tant que feedback pendant l'interaction, le gain de temps peut être substantiel pour conserver sa fluidité, mais pas dans le cas d'une lecture orale pour signaler des objets textuels, où le gain de temps sur une lecture de plusieurs minutes serait négligeable<sup>95</sup>. Ce résultat nous amène cependant à la question de l'utilisation d'icônes audio en parallèle du discours oral ; aucune étude à ce jour n'évalue l'acceptabilité ou les performances d'une telle configuration, qui pourrait permettre d'utiliser les

---

<sup>95</sup> Par exemple, cette page fait plus de 500 mots, ce qui représente environ 3 minutes de lecture à 170/180 mots par minute (débit par défaut appliqué aux synthèses vocales par les lecteurs d'écran).

icônes audio pour signaler des méta-informations à un niveau intra-paragraphique. Ainsi concernant l'utilisation des icônes audio, nous nous restreindrons à les positionner avant ou après un objet textuel pendant la restitution.

Les icônes audio abstraites peuvent ainsi représenter des informations locales de MFM comme des statuts ou propriétés, ou encore signaler les démarcations entre objets en utilisant une icône dédiée. Elles pourraient également être utilisées pour représenter certaines informations transverses (ex. nombre de sections dans le chapitre courant), dans la mesure où la complexité de l'information à représenter reste limitée, où le nombre d'icônes audio définies est restreint et où les icônes sont aisément discriminables.

Une seconde modalité audio utilisable pour restituer la MFM est la spatialisation du son, qui consiste à simuler la localisation spatiale d'un flux audio, cela étant applicable aux voix de synthèse ou aux icônes audio. Les capacités humaines de localisation du son sont assez limitées en termes de résolution, et leur discrimination est la plus efficace sur le plan horizontal en face de l'auditeur [Goose and Möller 1999]. De plus, la spatialisation du son requiert l'utilisation d'une fonction de transfert « moyenne », qui peut donner des résultats différents d'un individu à l'autre en termes de perception. Cependant, il est possible de définir différentes sources sonores pour typer l'information audio, en associant chaque source à un type d'information particulier (ex. lire le contenu des titres depuis une source sonore particulière). On pourrait également utiliser les déplacements entre ces sources sonores pour représenter des changements de type ou des déplacements au sein du document. L'avantage de la spatialisation est qu'elle a lieu en parallèle de la lecture orale et dans une modalité non-verbale. On peut par exemple l'utiliser pour restituer les démarcations entre objets par changement de source sonore, ou encore associer certaines sources sonores à des statuts ou propriétés particulières. Une limite majeure est qu'il est absolument nécessaire que les auditeurs portent un casque audio binaural afin de correctement percevoir la spatialisation du son, celle-ci reposant sur des modulations calculées relativement à la position de la tête de l'auditeur dans l'espace.

Modulation du son	Avantages	Inconvénients
icônes audio abstraites	<ul style="list-style-type: none"> <li>- discrimination aisée sous réserve de leur bonne conception.</li> <li>- possibilité de définir des familles d'icônes à travers des propriétés communes, pour faciliter leur reconnaissance.</li> <li>- pertinentes en situation de charge cognitive</li> </ul>	<ul style="list-style-type: none"> <li>- nécessitent un apprentissage.</li> <li>- peuvent pénaliser l'aisance de lecture et engendrer des frustrations si utilisées en trop grand nombre.</li> </ul>
Spatialisation du son	<ul style="list-style-type: none"> <li>- modalité parallèle au discours oral.</li> </ul>	<ul style="list-style-type: none"> <li>- nécessite un apprentissage.</li> <li>- différences interindividuelles pour la reconnaissance.</li> <li>- discrimination limitée, sauf sur le plan horizontal.</li> <li>- nécessité de porter un casque audio binaural.</li> </ul>

Tableau 16 - Avantages et inconvénients des icônes audio et du son spatialisé pour la restitution

### 3.2.d. Modalités tactiles

Une autre possibilité pour restituer les informations de MFM, qui sera brièvement évoquée dans cette section mais que nous n'avons pas reprise dans nos travaux, est d'utiliser le mode tactile à

travers différentes modalités comme une plage braille, des matrices de picots, des feuilles de papier en relief ou encore des dispositifs vibratoires. Nous ne présenterons pas ici les dispositifs utilisant le mode kinesthésique et notamment les dispositifs à retour d'effort évoqués dans l'état de l'art, encore assez peu répandus, qui relèvent d'un domaine d'étude éloigné du champ de ces travaux et n'ont à notre connaissance jamais été utilisés pour la lecture de documents.

Une plage braille pourrait être utilisée pour afficher un certain type d'informations, ou de manière plus générale dédier l'oral à certains types d'information et le braille à d'autres. L'avantage du braille est sa persistance, on pourrait par exemple y afficher le titre de la partie en cours de lecture ; les inconvénients de ce type de dispositifs ont déjà été évoqués et sont repris dans le Tableau 17 ci-après.

Une matrice de picots de grande taille pourrait permettre d'afficher la disposition des objets de MFM sur la page, ou encore d'afficher des schémas de picots qui pourraient représenter iconiquement certaines informations, ce qui est d'ailleurs réalisable sur des matrices de petite taille [Pietrzak 2008; Gapenne et al. 2003]. La même chose est réalisable à moindre coût à l'aide de papier thermogonflable pouvant être mis en relief, méthode déjà employée pour restituer la mise en forme de pages web [Bou-Issa 2010].

Enfin, les dispositifs vibratoire (ex. bracelet ou souris vibrante, *smartphone*) peuvent être vecteurs d'information. Les moteurs vibrants peuvent varier en fréquence, en amplitude et dans le temps en créant des rythmes, afin de générer différents schémas vibratoires. Ces dispositifs semblent particulièrement pertinents pour indiquer des transitions entre différentes phases d'une tâche [Kontarinis et al. 1995] ; ces auteurs évoquent le fait qu'ils sont souvent utilisés dans des tâches exploratoires. Ainsi, un dispositif vibratoire permettrait de restituer les démarcations entre objets de MFM dans une modalité non-verbale, mais pourrait être aussi utilisé pour définir des schémas vibratoires pouvant représenter différents types d'informations comme les *Tactons* de [Brewster and Brown 2004]. Ces dispositifs présentent aussi l'avantage de pouvoir être utilisés parallèlement au discours oral.

Dispositif tactile	Avantages	Inconvénients
Plage braille (40 à 80 caractères)	<ul style="list-style-type: none"> <li>- tangibilité et donc persistance.</li> <li>- possibilité d'afficher du braille.</li> </ul>	<ul style="list-style-type: none"> <li>- limité à une seule ligne de texte.</li> <li>- nécessite l'apprentissage du braille.</li> <li>- vitesse de lecture inférieure à l'énonciation orale par une voix de synthèse.</li> <li>- coût élevé.</li> </ul>
Matrice braille	<ul style="list-style-type: none"> <li>- tangibilité et donc persistance.</li> <li>- possibilité de représenter des icônes tactiles.</li> <li>- possibilité de représenter des formes.</li> <li>- possibilité d'afficher du braille.</li> </ul>	<ul style="list-style-type: none"> <li>- coût très élevé pour les grandes matrices.</li> <li>- implique des allers retours avec le poste de travail.</li> <li>- nécessite un apprentissage des icônes tactiles.</li> <li>- lenteur de l'exploration tactile.</li> </ul>
Papier thermogonflable	<ul style="list-style-type: none"> <li>- tangibilité et donc persistance.</li> <li>- possibilité de représenter des icônes tactiles.</li> <li>- possibilité de représenter des formes.</li> <li>- possibilité d'afficher du braille.</li> </ul>	<ul style="list-style-type: none"> <li>- production manuelle.</li> <li>- implique des allers retours avec le poste de travail.</li> <li>- nécessite un apprentissage des icônes tactiles.</li> <li>- lenteur de l'exploration tactile.</li> </ul>

Dispositif vibratoire	- parallèle au discours oral. - possibilité de représenter des icônes vibratoires.	- nécessite un apprentissage des icônes vibratoires.
-----------------------	---	--

**Tableau 17 - Avantages et inconvénients de modalités tactiles pour la restitution**

### ***3.2.e. Restitution à travers la définition de commandes utilisateur***

Enfin, une dernière possibilité pour restituer les informations de MFM, qui ne représente pas une modalité de restitution en soi mais un moyen d'accéder à l'information, est de restituer l'information métatextuelle non pas pendant la lecture « passivement » mais activement à travers des commandes utilisateurs. Ces commandes pourraient être activées par des combinaisons de touches du clavier, par des commandes vocales ou des gestes particuliers. L'idée est de laisser la possibilité à l'utilisateur d'accéder ou non à une information donnée pendant la lecture, en associant l'accès à chaque information à une commande particulière (ex. CTRL+T pour connaître le titre de la section en cours de lecture). Les informations correspondantes sont ensuite restituées dans l'une des modalités présentées précédemment dans cette section.

Pendant la lecture d'un objet textuel, le lancement de différentes commandes permettrait d'accéder à son statut ou ses différentes propriétés ; le cas échéant les commandes signaleraient l'absence d'information (ex. cette section n'a pas de titre). Il en va de même pour les relations ciblant ou partant de l'objet en cours de lecture. D'autres commandes permettraient d'accéder aux informations transverses de MFM. Cette approche permet d'éviter de surcharger l'utilisateur d'informations, en lui laissant le soin de choisir lesquelles sont pertinentes pour la tâche qu'il réalise. Les lecteurs d'écran adoptent d'ailleurs une restitution presque exclusivement basée sur le discours et son accès/navigation via des commandes.

Avantages	Inconvénients
- ne rajoute aucune information à la restitution et ainsi évite une surcharge cognitive directe.	- nécessite l'apprentissage des raccourcis clavier (ou des autres types d'entrée le cas échéant). - source d'erreurs si les commandes ne sont pas lancées au bon moment pendant la lecture (qui est passée à l'objet suivant). - mêmes les utilisateurs experts n'utilisent pas l'intégralité des raccourcis clavier à leur disposition [Lane et al. 2005]

**Tableau 18 - Avantages et inconvénients des commandes utilisateur pour la restitution**

### ***3.2.f. Comment combiner les modalités de restitution***

Dans l'optique d'une restitution la plus optimale possible, l'idée est de combiner les différentes modalités présentées afin d'exploiter leurs avantages et de minimiser leurs inconvénients. Par exemple si on ne restitue les méta-informations qu'avec des icônes audio, l'inconvénient cité (voir 3.2.c) « pénalisation de l'aisance d'écoute » sera beaucoup plus prégnant que si l'on se contentait d'en utiliser quelques une pour la restitution. Les inconvénients cités pour chaque modalité sont donc à mettre en perspective avec leur utilisation en contexte et notamment leur fréquence d'occurrence. Ou pour reprendre un paradigme déjà évoqué, différentes formes d'un même phénomène de MFM (ex. discursif vs. visuel) à l'écrit peuvent être équivalentes sur le plan informationnel mais pas sur le plan fonctionnel.

De manière générale, le choix des modalités de restitution doit suivre un compromis entre l'utilisation du verbal et l'utilisation du non-verbal, c'est-à-dire un compromis entre apprentissage nécessaire (du vocabulaire non-verbal) et bénéfices retirés de l'utilisation de modalités non-verbales. Ne restituer les méta-informations que verbalement ne nécessiterait aucun apprentissage particulier mais n'exploiterait aucun bénéfice temporel ou cognitif des autres modalités présentées. À contrario, n'utiliser que des modalités non-verbal pourrait être optimal (si leur utilisation est pertinente au regard de leurs caractéristiques), mais nécessiterait un apprentissage important de la part des utilisateurs. Cette idée de compromis est illustrée par les travaux de [Hussain et al. 2014] : ces auteurs ont montré que la méthode la plus efficace et satisfaisante de feedback audio était celle mélangeant feedback verbal et feedback non-verbal (comparativement à un feedback uniquement verbal et uniquement non-verbal), dans le cas d'une application d'aide à la navigation piétonne pour les non-voyants. L'heuristique qu'ils appliquaient pour l'utilisation du feedback non-verbal (ici des *spearcons*<sup>96</sup>) était la fréquence d'occurrence des items du feedback de l'application : les items les plus fréquents étaient remplacés par une *spearcon* correspondante. Nous avons également utilisé ce critère de la fréquence d'occurrence des informations à restituer pour concevoir une stratégie optimale, présentée dans la partie 4, où l'apprentissage devait être moindre étant donné le cadre expérimental visé.

Enfin, en ce qui concerne l'utilisation de modalités non-verbales, on peut supposer que dans un cadre d'utilisation régulière ou au moins répétée de dispositifs d'accès non-visuel aux documents, on pourrait introduire un vocabulaire non-verbal de plus en plus complexe avec le développement de l'apprentissage (d'une certaine forme de « littéracie ») afin de tendre vers une restitution optimale. Il semble d'ailleurs intéressant pour de futures études d'évaluer l'évolution des performances et de l'acceptabilité de tels dispositifs qui seraient utilisés fréquemment, et de voir dans quelle mesure la complexification de la grammaire non-verbale peut être bénéfique.

### **3.3. Quand restituer les informations de MFM ?**

---

Les métaphrases peuvent porter sur un seul objet textuel (informations locales), mettre en relation plusieurs objets ou bien porter sur un ensemble d'objets. Nous avons vu quelles étaient les modalités de restitution possibles de ces informations, dans le cadre d'une lecture par synthèse vocale des textes, et plusieurs critères pour filtrer et reformuler les informations de MFM (que restituer ?). La question que nous allons développer est, pour chacun des cas évoqués au début de ce paragraphe, quand restituer les informations portant sur les objets textuels ?

Cette question peut être reformulée de la manière suivante : dans le cadre d'un accès non-visuel et donc séquentiel, doit-on restituer des méta-informations relatives aux objets textuels avant, après ou pendant la lecture orale de ces derniers ? Choix qui nous est en effet imposé par la nature séquentielle des modes non-visuels. À noter que toutes les modalités présentées dans la section 3.1 ne sont pas utilisables aux 3 moments de la lecture évoqués ; par exemple un schéma prosodique est lié au discours oral et donc ne peut être utilisé que *pendant* la lecture d'un objet.

Donner une méta-information *avant* la lecture d'un objet correspondant permet de pré-activer les fonctions cognitives de traitement et de représentation adéquates. Ce postulat est appuyé par la

---

<sup>96</sup> *Spearcons* signifie *Speech (auditory) Icons*, et sont créées par accélération extrême du débit de la synthèse vocale sur la lecture d'un ou plusieurs mots. Voir Chapitre IV pour plus de détails.

littérature sur les vues d'ensembles (*previews* ou *overviews* en Anglais) dans les textes, qui sont des dispositifs visant à annoncer au lecteur la structure organisationnelle du texte au début de la lecture, sous forme verbale (par ex. un plan du texte) ou graphique (une carte conceptuelle<sup>97</sup>). En effet plusieurs études [Lorch 1989; Spyridakis 1989] ont montré que les vues d'ensembles aident le lecteur à sélectionner, organiser et récupérer l'information en mémoire; à noter que ces effets ne se produisent que lorsque les vues d'ensemble sont présentées en début de texte, alors qu'une présentation après la lecture n'a que peu, voire pas d'effet. Ces résultats appuient l'importance de pouvoir anticiper la structure du texte ; cela nous semble particulièrement important lors de son accès non-visuel : l'absence de ces dispositifs en lecture visuelle laisse toujours la possibilité au lecteur de construire des représentations équivalentes en balayant le texte, ce qui n'est pas possible lors d'un accès par synthèse vocale, par exemple.

Ainsi donner une méta-information avant la lecture du ou des objets correspondants sera particulièrement pertinent lorsque cette méta-information dénote de la structure ou de l'organisation des objets subséquents, à la manière des vues d'ensembles pour la macrostructure du texte mais ici pour une structure plus locale. Par exemple, préciser que l'objet textuel qui va être lu sera un titre (à l'aide d'une icône audio ou d'une phrase) permettra à l'auditeur de l'insérer en tant que nœud dans sa représentation globale du contenu, et de traiter les objets suivants comme étant relatifs au thème énoncé par le titre. Donner une méta-information en parallèle de l'oralisation d'un objet textuel peut se faire de 2 façons différentes : soit de manière « discrète » temporellement parlant en utilisant une modalité non-verbale par exemple avec une icône audio, un schéma vibratoire ou bien un schéma prosodique, ou de manière « continue » en spatialisant la lecture du texte (simulant ainsi une source sonore dans l'espace) ou en attribuant une voix particulière à sa lecture. Cette dernière méthode permet de rendre une ou plusieurs méta-informations disponibles pendant toute la durée de la lecture de l'objet, tandis que la méthode discrète ne rend l'information accessible que pendant une partie de la durée de la lecture. On préférera ainsi utiliser la méthode discrète avant ou après la lecture de l'objet, ou alors pour donner une méta-information portant sur un sous-objet de l'objet textuel en cours de lecture (ex. une emphase). Enfin, restituer une méta-information après la lecture d'un objet semble approprié dans le cas d'informations complémentaires à l'interprétation de son contenu, mais qui ne remettent pas en cause la sémantique de l'objet ou sa participation à la structure du texte, ce qui obligerait un travail cognitif de remaniement de la représentation élaborée.

À noter que cette réflexion est applicable aussi bien à la restitution d'informations de MFM qu'au signalement de la présence d'information. Au lieu de restituer directement une information, à cause de sa complexité par exemple, on peut simplement signaler la possibilité d'accéder à une information non restituée en utilisant par exemple une icône audio dédiée, information qui serait alors restituée au lancement d'une commande utilisateur dédiée. Cette méthode peut également être utilisée sans signalement de la présence d'information, en laissant l'utilisateur explorer (ou pas) les informations associées à un objet. Par exemple, une commande pourrait donner le titre de la section en cours de lecture, sans qu'il y ait besoin de signaler à chaque section la présence de cette information ; dans le cas d'une section sans titre, la commande renverrait alors un message d'erreur. Cela reste la

---

<sup>97</sup> Aussi appelées cartes heuristiques ou *mind maps*, ce sont des diagrammes représentant les concepts ou idées (ex. dans des bulles) et les liens entre ces concepts (ex. par des flèches ou traits), et ayant le plus souvent un concept ou une idée centrale.

méthode la plus neutre en termes de restitution, en laissant l'utilisateur libre de décider s'il souhaite ou non obtenir certains types d'information ; cette méthode implique cependant que l'utilisateur connaisse les commandes et les actions associées.

### 3.4. Restitution de la dimension fonctionnelle de la MFM

Comme évoqué dans la partie 2, l'inscription graphique du texte et de sa mise en forme permettent l'émergence de fonctions de lecture visuelle comme la lecture en diagonale, l'anticipation de la fin des objets textuels pendant la lecture, la lecture en diagonale, ou encore l'accès sélectif à certains types de phénomènes de MFM. La question de leur restitution nécessiterait tout d'abord une étude plus systématique de ces fonctions (nous en avons évoqué quelques-unes dans le Chapitre I et dans celui-ci), par exemple en utilisant un *oculomètre* auprès de lecteurs voyants parcourant un texte mis en forme. Ici, nous nous sommes centrés sur une seule fonction à titre d'exemple qui nous paraissait pertinente dans le cadre de la compréhension de texte, et qui a été restituée pour les évaluations présentées dans le Chapitre V : l'anticipation de la fin des objets pendant la lecture.

Ce phénomène, qui se produit en lecture visuelle, voit la vitesse de lecture ralentir avant une rupture visuelle (ex. fin de paragraphe), afin de mettre en place des processus cognitifs intégratifs [Hyönä 1995]. En effet, les ruptures graphiques marquent très souvent un changement thématique dans le discours, d'où la nécessité d'intégrer les informations précédemment lues afin de mieux pouvoir les relier aux informations de l'objet textuel suivant. L'anticipation de la fin des objets a ainsi des avantages cognitifs, rendus possibles par la vision périphérique qui permet justement d'anticiper la fin des objets. L'idée est ici de restituer cette fonction de lecture en spatialisant la voix de synthèse lisant de texte. Plus précisément, la lecture des objets commencera alors toujours depuis une source sonore située sur la gauche de l'auditeur, puis se déplacera vers la droite en arc de cercle pour arriver à un emplacement situé sur la droite de l'auditeur.

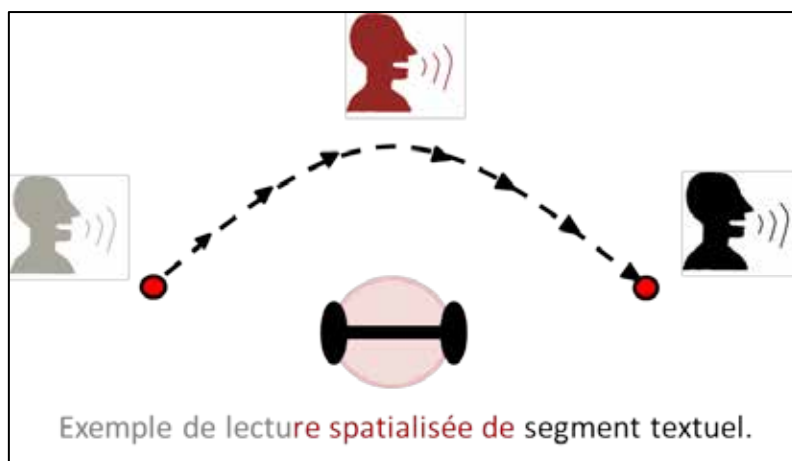


Figure 34 - Illustration de la restitution de la fonction d'anticipation à l'aide de son spatialisé

La lecture commence et se termine toujours aux mêmes emplacements dans l'espace ; le déplacement continu décrit par la source sonore permet ainsi à l'auditeur d'anticiper la fin de l'objet textuel, et accessoirement de connaître rapidement la taille de l'objet en évaluant la vitesse de déplacement de la source sonore. De plus, ce procédé permet de restituer la démarcation entre objets textuels à travers le changement de localisation de la source sonore qui revient sur la gauche à chaque nouvel objet.

La restitution de cette fonction, utilisée pour l'évaluation dans différentes stratégies de lecture, est présentée ici à titre d'exemple. D'autres fonctions de lecture visuelle permises par la MFM gagneraient à être restituées, ou tout du moins étudiées. On peut citer par exemple la lecture en diagonale sur laquelle certains travaux se sont déjà penchées en proposant des résumés de texte ou différents heuristiques de lecture [Gadde and Bolchini 2014; Arne et al. 2008], ou encore l'accès sélectif à certains objets en fonction de leur mise en forme, qui est déjà une fonction proposée par les lecteurs d'écran pour différents types d'objet (ex. accès à la liste des titres, des liens, des énumérations de la page courante). Nous pensons également qu'étudier la navigation non-visuelle en s'inspirant des mouvements oculaires et notamment des saccades régressives [Vitu 2005] pourrait être une source intéressante de nouveaux paradigmes de navigation, qui sont actuellement cantonnés à une navigation mot à mot, phrase à phrase, de paragraphe en paragraphe ou bien en accès sélectif.

La restitution de la dimension fonctionnelle de la MFM se heurte à la même problématique que la restitution des informations véhiculées par cette dernière. En effet, il existe de *nombreuses* fonctions différentes qui font écho à la flexibilité du système visuel et à la diversité des marques de MFM, qui sont de plus souvent spécifiques à certaines tâches ou certaines situations de lecture (une tâche d'exploration ou de recherche d'information n'observera pas les mêmes types de mouvements oculaires qu'une tâche de lecture linéaire). Ainsi, une méthodologie similaire à celle présentée dans la section 3.1 pour les informations de MFM devrait être appliquée afin de sélectionner les fonctions les plus pertinentes à restituer.

## **4. Proposition d'une stratégie de restitution de la MFM**

---

---

Cette partie du Chapitre constitue un exemple d'application des critères et réflexions exposées dans les parties précédentes autour de la restitution des informations de MFM formalisées. En suivant la typologie définie dans la partie 2, on présentera les différents choix réalisés pour chaque type d'information à restituer ainsi que les raisons argumentées de ces choix. On décrira notamment comment restituer ces informations, et pourquoi conserver ou supprimer tel ou tel type d'information. L'ensemble constitue une première *stratégie* de restitution qui a été réutilisée partiellement pour la définition des stratégies de restitution évaluées au sein de 2 expérimentations. Cette stratégie a été définie pour une tâche de compréhension d'un document expositif lu par synthèse vocale, qui représente le cadre expérimental principal de nos travaux. Ce choix est basé sur les résultats de la littérature portant sur les effets des *signaux visuels*, montrant leur rôle prépondérant dans la compréhension [Lemarié et al. 2012; Lemarié et al. 2006; Lorch 1989] (voir Chapitre I). De plus les textes expositifs, avec par exemple les manuels scolaires, présentent fréquemment des mises en forme riches ce qui les rend d'autant plus pertinents au regard de notre objectif d'évaluer plusieurs stratégies de restitution de la mise en voir (évaluations présentées au chapitre suivant). Nous avons employé le terme *stratégie* pour désigner nos choix de conceptions quant à la restitution de la MFM, afin de souligner l'importance de la prise en compte stratégique (et holistique) des caractéristiques des modalités, des informations à restituer mais aussi du contexte d'utilisation. Ainsi, la partie précédente était nécessaire dans l'optique de concevoir des stratégies de restitution pertinentes, dont nous présenterons ci-après une application à une stratégie particulière.



## 4.1. Restitution des informations locales de MFM

---

On présentera dans cette section la restitution des démarcations entre objets textuels, des statuts des objets et de leurs propriétés.

La **démarcation** entre objets textuels est ici restituée à travers le déplacement de la source sonore décrite dans la section 3.4, qui visait initialement à restituer la fonction d'anticipation de la fin des objets. Ainsi, le changement d'emplacement de la source sonore de gauche à droite permet de démarquer les objets entre eux pendant la lecture. Cette méthode permet de ne pas rallonger le temps de parcours des documents, les démarcations entre objets étant nombreuses dans un document mis en forme, et ne perturbe à priori pas l'interprétation du contenu verbal. L'acceptabilité de la synthèse vocale « spatialisée » reste cependant à investiguer, notamment lorsqu'elle est utilisée pour l'oralisation de documents longs.

Les **statuts** des objets textuels seront restitués différemment en fonction de leur importance d'après les critères définis dans la section 3.1, et donc notamment en fonction de leur importance pour la tâche de compréhension. Les statuts les plus importants pour la compréhension du contenu (notamment le statut de « titre ») seront restitués à l'aide d'icônes audio jouées avant la lecture de l'objet textuel; on pourra définir des familles d'earcon pour des statuts similaires (ex. titre de section/chapitre et titre de rubrique). Les autres statuts seront restitués en insérant du texte, dont la lecture pourra être accélérée à la manière de *spearcons* si le statut a une fréquence d'occurrence élevée dans le texte. Enfin, lorsque les objets sont imbriqués au sein du discours (intra-phrastique), alors on utilisera soit les propriétés de la voix pour restituer un statut particulier (ex. modification de la hauteur et de la vitesse d'énonciation pour les emphases, ou utilisation d'un schéma prosodique particulier), plutôt pour les objets textuels importants pour la compréhension afin de les rendre plus saillants, soit en rajoutant du texte de manière intégrée au discours (ex. remplacer une parenthèse à l'aide du texte « par exemple ») afin simplement d'assurer la bonne compréhension du contenu.

Les **propriétés** des objets, par exemple le niveau hiérarchique d'un titre, seront restitués de 2 façons différentes en fonction de leur importance : les plus importantes pour la compréhension seront intégrées sous forme de texte lu par une voix différente (une voix dédiée au métatexte à la manière d'une voix-off), les moins importantes seront intégrées directement au texte ou supprimées.

## 4.2. Restitution des informations transverses de MFM

---

Concernant les **relations** entre objets textuels, on distingue plusieurs restitutions différentes en fonction de la fréquence d'occurrence des relations et de leur nature.

Les relations de *composition* et la relation *intituler* concernent toutes deux, pour la plupart des documents, la quasi-totalité de leurs objets textuels. En effet la présence d'un titre implique par convention la distinction d'une section subséquente, *intitulée* par le titre qui la précède immédiatement ; de plus chaque section est un objet *composée* d'au moins 1 autre objet si la section n'est pas vide. Ainsi, afin d'éviter de surcharger le contenu informationnel du document on laissera l'utilisateur accéder activement à ces relations à l'aide de commandes dédiées. Concernant d'autres relations peu importantes pour la compréhension mais avec des fréquences d'occurrences plus faibles, comme les relations de *renvoi* ou *d'annotation* (ex. notes en marge), on signalera la présence

de ces relations par une icône audio après la lecture de chacun des objets textuels ciblés par la relation. L'utilisateur pourra ensuite à l'aide d'une commande dédiée accéder à l'objet « origine » de la relation. Cette méthode permet de signaler la présence de ces relations, ce qui étant donné leur faible fréquence d'occurrence semble nécessaire, sans entraver la lecture. Enfin, les relations les plus importantes impactant la compréhension de l'organisation des objets textuels et du discours (ex. *illustrer, exemplifier*) seront restituées verbalement à l'aide de métaphrases insérées après les objets ciblés par les relations et avant les objets d'origine de ces relations (ex. « illustration du paragraphe précédent »).

Enfin, pour les informations **transverses** de MFM, on ne restituera que celles concernant la structure organisationnelle du document, c'est-à-dire l'agencement et la composition de la titraille et celle des énumérations. Ces informations transverses seront restituées avant la lecture des objets textuels sur lesquels elles portent, et de manière verbale à l'aide de métaphrases discursives explicitant ces informations. On retrouve ce type de procédé dans l'étude des signaux visuels sous le terme vues d'ensemble (*overview*) [Lemarié et al. 2006], qui permettent (comme évoqué précédemment en 0) d'améliorer la compréhension du texte lorsqu'ils sont placés avant l. Par exemple, avant une nouvelle partie du document on pourra lire la métaphore « cette partie traitera tout d'abord de... dans la première section A, puis de ... dans la seconde section B ». La restitution des informations transverses peut être particulièrement utile pour aider l'auditeur à construire une première représentation de l'organisation générale du document, et donc de mieux le comprendre. Cependant, leur restitution peut désorienter les auditeurs étant donné que les objets visés par ces métaphrases sont souvent éloignés de l'emplacement (ou du moment dans la lecture) où l'on insérera ces informations transverses. Enfin, on pourrait également utiliser des « vues » du texte dédiées à la restitution d'informations transverses, accessibles à travers des commandes dédiées ; ces vues contiendraient l'ensemble des objets visés par les métaphrases transverses et permettrait l'exploration active de cette liste d'objets textuels. Par exemple, pour une stratégie de restitution implémentée dans un cadre expérimental, on utilisait une vue « table des matières » reprenant l'ensemble des titres du document.

### **4.3. Commentaires sur la stratégie de restitution**

---

La stratégie de restitution présentée de cette section a été conçue pour un cadre expérimental centré sur l'accès à des documents expositifs pour une tâche de compréhension. La stratégie devait donc être appréhendable rapidement par les participants aux expérimentations, et ainsi ne demander qu'un apprentissage limité en ce qui concerne le « vocabulaire » non-verbal utilisé. Cependant, des stratégies plus complètes, étendues à des types d'objets plus précis et plus atypiques pourraient tout à fait être définies ; l'influence du développement de l'expertise sur l'efficacité des stratégies de restitution, avec notamment le développement d'un traitement cognitif optimisé des indices non-verbaux par les utilisateurs, est un sujet d'étude intéressant que l'on pourrait relier au développement de la littéracie pour les lecteurs voyants qui vont par exemple apprendre à mieux utiliser la mise en forme des textes.

L'utilisation de plusieurs modalités pour la restitution permet d'optimiser cette dernière en prenant en compte leurs différentes caractéristiques, avantages et inconvénients, dans l'implémentation. De plus, l'utilisation *complémentaire* de plusieurs modalités peut amener qualitativement une plus grande satisfaction des utilisateurs du système [Bellik 1995]. Cet auteur souligne d'ailleurs

l'importance du dynamisme de la restitution, afin d'utiliser « la meilleure modalité ou la meilleure combinaison de modalités pour transmettre une information, et ce en fonction du contexte courant. »(p.200). Cette remarque souligne l'importance de l'adaptation au contexte c'est-à-dire à la fois à la tâche à réaliser et à l'utilisateur, grâce à une restitution paramétrable.

La navigation dans le document (le déplacement du *curseur virtuel* de lecture) fait partie intégrante de la restitution ; cependant, nous ne l'avons ici considérée que du point de vue de la MFM, en proposant 3 manières de naviguer : séquentiellement d'objet en objet (aller à l'objet textuel précédent ou au suivant), d'après les relations entre objets textuels (aller à l'objet qui annote l'objet courant), et enfin une navigation dans certaines *vues* du document reprenant des informations de MFM transverses (par exemple une table des matières). D'autres fonctions de navigation dans la structure hiérarchique du document pourraient être pertinentes ; [Smith et al. 2004] proposent plusieurs fonctions de ce type, ici dans le cas de la navigation dans une arborescence de fichiers, avec des résultats encourageants en termes de performance et de satisfaction des utilisateurs. On peut notamment citer la navigation entre « cousins » : dans le cas des sections d'un document, cela correspond à naviguer entre les sections d'un même niveau hiérarchique quelle que soit le titre englobant (ex. naviguer de la section III.4 à la section IV.1 directement).

Enfin, nous avons défini dans les stratégies de restitution implémentées (celle proposée ici, et celles que nous avons évaluées qui seront présentées au chapitre suivant) un ordre de lecture par défaut des objets textuels du document. Sans avoir besoin de naviguer, les utilisateurs peuvent laisser défiler la lecture des objets textuels successifs du document. Cet ordre de lecture était défini tout naturellement pour les documents utilisés étant donné qu'il était rendu explicite par l'agencement des objets textuels sur la page (qui se succédaient, de haut en bas et de gauche à droite). Or, l'ordre de lecture à proposer par défaut pour un document est une vraie question de recherche (d'ailleurs étudiée dans le cas de la numérisation de journaux [Ferilli et al. 2014]), lorsque l'on sort du cadre prototypique de documents linéaires, par exemple avec certaines pages web (voir Figure 35). En effet on peut retrouver sur le web des pages aux objets textuels disséminés sur l'espace de la page, qui peuvent être lus dans n'importe quel ordre, car ne présentant pas de relation de MFM particulière.

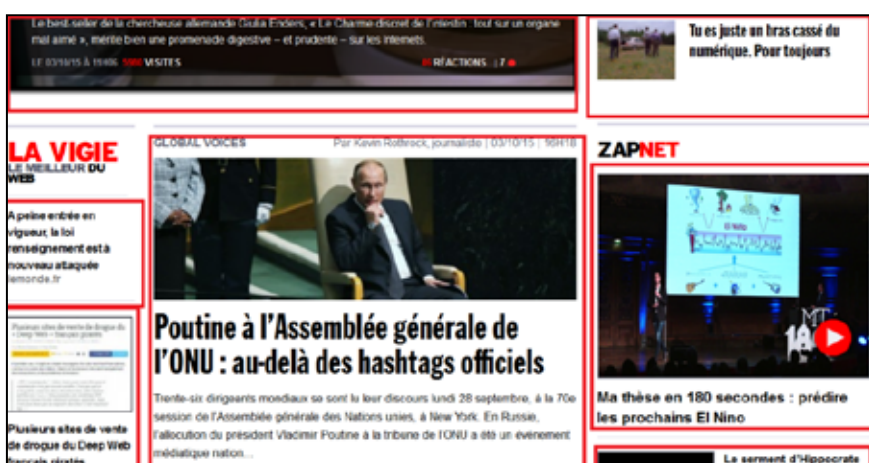


Figure 35 - Capture d'écran d'un site web journalistique

Le plus souvent ce type de phénomènes ne se rencontrent pas pour des contenus de type expositif ; on sort ainsi du cadre de la compréhension de contenu, visé par ces travaux. Il sera toutefois pertinent de se pencher sur cette question de l'ordre de lecture des objets textuels dans le cas

d'autres tâches, comme par exemple la recherche d'information, où en fonction des objectifs de l'utilisateur et du contexte d'accès on pourrait définir un ordre de lecture particulier, comme par exemple placer les objets textuels à priori les plus pertinents en début de page comme l'ont fait [Yesilada et al. 2007].

## 5. READ : une plateforme expérimentale pour implémenter et évaluer des stratégies d'accès non-visuelles aux documents

### 5.1. Introduction : pourquoi cette plateforme ?

Afin de pouvoir évaluer différentes stratégies de restitution de la MFM, dont celle présentée à la partie 4 précédente, nous avons développé une plateforme logicielle fonctionnant sous Windows, dénommée *READ (REstitution de l'Architecture des Documents)* [Sorin, Lemarié and Mojahid 2015]. Cette plateforme constitue un logiciel d'accès non-visuel *multimodal* basé sur une lecture par synthèse vocale, mais dont la flexibilité de l'architecture permet aisément d'implémenter n'importe quelle autre modalité pour la lecture. Nous présentons READ comme une plateforme expérimentale étant donné qu'elle est très adaptative, et permet donc différentes configurations dédiées. Cependant, parce qu'elle nécessite une mise en place manuelle pour sa configuration et l'adaptation des documents à « lire », c'est un outil plutôt réservé à un cadre expérimental de recherche pour le moment.

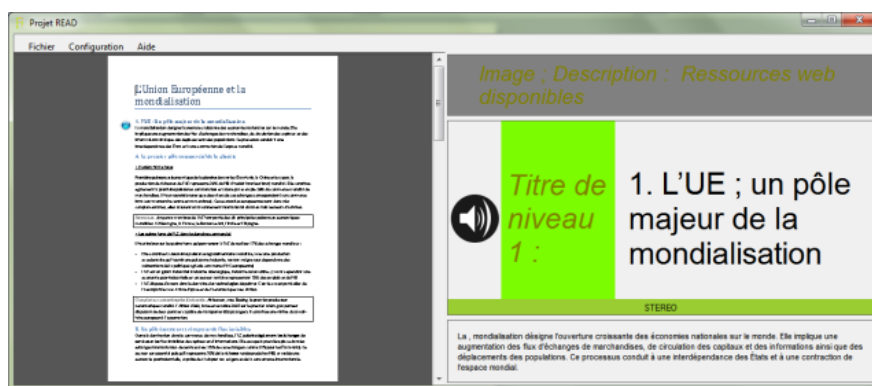


Figure 36 - Capture d'écran illustrative du logiciel READ

La création et la publication d'un tel outil d'accès non-visuel aux documents fait écho à plusieurs problématiques retrouvées dans la littérature. Tout d'abord, des progrès restent à réaliser dans le domaine de l'accessibilité des contenus numériques et des outils d'assistance : l'utilisation des lecteurs d'écrans est source de nombreuses frustrations [Lazar et al. 2007] ; de plus les normes d'accessibilité des contenus ne règlent qu'une partie du problème [Power et al. 2012]. Dans ce contexte, plusieurs paradigmes de « lecture » (accès non-visuel aux textes) sont expérimentés ; on peut citer notamment la multimodalité pour la restitution, utilisée aussi bien pour l'accès aux textes [El-glaly and Quek 2014] que pour l'accès aux contenus complexes comme les mathématiques [Bates and Fitzpatrick 2010]. On voit également émerger des essais de discours concourants (plusieurs voix lisant en même temps des contenus différents) [Guerreiro and Gonçalves 2014], l'utilisation d'icônes

audio remplaçant des segments de texte [Hussain et al. 2014] ou encore l'utilisation d'audio spatialisé pour la lecture [Goose and Möller 1999].

L'évaluation de tels paradigmes d'accès nécessite le développement d'outils dédiés, ou alors la modification d'outils existants comme les lecteurs d'écrans à condition qu'ils soient publiés en *open source*<sup>98</sup>. Même lorsque c'est le cas, ces outils présentent souvent une architecture complexe (un aperçu de l'architecture globale des lecteurs d'écran est présentée par [Blenkhorn 2000]) qui peut rendre le développement coûteux (en temps et en ressources) et donc inadapté à un cadre de recherche où les objectifs principaux sont expérimentaux.

Nous pensons que READ pourrait fournir une alternative intéressante à des développements dédiés pour ce type de projet, en permettant d'implémenter différentes modalités de lecture, de configurer à souhait la restitution du document et d'implémenter aisément de nouvelles fonctions de navigation ou de présentation de l'information. En fonction du projet, l'implémentation peut ne nécessiter aucun développement en passant simplement par la modification des fichiers de configuration. Dans d'autres cas, des modules C# peuvent être greffés à l'application pour y incorporer des comportements ou modalités de sortie particuliers. À notre connaissance, il n'existe aucune plateforme expérimentale offrant une telle modularité et flexibilité pour l'accès non-visuel aux documents. Les outils existants comme le navigateur audio CYNTHIA [Raynal and Serrurier 2002] ne permettent qu'une configuration limitée de la restitution.

READ fonctionne sous le système d'exploitation MS-Windows, et sera prochainement publié sous licence CC BY-NC-SA<sup>99</sup> ; une perspective future est de voir émerger plusieurs versions et extensions de l'application suite à des développements collaboratifs et leur mise en commun à travers des plateformes de dépôt publiques comme *GitHub*<sup>100</sup>.

Enfin, la stabilité du logiciel a été éprouvée à travers les expérimentations rapportées dans le chapitre suivant ; durant les 40 heures totales d'utilisation du logiciel (portant sur 2 documents différents) par 60 volontaires voyants et 12 non-voyants, aucune erreur ni crash n'est apparu et la lecture est restée stable et fluide.

## 5.2. Présentation générale de READ

---

### 5.2.a. Fonctionnement et objectifs

---

READ est donc un outil pour la recherche permettant d'implémenter et d'évaluer différentes stratégies d'accès non-visuel à des documents textuels. L'accès aux documents dépend entièrement du, ou des, configurations qui auront été définies dans des fichiers dédiés ; on autorise la coexistence de plusieurs fichiers de configurations afin de pouvoir tester plusieurs stratégies différentes sur une même expérimentation, mais aussi de pouvoir changer de stratégie en cours de lecture, auquel cas on peut voir ces stratégies comme des « modes » de lecture. L'intérêt de disposer de plusieurs *modes* de lectures, pouvant faire partie d'une stratégie d'accès globale pour une expérimentation,

---

<sup>98</sup> L'open source est un modèle de développement de logiciels pouvant aboutir à plusieurs types de licence gratuites avec différentes restrictions, donnant cependant accès au code source de l'application et laissant donc possible la modification du programme par des tiers.

<sup>99</sup> Creative Commons Attribution (by) NonCommercial ShareAlike

<sup>100</sup> <https://github.com/>

est de pouvoir adapter chaque mode à une situation de lecture particulière (ex. recherche d'information) et éviter ainsi de surcharger la restitution d'informations plus ou moins pertinentes.

Chaque configuration spécifie comment l'accès aux documents doit être réalisé, c'est-à-dire comment transformer le texte et les métadonnées du document en une restitution non-visuelle ; il est donc essentiel que les documents traités par le système respectent un format commun, celui dont les paramètres de restitution sont décrits dans la configuration. En d'autres termes, afin de pouvoir être accédé via READ, les documents textuels d'origine (documents sources sur la Figure 37) doivent être au format XML (documents formatés) et respecter un certain formalisme, c'est-à-dire un vocabulaire de balises et d'attributs<sup>101</sup>. Ce formalisme n'est pas arrêté et doit simplement correspondre au vocabulaire employé dans les fichiers de configuration afin d'être pris en compte dans la restitution. Dans notre cas nous avons choisi un vocabulaire permettant de décrire les informations de MFM précédemment extraites par le processus de (voir section 5.3).

Dans le cas de notre implémentation de stratégies de restitution de la MFM (dont les évaluations sont présentées au chapitre suivant), le principal mode de restitution du contenu des documents était la lecture par synthèse vocale. Les fichiers de configuration XML permettaient ainsi de gérer l'utilisation de différentes voix de synthèse (quelle(s) voix, avec quels paramètres), la lecture d'icônes audio pendant l'accès ainsi que la spatialisation du son (des voix ou icônes audio). D'autres modalités de sortie (ex. plage braille, dispositif vibrant) peuvent être implémentées en développant des plugins (extensions) dédiés ; cette architecture modulaire permettra à READ d'être réutilisé pour de futures évaluations. Nous détaillerons les spécificités de la configuration et des fichiers correspondants par la suite.

---

<sup>101</sup> En XML et dans les langages de markup en général chaque balise peut contenir un ensemble d'attributs ; par exemple `<balise attribut1="valeur1" attribut2="valeur2"> texte.... </balise>`.

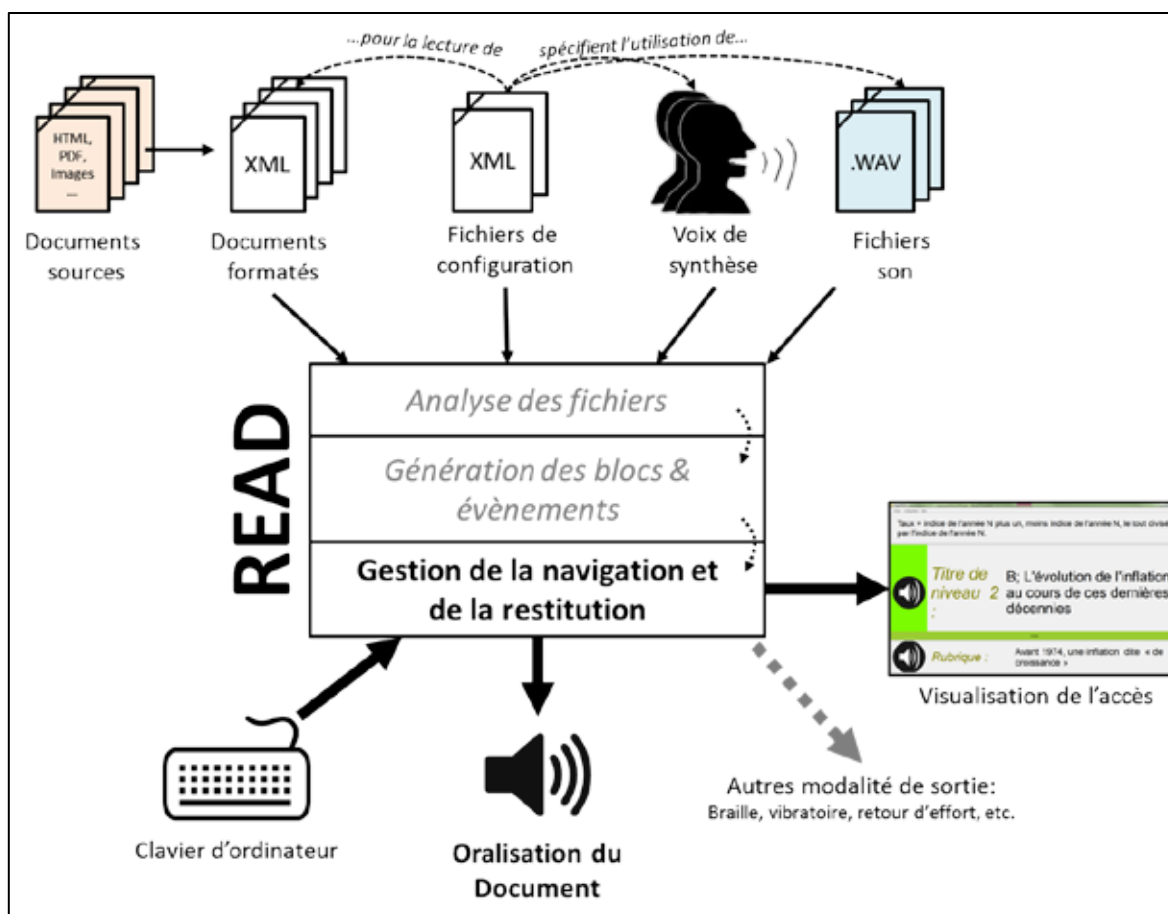


Figure 37 - Schéma du fonctionnement global de READ

À partir d'un document formaté et pour un fichier de configuration donné<sup>102</sup>, le système READ va générer un ensemble de *blocs*, qui correspondent dans notre cas aux objets textuels du document, eux même constitués d'un certain nombre d'*événements* de lecture. READ dispose d'un module de visualisation qui permet d'afficher les blocs et événements en cours de lecture ; on retrouve sur la capture d'écran de la visualisation de READ à droite de la Figure 37. On retrouve 3 blocs affichés verticalement sur 3 bandeaux ; le bloc du milieu (celui en cours de lecture) comprend 3 événements (affichés horizontalement) : 1 icône audio, 1 lecture de métatexte et 1 lecture de texte. A minima, un bloc contiendra un seul événement de texte qui sera (par défaut dans l'implémentation actuelle) la lecture par synthèse vocale de son contenu. La représentation d'un document sous forme de *blocs* constitués d'*événements* est une manière générique de représenter une séquence d'informations. Ce mode de représentation est lié à la nature séquentielle de l'accès à l'information dans des modes non-visuels. Par exemple pour la restitution de la MFM, nous avons écrit un fichier de configuration qui comprend 3 types d'événements utilisables pour la construction des blocs : la lecture d'icônes audio, la lecture de métatexte et l'inclusion de texte au discours. Ces événements pourront avoir, au cas par cas, des paramètres particuliers (paramètre de la voix, changement de voix, spatialisation de la lecture) et inclure les attributs du contenu XML en cours de lecture.

Avant de pouvoir générer les blocs et événements, READ commence par analyser le document formaté en XML ainsi que le(s) fichier(s) de configuration ; cette phase consiste à lire les fichiers et

<sup>102</sup> Dans le cas de nos expérimentations un document formaté était un document MFM dont les balises et attributs représentaient une formalisation de la MFM du document source.

représenter leurs contenus respectifs en mémoire. La configuration (courante) est ensuite appliquée au contenu du document, générant ainsi une séquence de blocs comportant chacun un ou plusieurs évènements. Les fichiers de configuration doivent donc spécifier, entre autres, quelles balises du document vont être génératrices de nouveaux blocs de lecture, et quels évènements doivent être associés à chaque balise. Ci-après un extrait d'un fichier de configuration.

```
<!-- item d'une énumération -->
<item mode="spatial" newblock="true">
  <event type="txt">
    <attribute name="role" />;
    <short_break/>
  </event>
  <event type="txt">
    <inner_txt/>
    <long_break/>
  </event>
</item>
```

Figure 38 - Extrait d'un fichier de configuration

Dans l'exemple donné par la Figure 38, la configuration spécifie que chaque balise `<item>` rencontrée dans le document XML formera un nouveau bloc de lecture (`newblock="true"`) ; chaque « bloc-item » contiendra alors deux évènements : le premier sera de type *lecture de texte* et contiendra l'attribut « rôle » de l'item lorsque celui-ci existe, et le second, également de type *lecture de texte*, contiendra le texte présent entre la balise `<item>` ouvrante du document et la balise fermante la plus proche (noté `<inner_txt/>` dans le fichier). De plus, chaque évènement de lecture sera ici suivi d'une pause dans l'énonciation définies avec les balises `<short_break/>` et `<long_break/>` ; la durée de ces pauses est définie dans une partie dédiée du fichier de configuration. Enfin, la configuration peut également inclure des éléments de relations entre objets de MFM, relations qui sont définies dans le document formaté. Pour mieux comprendre le rôle de la configuration de la lecture, ci-après sur la Figure 39 un exemple complet de restitution d'un extrait de document.

Une fois la séquence de blocs de lecture générée, la lecture (restitution) du document débute au premier bloc de la séquence. Les flèches du clavier sont utilisées par défaut pour naviguer linéairement entre les blocs et évènements de lecture ; des touches spécifiques peuvent être utilisées pour naviguer entre les blocs d'après les relations de MFM définies dans le document, ou pour effectuer des actions de navigation particulières.



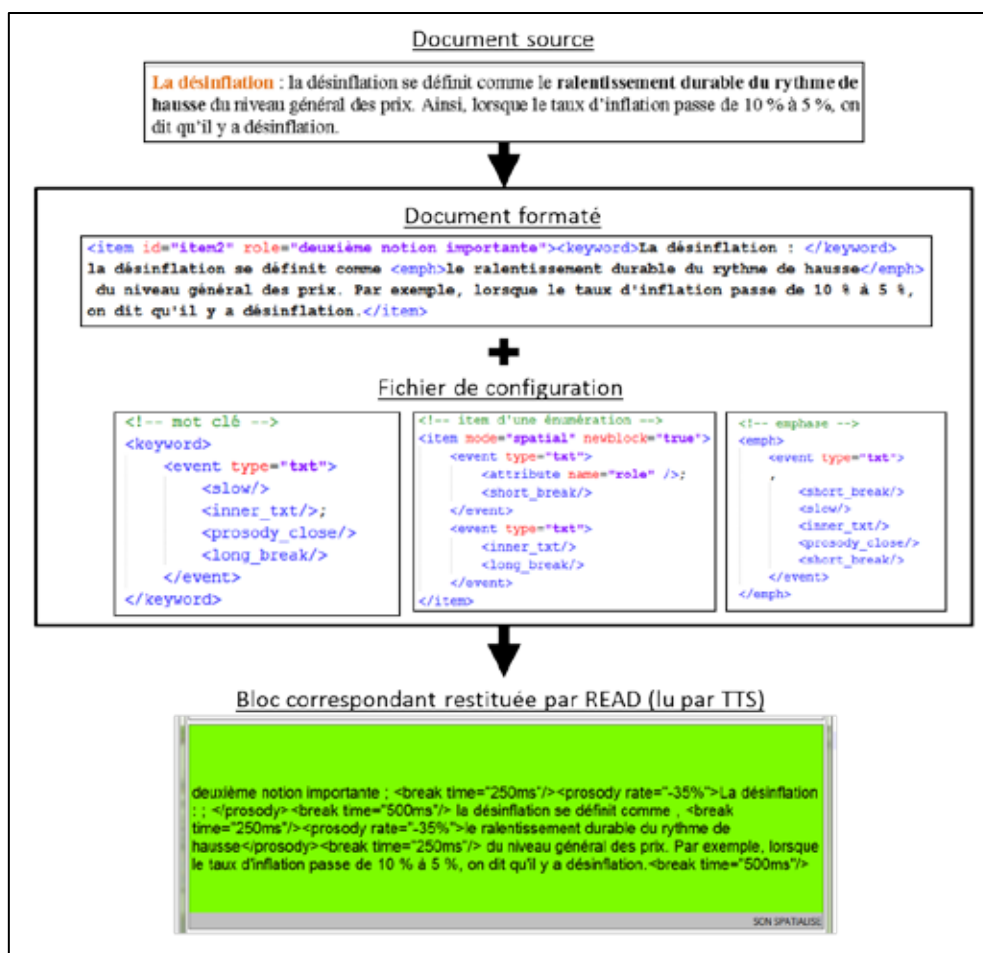


Figure 39 - Illustration du résultat de la configuration de la lecture d'un objet textuel

Afin d'en donner un vue d'ensemble, nous allons maintenant passer en revue les différentes fonctionnalités et caractéristiques de READ.

Analyse et génération : l'analyse consiste à représenter en mémoire le contenu du document formaté en cours de lecture, ainsi que du ou des fichiers de configurations présents dans un dossier dédié. La configuration sélectionnée dans l'interface est ensuite appliqué au contenu du document, générant une séquence de blocs composés de différents évènements de lecture.

Configuration de la restitution : permet de définir les types d'évènements utilisables (ex. icônes audio), les paramètres de ces évènements (volume, vitesse, paramètres des voix), et à quelles éléments du document associer ces évènements. Ainsi, à chaque fois qu'une des balises listées dans le fichier de configuration sélectionné apparait dans le document en cours de lecture, READ génère la liste d'évènements correspondants. Les possibilités de configuration offertes par READ permettent de définir finement différentes stratégies de restitution du contenu des documents utilisés. Dans un cadre expérimental cela peut correspondre à différentes conditions, comme ce fut le cas pour les expérimentations présentées au chapitre suivant.

Visualisation : comme on peut l'apercevoir sur les différentes figures précédentes, READ fournit une visualisation basique du bloc de lecture courant ainsi que des évènements qu'il inclut, et de même pour les blocs précédant et suivant le bloc courant (l'évènement en cours de lecture est affiché sur fond vert). Ces 3 visualisations correspondent aux 3 bandeaux que l'on aperçoit sur la droite des

captures d'écran de READ. Enfin, comme montré sur la Figure 36, si une version PDF du document d'origine existe, READ permet son affichage dans un encart sur la gauche de la fenêtre, et la synchronisation de cet affichage avec la navigation dans le document. Les visualisations présentées ici ont pour but de pouvoir contrôler la restitution durant les expérimentations, vérifier la configuration et l'absence d'erreurs par exemple, mais aussi de pouvoir utiliser READ en tant que démonstrateur.

Navigation : Les utilisateurs peuvent choisir de laisser READ restituer linéairement le document bloc par bloc, mais peuvent également naviguer entre les blocs représentant le contenu du document à l'aide de raccourcis clavier. Différentes fonctions de navigation ont été définies dans l'implémentation réalisée pour restituer la MFM ; il est également possible de rajouter des fonctions de navigation en développant des extensions au système (voir ci-après dans cette section).

Développements d'extensions (plugins) au système : Le développement d'extensions à READ permet d'étendre ses capacités de restitution (ex. nouvelles modalités de sortie) ou de définir de nouvelles fonctions de navigation. Cette modularité devrait permettre à READ d'être utilisé dans différents contextes expérimentaux, en nécessitant un minimum de développements pour s'interfacer avec un maximum de dispositifs.

Fichiers « trace » : à des fins d'analyses, chaque action de l'utilisateur et chaque évènement du système est conservé dans un fichier trace au format texte ; y sont inscrits : l'heure exacte du déclenchement, la nature de l'évènement et ses différents paramètres. Ces fichiers permettent d'analyser les comportements des utilisateurs.

### ***5.2.b. Données techniques sur le système et son fonctionnement***

---

READ a été développé en C#, langage créé par Microsoft, qui fonctionne à l'aide du framework .NET<sup>103</sup> 3.5 ; toute machine sur laquelle est installé .NET 3.5 ou une version plus récente peut donc utiliser READ. Le choix de ce langage a été motivé par sa gestion native des voix SAPI (Speech Application Programming Interface), qui est un standard pour le développement et l'utilisation des voix de synthèse sous Windows. À noter que ce standard est suivi par de nombreuses voix de synthèse, et qu'il est pris en compte par la quasi-totalité des lecteurs d'écran existant. Ainsi n'importe quelle voix de synthèse SAPI4 (ou SAPI5) installée sur la machine peut être réutilisée dans READ. Pour la spatialisation du son, READ utilise la bibliothèque XNA Audio (réutilisation libre et gratuite) qui fait partie d'un ensemble d'outils dédiés initialement au développement de jeux-vidéos : XNA Framework.

Les documents accédés via le système doivent être en XML (eXtended Markup Language) ; le choix de ce format est lié à son interopérabilité, et à la nature des langages de *markup* qui permettent d'imbriquer des balises entre elles et donc de reproduire l'imbrication visuelle ou logique des objets textuels. Les documents XML à restituer via READ ne sont pas contraints par un vocabulaire XML particulier, on pourrait par exemple utiliser des documents HTML (HTML étant un sous-langage XML). La seule contrainte que doit respecter le vocabulaire de balises utilisées dans le document (son format) est qu'il doit être en correspondance avec le vocabulaire utilisé dans les fichiers de configuration ; les balises ne respectant pas ce critère seront ignorées (non restituées). Cependant, si

---

<sup>103</sup> .NET a plusieurs composants ayant différentes licences; toutes sont cependant libres de droit pour la réutilisation.

techniquement READ n'est pas limité par les formats (on pourrait par exemple l'utiliser pour accéder à des documents HTML), la pertinence de la restitution fournie par READ dépend entièrement de la structure du document, de sa cohérence, et des métadonnées qui y sont incorporées. En d'autres termes, READ peut techniquement lire tout fichier balisé, mais la pertinence de la restitution et son évaluation sont à la charge des chercheurs ou testeurs qui mettent en place ces situations de restitution.

Les fichiers de configuration listent un ensemble de balises (ex. <item>) et permettent d'associer à chacune une série d'évènements de lecture, avec éventuellement des paramètres particuliers (ex. son spatialisé) ; les évènements peuvent être paramétrés d'après les attributs des balises. Ainsi, à chaque fois qu'une des balises listées dans le fichier de configuration sélectionné apparaît dans le document en cours de lecture, READ génère la liste d'évènements correspondants. La configuration peut également être conditionnelle : si une balise possède un attribut avec une certaine valeur, alors lui attribuer cette configuration, si non, lui attribuer une configuration différente. Si une balise inconnue est rencontrée dans le document, alors suivant ce que spécifie le fichier de configuration la balise sera ignorée ou on lui appliquera une configuration « par défaut ». Nous avons par exemple utilisé un fichier de configuration permettant de ne restituer que la table des matières du document lu, ignorant ainsi les balises qui ne correspondent pas à la titrairie.

En ce qui concerne la navigation, certaines fonctions ont été implémentées (voir section suivante 5.3) et associées à des raccourcis clavier. La navigation est sécurisée et lorsque l'utilisateur effectue une action « interdite » (ex. vouloir aller au bloc précédant le premier bloc de lecture), le message d'erreur vocal correspondant est lu. À noter que les messages d'erreur sont également configurables. En l'état actuel, le système ne permet pas de naviguer mot à mot ou phrase à phrase mais uniquement entre évènements et blocs. La raison est technique : afin que les synthèses vocales lisent les phrases avec un schéma prosodique adapté (schémas utilisés par défaut pour le « naturel » des voix), elles doivent prendre en entrées des phrases complètes ; si on leur envoie séparément chaque mot d'une phrase, la voix sera perçue comme « robotique » voir « incohérente » dans sa prononciation relativement à la syntaxe des phrases. Or, afin de pouvoir retourner au mot ou à la phrase précédente, il sera nécessaire dans de futurs développements de contourner cette limite. Ces fonctions n'étaient pas utiles aux expérimentations, et pour cette raison, elles n'ont pas encore été implémentées.

Enfin, concernant le développement d'extensions au système pour l'utilisation de nouvelles modalités de sortie ou l'implémentation de nouvelles fonctions de navigation, elles doivent être écrites en C#, implémenter les interfaces de programmation définies, et être placées dans le dossier du programme réservé aux plugins. Au démarrage, le programme prend alors en compte tous les plugins du répertoire et met à jour l'application en fonction. Nous produirons une documentation pour le développement des plugins à destination des futurs usagers.

### **5.3. Formats et modalités utilisées pour restituer la MFM**

---

Dans les sections précédentes nous avons présenté le fonctionnement général du système ainsi que divers points techniques quant à son développement. Dans cette section nous allons préciser les formats de fichiers et les modalités de restitution utilisées avec READ dans le cadre de nos expérimentations sur la restitution de la MFM.

Chaque document à restituer dont la MFM a été formalisée au préalable devra être au format XML. Le vocabulaire de balises et d'attributs utilisé pour décrire la MFM dans notre corpus de documents est disponible en Chapitre V4.5. Ce vocabulaire comprend des balises délimitant les objets textuels (leur contenu textuel) et leur donnant leurs statuts de MFM (ex. `<item>...</item>`) et des attributs permettant d'explicitier les propriétés des objets et leurs identifiants (ex. `<titre niveau="2" id="titre1">...</titre>`). L'imbrication des balises permet de rendre compte de l'imbrication visuelle et logique des objets textuels (ex. `<section> <paragraphe>...</section> </paragraphe>`). Enfin, les relations entre objets seront décrites à l'aide de balises spécifiques, qui contiennent à la fois la/les identifiants des objets cibles de la relation, la/les identifiants des objets d'origine de la relation, et enfin le verbe performatif associé (ex. `<rel origine="ndbp1" cible="label1;bdt3" verbe="annoter"/>`). Pour le moment, le vocabulaire XML appliqué aux documents du corpus utilisés dans le cadre de nos expérimentations est un vocabulaire ad-hoc. Dans le futur, à la suite d'études de la formalisation de la MFM sur différents corpus de documents, il pourra être intéressant d'écrire la DTD (Document Type Definition) contraignant l'organisation des balises et la présence de leurs attributs afin de permettre un échange de documents annotés et leur utilisation à une grande échelle.

En ce qui concerne les modalités de restitution implémentées dans READ, il est pour le moment possible de jouer sur les paramètres suivants dans les fichiers de configuration :

- § Les voix de synthèse, dont la plupart des paramètres sont ajustables à l'aide de balises SSML (Speech Synthesis Markup Language)<sup>104</sup> insérées dans le texte à lire, qui peuvent être définies dans les fichiers de configuration.
  - Changement de la voix de lecture utilisée (parmi les voix SAPI installées).
  - Modification du volume de la voix.
  - Modification du *pitch* (hauteur) de la voix.
  - Rajout de pauses dans l'énonciation de différentes durées.
  - Utilisation de schémas prosodiques prédéfinis (le SSML permet d'utiliser nativement certains schémas prosodiques et sont donc théoriquement utilisables dans les fichiers de configuration ; cependant les voix de synthèses utilisées pour les tests et/ou la librairie *System.Speech* utilisée ne semblent pas prendre en compte ces schémas).
- Les icônes audio stockées dans des fichiers WAV.
- L'ajout de texte en fonction des propriétés des objets textuels.
- La spatialisation de toute sortie sonore (parole ou icônes audio), qui peut être réalisée en définissant des sources fixes ou mouvantes.

De futurs développements devraient permettre d'utiliser plusieurs événements en parallèle pendant la lecture d'un bloc (ex. jouer une icône audio et lire le texte en même temps). Cette fonctionnalité n'a pas été implémentée car elle n'était pas utile du point de vue de nos expérimentations.

Enfin, nous avons implémenté les fonctions de navigation suivantes : aller au bloc suivant, aller au bloc précédent, aller à l'évènement suivant/précédent. D'autres fonctions permettent de naviguer en fonction des relations entre blocs ont été implémentées : aller au renvoi annotant le bloc courant (si elle existe), aller au titre de la section courante, aller à la note de bas de page annotant le bloc

---

<sup>104</sup> Langage de Markup standardisé permettant le contrôle des voix de synthèses dans des applications Web, livres audio et serveurs vocaux, par exemple.

courant. Il est également possible de lire un message listant la hiérarchie de titres intitulant la section courante (pour cette section le message contiendrait les titres du chapitre IV, de la partie V et de la section 3). Enfin l'utilisateur peut changer la configuration appliquée au document et ce en cours de lecture. Toutes les fonctions citées ici n'étaient cependant pas disponibles pendant les expérimentations ; pour plus de détails voir le détail des stratégies de restitution de la MFM évaluées dans le Chapitre V.

## **5.4. Perspectives et limites**

---

La principale limite de READ est que l'édition des fichiers de contenu et des fichiers de configuration se fait pour le moment manuellement, ce qui peut être un processus relativement long en fonction de la taille du document et du nombre de types d'objets textuels à configurer ; l'avantage est qu'il n'y a pas de restrictions sur le document source puisqu'il doit être manuellement transposé et annoté en XML. Cependant, les contraintes et limites de l'édition manuelle semblent acceptables pour un usage à des fins de recherche. Une autre des limites de READ est que le système a été conçu pour un accès non-visuel aux contenus ; la visualisation qu'il propose n'est pas paramétrable et ne peut être modifiée qu'en modifiant le code source. Ainsi pour mettre en place des stratégies de restitution visuelle ET non-visuelle, pour tests auprès de malvoyants par exemple, il serait alors nécessaire de modifier le code source et d'effectuer les développements nécessaires à la visualisation souhaitée. Enfin, l'architecture interne n'a pas été prévue pour prendre en compte des entrées utilisateur autres que les raccourcis clavier ; bien que la difficulté de tels développements soit relativement minime, ils ne peuvent pas se faire par le biais de plugins et impliquent nécessairement de modifier le code interne au logiciel.

Cependant, READ offre de nombreuses possibilités de restitution qui pourraient être utilisées pour conduire des expérimentations à des fins très diverses comme l'apprentissage de documents pédagogiques, la recherche d'information dans des pages web ou encore l'exploration de documents inconnus, et ce en utilisant des paradigmes de navigation et de restitution potentiellement novateurs, sans avoir eu à développer un outil dédié. Ces restitutions pourront, sous réserve du développement des plugins correspondants et de l'existence de drivers Windows pour les périphériques, utiliser des dispositifs vibratoires, une plage braille ou encore des dispositifs à retour d'effort. Nous prévoyons de publier READ sous une licence libre (Non-Commercial Share Alike) lorsque les derniers développements importants seront achevés.

## **6. Conclusion**

---

---

L'accès à l'information dans des modes non-visuels présente des caractéristiques radicalement différentes de l'accès visuel, dans notre cas de la lecture de texte écrit. La séquentialité de l'accès à l'information dans ces modes, et la charge cognitive qui peut en résulter, impose de faire des choix sur ce qu'il faudrait restituer et comment. Ainsi, choisir une stratégie de restitution ou plus généralement choisir des méthodes et moyens d'accès non-visuels à l'information est une tâche complexe, lorsque l'on considère l'ensemble des modalités et des différents paramètres de la restitution sur lesquels on peut jouer pour tendre vers un optimum théorique. Un moyen de tendre vers cet optimum en pratique est de considérer pour la restitution à la fois la tâche à réaliser par l'utilisateur mais aussi ses caractéristiques, afin de pouvoir réduire le champ d'étude.

Dans ce chapitre nous avons étudié plusieurs modalités orales utilisables pour la restitution afin de pouvoir tirer le meilleur parti de leurs spécificités. Nous avons également présenté plusieurs réflexions et heuristiques donnant des critères de décision pour choisir le méta-contenu à restituer, et quand le restituer dans le cadre d'un accès séquentiel. L'ensemble de ces heuristiques et réflexions ont été utilisées pour proposer une stratégie de restitution dans la partie 4 de ce Chapitre. Cette stratégie ne représente pas un optimum absolu mais une proposition d'implémentation des réflexions présentées précédemment dans ce chapitre ; nous l'avons partiellement réutilisée dans la stratégie de restitution dite optimale que nous évaluons au Chapitre V. Nous nous sommes cependant limité au mode audio pour la stratégie de restitution proposée, qui, bien que présentant plusieurs avantages sur le braille (évoqués en introduction), reste problématique dans certains contextes d'utilisation. L'oral étant le seul mode de communication directe entre déficients visuels et non déficients, un périphérique informatique proposant une restitution orale peut rentrer en concurrence avec le discours oral par exemple de l'enseignant, dans un contexte scolaire. L'oral pose donc un problème attentionnel, étant donné que le traitement parallèle de deux discours concourants est extrêmement limité, et dans la plupart des cas ignore complètement l'un des deux discours [Allport et al. 1972], exception faite des effets *pop-out* attentionnels associés à certains mots clés (ex. le nom de l'auditeur) plus connus sous le nom effet *cocktail-party*.

L'utilisation du mode audio pose également certains problèmes pratiques dans son utilisation ; pour des raisons de confidentialité ou de discrétion [Ahmed 2015] il peut s'avérer nécessaire d'utiliser des écouteurs ou un casque audio pour l'écoute. Leur utilisation peut occulter les sons de l'environnement posant, au mieux, des problèmes d'accès au discours oral des interlocuteurs, et au pire des problèmes de sécurité (ex. en tant que piéton en ville). Le problème inverse peut également se poser : un environnement trop bruyant peut empêcher l'accès à la restitution orale du périphérique. Le problème de l'occultation de l'environnement peut cependant être partiellement résolu en ne portant qu'un écouteur sur les deux<sup>105</sup>, ou en utilisant des casques à *conduction osseuses*, qui transmettent le son par les vibrations de transducteurs électromécaniques en contact avec les os crâniens occipitaux ou temporaux<sup>106</sup>. Ce type de casque a l'avantage de ne pas obstruer les conduits auditifs, et d'être (selon les constructeurs) plus efficaces en environnement bruyant [Jzool.com n.d.]. Quoi qu'il en soit, le port d'un casque ou d'écouteurs peut être mal accepté par les personnes déficientes visuelles car le dispositif peut être perçu comme soulignant leur déficience et donc représentant un marqueur d'exclusion sociale ; [Nevile 2005] évoque le fait que les personnes handicapées ne veulent pas être traitées comme telles et souhaitent avoir les mêmes moyens que tout un chacun pour participer à la vie sociale. Ainsi, à cause des limites posées par le mode audio, une première application de la stratégie de restitution proposée est l'étude et la compréhension de documents pédagogiques dans un cadre de travail à la maison.

Pour revenir aux réflexions et heuristiques présentées dans le cadre de la conception de stratégies de restitution de la MFM, au-delà de leur application à la restitution de la MFM des documents, ces dernières sont généralisables dans une certaine mesure à l'accès non-visuel *multimodal* aux contenus textuels. Cet accès repose actuellement sur le braille, la synthèse vocale et différentes fonctions de navigation, et pourrait donc bénéficier de l'utilisation conjointe d'autres modalités. De

<sup>105</sup> Cette solution pose cependant problème si la restitution utilise du son spatialisé (voir Chapitre IV3.2.c) qui nécessite le port des deux écouteurs.

<sup>106</sup> Un exemple d'un tel casque peut être visionné ici : [http://jzool.com/en/page/bone\\_conduction](http://jzool.com/en/page/bone_conduction)

plus, les informations de MFM dont on étudiait ici la restitution ne constituent qu'une des « couche » d'informations qu'il est possible d'extraire d'un document écrit ; la conception de stratégies de restitution (ou d'accès non-visuel) pourrait également prendre en compte d'autres couches d'informations comme par exemple les intentions pédagogiques ayant structuré les manuels scolaires, ou les rôles de certains objets HTML pour la navigation intra et inter pages Web.

Les travaux sur la multimodalité pourraient contribuer à la restitution d'informations relatives aux textes. [Bellik 1995] montre d'ailleurs que l'utilisation de plusieurs modalités dans une interface non-visuelle permet une bonne efficacité et une satisfaction des utilisateurs. Une autre perspective est l'utilisation d'autres modalités que le clavier en entrée, par exemple avec des commandes vocales qui pourrait diminuer les contraintes d'apprentissage de raccourcis clavier en offrant une interface d'entrée plus naturelle dans le sens où elle suggère un dialogue oral.

Restituer l'information de mise en forme permettrait des avancées fonctionnelles dans l'accès non-visuel aux textes (nous verrons ce qu'il en est dans le Chapitre V), mais pourrait également permettre de fournir une expérience perceptive ayant une composante esthétique et peut être même sociale. Ce n'est qu'une hypothèse, mais l'utilisation de plusieurs modalités dans le cadre de l'accès non-visuel aux textes (icônes audio avec une musicalité, voix de synthèse utilisant des schémas prosodiques avancés) pourrait enrichir la perception des utilisateurs par une expérience plus immersive et plus riche, qui pourrait être comparable à la lecture d'un texte écrit mis en forme. Si cette utilisation débouche à terme sur la mise en place de standards, de certaines conventions partagées pour l'accès non-visuel et multimodal aux textes, alors l'expérience de « lecture non-visuelle » pourrait être qualitativement améliorée car socialement partagée [Declerck et al. 2009].

Enfin, READ représente une contribution de nos travaux à la recherche sur l'accessibilité et les interactions non-visuelles. Cette plateforme, unique dans son genre à notre connaissance, offre de nombreuses possibilités de restitution qui pourraient être utilisées pour conduire des expérimentations à des fins très diverses. READ pourrait être utilisé pour implémenter des stratégies d'accès visant l'apprentissage de documents pédagogiques, la recherche d'information dans des pages web ou encore l'exploration de documents inconnus, et ce en utilisant des paradigmes de navigation et de restitution potentiellement novateurs, sans avoir eu à développer un outil dédié. Ces restitutions pourront, sous réserve du développement des plugins correspondants et de l'existence de drivers Windows pour les périphériques, utiliser des dispositifs vibratoires, une plage braille ou encore des dispositifs à retour d'effort. Nous prévoyons de publier READ sous une licence libre (Non-Commercial Share Alike) lorsque les derniers développements importants seront effectués. À noter que la stabilité du logiciel a été éprouvée durant les expérimentations rapportées dans le Chapitre suivant ; durant les 40 heures totales d'utilisation du logiciel par des volontaires, aucune erreur ni crash n'est apparu et la lecture est restée stable et fluide.

# Chapitre V - Évaluations de la restitution orale de la mise en forme matérielle

**Résumé :** Ce chapitre présente les résultats de 3 expérimentations portant sur la restitution de la sémantique de la mise en forme matérielle de textes à l'oral. L'objectif général de ces 3 études empiriques était d'évaluer l'impact de la restitution des architectures textuelles sur la compréhension d'un texte expositif oralisé par synthèse vocale ; cette approche a été motivée par de nombreux travaux en psycholinguistique et psychologie cognitive des apprentissages montrant l'importance des signaux visuels (i.e. de la mise en forme) sur la compréhension. La première expérimentation étudie la restitution des titres pendant la lecture, et a été réalisée auprès de 9 sujets déficients visuels. Les résultats montrent que la restitution à l'oral des seuls indices de structure liés aux titres n'a d'effet ni sur la charge cognitive associée à la compréhension du texte, ni sur la compréhension. La deuxième et la troisième expérimentation visaient à évaluer l'impact de 4 stratégies de restitution des architectures textuelles d'un document, testées respectivement auprès de 60 participants non-déficients et 12 participants non-voyants. Les résultats montrent notamment que pour les non-déficients comme pour les déficients visuels, une stratégie est significativement plus efficace que les autres, et ce à plusieurs niveaux. La comparaison des 4 stratégies et des deux expérimentations a permis de définir les cas dans lesquels la restitution à l'oral de la MFM est bénéfique, et quels sont les paramètres à prendre en compte. Plus généralement, ces résultats valident nos hypothèses et l'intérêt de placer la MFM au centre de la problématique plus générale d'accès non-visuel aux documents, cela ayant plusieurs implications pour la conception des technologies d'assistance et pour la recherche.

## Sommaire du Chapitre V

<b>1. Introduction</b> .....	<b>168</b>
<b>2. Concepts clés</b> .....	<b>170</b>
<b>3. Évaluation de la mise en saillance des titres à l'oral auprès de sujets non-voyants</b> .....	<b>172</b>
3.1 Hypothèse éprouvée .....	172
3.2 Méthode.....	172
3.3 Résultats .....	174
3.4 Discussion .....	181
<b>4. Évaluation de 4 stratégies de restitution de la MFM auprès de sujets voyants</b> .....	<b>182</b>
4.1 Présentation des hypothèses .....	183
4.2 Méthode.....	185
4.3 Résultats attendus.....	193
4.4 Résultats .....	194
4.5 Discussion .....	219
<b>5. Évaluation de 2 stratégies de restitution de la MFM auprès sujets non-voyants</b> .....	<b>224</b>
5.1 Présentation .....	224
5.2 Méthode.....	225
5.3 Résultats .....	226
5.4 Discussion .....	247
<b>6. Conclusion sur l'ensemble des résultats</b> .....	<b>250</b>



## 1. Introduction

---

Dans les chapitres précédents nous avons défini ce qu'est la mise en forme d'un texte, montré qu'elle était porteuse de sens, et recensé un certain nombre de travaux ayant révélé l'impact de la signalisation visuelle de la structure d'un texte sur la compréhension. Ces éléments conduisent à penser que lors de l'oralisation automatique d'un document fortement structuré sur le plan visuel, il sera bénéfique de rendre compte de ces informations pour l'apprenant déficient visuel. Nous avons donc proposé une méthodologie de formalisation des architectures textuelles véhiculées par la mise en forme, et montré comment en concevoir une restitution orale pertinente. Enfin, nous avons développé un outil permettant une lecture<sup>107</sup> par synthèse vocale « augmentée » des documents, pour évaluer différentes stratégies de restitution de la mise en forme matérielle (MFM).

L'étude de la mise en forme des documents, aussi appelée « signaux visuels » en psycholinguistique, a mis en évidence son importance pour le lecteur, principalement en termes de mémorisation et de compréhension des textes expositifs<sup>108</sup> [Lorch 1989; Wright 2000; Lemarié et al. 2008] (voir Chapitre I pour plus de détails). Notre approche générale se base sur ces résultats : nous souhaitons évaluer l'impact de la restitution la MFM lors d'un accès au texte par synthèse vocale.

Le choix de la synthèse vocale s'est imposé pour concevoir une restitution de la MFM dans le cadre d'un accès non-visuel aux documents, et ce pour plusieurs raisons. La première est qu'au-delà de l'accessibilité des documents pour les déficients visuels, le choix de la modalité orale permet d'étendre ces travaux à l'accès aux documents par des utilisateurs non-déficients dans des cas où la modalité visuelle n'est pas disponible ; cela rejoint un constat plus général qui est que les travaux et avancées dans le domaine de l'accessibilité ont un impact bien au-delà des populations en situation de handicap [Meiselwitz et al. 2009] . Ensuite, l'oral a été retenu plutôt que l'autre modalité principale d'accès à l'information pour les déficients visuels qui est le braille pour des raisons de littéracie, d'apprentissage et de vitesse d'accès à l'information. En effet, on estimait en 2009 que moins de 10% des déficients visuels considérés *legally blind*<sup>109</sup> lisaient le braille [National Federation of the Blind 2009] ; cela aurait également été un facteur limitant pour la recherche de volontaires déficients visuels. Ensuite, l'apprentissage du braille est un processus long et complexe qui nécessite une assiduité importante [Lusk and Corn 2006], tandis que l'oral suit un apprentissage naturel dans le développement humain. À noter également que la vitesse d'un bon lecteur braille [Hill 1999] est bien inférieure aux vitesses d'énonciations des synthèses vocales utilisées par les déficients visuels [Asakawa and Takagi 2003]<sup>110</sup>. De plus, [Zhao et al. 2008] montrent dans leur étude que l'oral est la modalité la plus utilisées par les déficients visuels ; ce résultat est cependant à nuancer : par exemple les élèves déficients visuels suivant une scolarité en inclusion semblent préférer le braille pour

---

<sup>107</sup> Par abus de langage et pour des raisons de facilité d'écriture, nous désignerons dans ce chapitre l'utilisateur accédant au document lu par synthèse vocale par le terme « *lecteur* », et l'*activité correspondante* la « *lecture* ». Dès lors, toute référence à la lecture visuelle sera faite explicitement pour éviter toute ambiguïté.

<sup>108</sup> Texte présentant et analysant des représentations conceptuelles. Le discours expositif vise à délivrer à ses destinataires des informations dans un domaine de connaissances donné, sans influencer leur jugement.

<sup>109</sup> Ce qui correspond à une acuité visuelle inférieure à 20/200<sup>ème</sup> pour le meilleur œil, voir : <http://www.lighthouse.org/about-low-vision-blindness/definition-legal-blindness/>

<sup>110</sup> Estimée à 100/120 mots par minute maximum pour un bon brailleux [Legge et al. 1999] et au minimum à 180 mots par minute pour les synthèses vocales, qui est la vitesse par défaut, pouvant aller jusqu'à 500 mots par minute.

l'apprentissage [Rao 2006] et l'oral pour la lecture d'autres contenus, pour les loisirs par exemple. Enfin, nous souhaitons écarter pour le moment la complexité supplémentaire qu'aurait amenée la multimodalité (braille + synthèse vocale par exemple) en termes d'évaluation.

La première expérimentation menée s'est concentrée sur la restitution des titres, qui sont des objets jouant un rôle prépondérant dans la compréhension des textes [Wilhite 1989; Lemarié et al. 2012] ; en effet ils sont nombreux dans les textes expositifs et permettent au lecteur de repérer et d'organiser les thèmes, activité essentielle à la compréhension [Lemarié et al. 2008]. Leur restitution à l'oral a d'ailleurs déjà été évaluée auprès de sujets non-déficients [Lorch et al. 2012], en montrant l'importance de leur signalement pour la compréhension. Ces auteurs ont notamment montré qu'effectuer une pause avant l'énonciation d'un titre et signaler son niveau hiérarchique permettent aux auditeurs de mieux retenir les thèmes du texte. On retrouve également dans la littérature plusieurs études portant sur la restitution orale d'objets particuliers comme les énumérations [Maurel et al. 2002; Giraud et al. 2010], utilisant la prosodie des voix de synthèse ou encore l'ajout de métatexte pour signaler ces différents objets et leurs propriétés ; ces études ayant notamment montré l'utilité des indices non-verbaux pour le signalement des objets. On peut également citer [Klatzky et al. 2006] qui a montré le bénéfice induit par l'utilisation d'indices non-verbaux en situation de charge cognitive. Le cas d'une tâche de compréhension d'un document oralisé est une situation génératrice de charge cognitive, à cause des contraintes relatives à la modalité (voir Chapitre II.3.2) qui entraînent plus de charge cognitive qu'à l'écrit, et à cause de la difficulté intrinsèque à la compréhension du contenu. L'objectif pour cette expérimentation était non seulement de valider les résultats de [Lorch et al. 2012] auprès de sujets déficients visuels, mais aussi d'évaluer différentes modalités de signalement non-verbales des titres dans le discours oral (à priori utiles au vu de la charge cognitive induite dans ce contexte), afin de les rendre plus saillants relativement au reste du contenu.

Notre objectif pour les deux autres expérimentations présentées dans ce chapitre était d'évaluer plusieurs stratégies de restitution<sup>111</sup> de la MFM, en nous appuyant sur les résultats de la littérature en ce qui concerne les objets cités ci-avant. En partant d'une stratégie de lecture ne restituant aucune information ni fonction liée à la MFM (condition *baseline*), nous avons défini 3 autres stratégies de restitution allant jusqu'à une restitution dite « optimale », c'est-à-dire essayant, sur la base des informations de mise en forme annotées, de restituer de manière optimale uniquement les informations pertinentes pour la réalisation de la tâche, qui consistait ici pour les sujets à comprendre le contenu et la structure du document. Une de ces deux expérimentations évaluait les stratégies définies auprès de sujets voyants, et l'autre auprès de sujets non-voyants. Même si ces travaux présentent des applications directes pour les déficients visuels, il nous semblait pertinent d'évaluer la restitution de la mise en forme sur une population de sujets non-déficients afin d'explorer le potentiel applicatif des résultats au-delà du cas de la déficience (ex. accès oral à l'information sur de petits écrans ou en double-tâche), mais aussi afin de pouvoir comparer les comportements des sujets déficients à ceux des sujets non-déficients pour une même tâche. Enfin, cela permettait de recueillir une plus grande quantité de données au vu des difficultés d'accès à une population réduite de volontaires déficients visuels.

---

<sup>111</sup> Par « stratégie de restitution » on entend une réflexion globale sur quelles informations restituer, quand, et comment (par quelles modalités). Plusieurs stratégies ont été définies pour les expérimentations 2 et 3 (voir Chapitre V4.2.c) sur la base de l'étude présentée dans le Chapitre IV.

Avec cette approche, nous souhaitions tester notre hypothèse globale selon laquelle la restitution de la MFM peut améliorer la compréhension. Étant donné qu'une tâche d'apprentissage comme la compréhension d'un texte implique la mise en place de nombreux processus et activités différentes par les apprenants, nous avons mesuré le plus de paramètres pertinents possible pour ces deux expérimentations, les principaux ayant été : la compréhension, la rétention du plan, la charge cognitive subjective associée à la tâche de compréhension ainsi que la désorientation perçue (le sentiment d'être perdu dans un document). Nous souhaitions ainsi mesurer à la fois la performance, qui dans notre cas peut être représentée par la compréhension et la rétention du plan, mais aussi la charge cognitive qui représente les ressources mentales engagées dans la tâche, et le sentiment de désorientation mesurant (par auto-évaluation) la capacité des sujets à pouvoir s'orienter, se repérer dans le document. L'ensemble de ces mesures étaient nécessaires, non seulement en raison de la complexité de la tâche de compréhension, mais aussi afin d'évaluer l'accessibilité des stratégies proposées en termes d'efficacité cognitive et de satisfaction des utilisateurs. Ce dernier argument nous paraît être d'importance, au vu des frustrations et de la fatigue mentale auxquelles peuvent être sujets les déficients visuels lors de l'accès aux contenus numériques, et à l'information de manière générale.

L'objectif final, à travers ce que l'on appelle ici la « restitution de la MFM », est à la fois de rendre disponible les informations relatives aux architectures textuelles, qui autrement ne le seraient pas, mais aussi d'améliorer l'accessibilité de ces documents en resituant ces informations de la manière la plus pertinente possible au regard de l'objectif de la tâche réalisée avec le texte, qui est ici une tâche de compréhension. Les résultats de ces travaux ont donc une portée immédiate sur la conception des systèmes d'accès non-visuel aux documents numériques, notamment dans le cadre de l'apprentissage.

## 2. Concepts clés

---

Avant de développer les différentes expérimentations réalisées et les résultats obtenus, il est nécessaire de présenter les concepts importants repris à la fois dans la conception des expérimentations et dans les résultats : la compréhension, la charge cognitive, la désorientation et l'utilisabilité.

Tout d'abord notre hypothèse principale porte sur l'impact de la restitution de la mise sur la **compréhension** d'un texte expositif (visant à informer le lecteur sur certains concepts) ; nous avons déjà vu d'après la littérature en psycholinguistique que cette hypothèse était vérifiée en lecture visuelle. La compréhension d'un tel texte se fait chez le lecteur à travers la construction d'une représentation globale cohérente des thèmes traités par le texte ; cette représentation étant logiquement influencée par le plan du texte [Lemarié et al. 2012], c'est-à-dire l'ensemble de ses titres et la hiérarchie de thèmes qu'ils définissent, nous avons mesuré dans toutes les expérimentations la rétention du plan par les sujets, c'est-à-dire ce dont ils se souvenaient de l'ensemble des titres.

Pour les trois expérimentations présentées, nous avons mesuré la **charge cognitive** subjective des sujets. La théorie de la charge cognitive [Sweller 1994] est basée sur le fait que l'humain a des ressources de traitement cognitives limitées (en mémoire de travail notamment). Mesurer la charge cognitive revient donc à mesurer la contrainte pesant sur le système cognitif pour une tâche donnée, les ressources engagées. Les mesures subjectives de la charge cognitive à travers des

questions d'auto-évaluation par exemple, sont faciles et économiques à administrer et à la fois robustes et sensibles à de faibles variations [Paas and Van Merriënboer 1994]. Si la charge cognitive représente globalement l'ensemble des ressources cognitives engagées pour la réalisation d'une tâche, la théorie distingue différentes composantes de la charge cognitive globale, qui ont été mesurées dans les deux dernières expérimentations. On distingue notamment la charge intrinsèque qui varie en fonction du contenu et de l'expertise du lecteur relative à ce contenu ; le terme intrinsèque renvoie au fait que cette composante dépend uniquement de la difficulté du contenu pour le lecteur, et pas d'autres facteurs externes. Une autre composante de la charge cognitive est la charge extrinsèque qui est liée à la manière dont l'information est présentée (ici en utilisant la modalité orale couplée à divers indices non-verbaux), et dépend donc de la consigne donnée aux lecteurs et de la manière dont le contenu document est présenté. De manière générale, on peut supposer que sauf cas particuliers, la restitution de la MFM va principalement influencer sur la charge extrinsèque car elle est directement liée au format de présentation de l'information. Enfin, une dernière composante représente les ressources utilisées pour la mise en place des processus d'apprentissage (par exemple, les ressources mentales utilisées pour extraire la structure thématique d'un texte). Ces trois composantes ont notamment été détaillées par [Paas et al. 2003]. Le cumul de ces composantes représente la charge cognitive globale, c'est-à-dire l'ensemble des ressources cognitives mobilisées pour la réalisation de la tâche.

**L'utilisabilité** est un concept très général utilisé au-delà du domaine des interfaces informatiques, englobant la facilité d'utilisation d'un système, sa capacité à être appréhendé rapidement, son efficacité, et bien d'autres. Il est défini par [Brooke 1996] comme la pertinence, l'adéquation avec laquelle un système (au sens large) peut permettre à son utilisateur d'atteindre un certain but dans un contexte donné. Dans les deux dernières expérimentations nous avons choisi de mesurer l'utilisabilité du système de lecture utilisé par les participants à l'aide d'un questionnaire standard (largement utilisé dans le domaine des interfaces homme-machine), afin d'évaluer le ressenti utilisateur vis-à-vis du système.

La **désorientation** est un concept qui a vu son émergence avec la navigation hypertexte [Kim and Hirtle 1995; McDonald and Stevenson 1996] et désigne pour un utilisateur le sentiment d'être perdu en termes de navigation (ne pas savoir où l'on se trouve, comment aller où l'on souhaite, ne pas savoir où aller). Cette nouvelle mesure pourrait permettre des analyses complémentaires relatives à un système de lecture pour les déficients visuels : si, pour les voyants la désorientation est liée principalement à la navigation hypertexte qui est définie comme une navigation non-linéaire à travers des liens entre pages (et donc d'une nature exploratoire [McDonald and Stevenson 1996]), la navigation au sein d'un même document pour un déficient visuel peut comporter de nombreuses phases de navigation similaires en plus de la navigation hypertexte. En effet l'absence d'une perception globale directe de l'espace de navigation fait que la représentation de cet espace se construit de manière *exploratoire*, au fur et à mesure de la lecture ; cela est également illustré par le fait que les principales fonctions de lecture d'un document des lecteurs d'écrans soient non linéaire (ex. pour une page web : accès par la liste des titres, accès par liste des liens, accès par champs de formulaires). La désorientation peut être mesurée en termes de ressenti subjectif, mais aussi par des mesures objectives, ici définies pour la navigation Web (domaine dans lequel a émergé le concept de désorientation), qui vont relater du temps total de navigation, du nombre de retours en arrière et de boucles effectuées entre les pages ou encore du nombre total de liens visités [Lin 2003; Giraud 2014].

Enfin la littérature montre que même si désorientation et utilisabilité sont corrélées, les deux sont complémentaires pour tirer des conclusions sur la conception du système utilisé [Ahuja and Webster 2001]. De plus, la même étude montre que la désorientation n'est à priori pas associée à des comportements de navigation particuliers (la désorientation ne permet pas de prédire les comportements de navigation) ; cependant [McDonald and Stevenson 1996] ont montré que la performance dans des tâches de navigation (recherche d'information, d'emplacement) est corrélée avec la désorientation. Ces deux derniers points montrent qu'il est important de mesurer la désorientation en plus des performances de navigation, mais qu'une simple observation des comportements ne suffit pas.

### **3. Évaluation de la mise en saillance des titres à l'oral auprès de sujets non-voyants**

---

Cette partie présente le protocole et les résultats d'une expérimentation portant sur l'évaluation de 3 stratégies de restitution des titres pendant une lecture par synthèse vocale. Les résultats de cette étude ont été publiés dans une conférence internationale [Sorin et al. 2014].

#### **3.1. Hypothèse éprouvée**

---

Les titres font partie des objets textuels les plus importants pour la compréhension d'un texte, et ont été largement étudiés dans le cadre de la lecture visuelle [Wilhite 1989; Lemarié et al. 2012]. En effet comprendre un texte consiste à en construire une représentation cohérente (voir Chapitre I), ce qui implique d'identifier les thèmes du texte et leur organisation, justement véhiculés par les titres dans un texte expositif. Les lecteurs d'écran signalent les titres verbalement par un énoncé métatextuel indiquant le niveau hiérarchique de ce titre ; l'importance de ce signalement pour la rétention des thèmes lors d'une lecture par synthèse vocale a été démontrée [Lorch et al. 2012]. Or, des travaux ont montré que les indices non-verbaux peuvent être plus efficaces que les indices verbaux pour restituer de l'information à l'oral [Giraud et al. 2010; Gellenbeck and Stefik 2009], et particulièrement en situation de charge cognitive élevée [Klatzky et al. 2006]. La compréhension d'un texte étant une activité complexe, ces derniers semblent être appropriés pour signaler la présence de titres dans ce cadre. De plus, utiliser des indices non-verbaux (comme des icônes audio) pourrait permettre une meilleure identification des titres grâce au changement de modalité marquant l'importance de l'objet énoncé.

L'hypothèse principale de cette expérimentation est donc que signaler les titres d'un texte par des indices non-verbaux permettrait d'améliorer l'identification, la rétention des titres et donc la compréhension générale de la structure du document.

#### **3.2. Méthode**

---

##### **3.2.a. Méthodes de restitution des titres**

---

Les titres du document pouvaient être restitués de 3 façons différentes en fonction de la condition. Le reste du document était lu normalement par la voix de synthèse (qui était une voix féminine) :

- § Condition de contrôle (CTRL) : Les titres étaient lus normalement par la voix de synthèse. Avant chaque titre était énoncé le niveau du titre de la manière suivante : « Titre de niveau N ». Cette condition reproduit la façon dont les titres sont signalés par les lecteurs d'écran.
- § Condition avec une seconde voix (2V) : Les titres étaient lus de la même manière que dans la condition de contrôle mais avec une voix masculine.
- § Condition avec son spatialisé (3DS) : Les titres étaient lus de la manière que dans la condition de contrôle, excepté qu'ils l'étaient depuis une source sonore localisée (simulée comme se trouvant) sur la gauche des participants. Le reste du texte était lu depuis une source sonore se situant sur la droite des participants.

Hypothèses opérationnelle : Les conditions 2V et 3DS devraient induire une meilleure rétention des thèmes du plan et de la compréhension, en permettant d'identifier plus facilement les titres du texte. Leur signalement par des modalités non-verbales (changement de voix et localisation du son) devrait être plus efficace en termes de charge cognitive pour une tâche de compréhension.

#### ***3.2.b. Participants***

---

Les 9 sujets ayant pris part à cette expérimentation étaient âgés en moyenne de 43 ans ( $SD = 15$ ), avec une étendue importante (17-64 ans). Cette population était composée de 5 sujets de sexe féminin et 4 de sexe masculin.

Tous les sujets présentaient une cécité totale ; 5 sur les 9 étaient aveugles de naissance. Tous les sujets utilisaient quotidiennement un ordinateur ou un bloc-notes braille, avec des voix de synthèse.

#### ***3.2.c. Matériel, tâches et consignes***

---

Les sujets disposaient d'un clavier d'ordinateur qu'ils pouvaient utiliser pour mettre en pause la lecture. Le contenu des documents était lu par synthèse vocale à l'aide d'une voix de synthèse issue du projet MBROLA<sup>112</sup>.

Trois documents tests étaient utilisés, un par condition expérimentale. Les thèmes des 3 textes expositifs utilisés étaient respectivement : les problèmes énergétiques, les incendies et les solutions énergétiques (énergies renouvelables). Ces thèmes ont été choisis afin que chaque participant puisse comprendre le contenu sans être expert dans le domaine, en d'autres termes que le contenu présente une difficulté moyenne pour les participants. Le contenu des documents provenait initialement d'un corpus expérimental utilisé précédemment en psychologie cognitive des apprentissages. Tous les documents présentaient tous une structure similaire (même profondeur de la hiérarchie définie par les titres, même nombre de parties et sous parties) et une taille identique à 100 mots près. La lecture de chaque document durait environ 5 minutes. Deux autres documents d'une durée de lecture de moins de 2 minutes étaient utilisés comme « distracteurs » (voir section suivante).

Les tests débutaient systématiquement par la signature d'un formulaire de consentement par les sujets. Les étapes suivantes étaient ensuite répétées 3 fois durant les tests :

---

<sup>112</sup> Projet visant à développer des voix de synthèse utilisables gratuitement à des fins non-commerciales, voir : <http://tcts.fpms.ac.be/synthesis/mbrola.html>

Étape	Consigne	Description
1. <i>Écoute passive d'un des 3 documents avec une des 3 méthodes de restitution des titres</i>	« Votre objectif pendant l'écoute de ce document sera d'essayer de comprendre au mieux le contenu du document ainsi que sa structure ».	Les sujets ne pouvaient pas naviguer dans le document et devaient écouter l'intégralité du contenu ; ils pouvaient cependant mettre la lecture en pause. En effet, nous souhaitons que les sujets n'accèdent qu'une seule fois aux informations du texte afin de pouvoir plus aisément comparer les mesures. À noter que l'ordre de présentation des conditions et des documents était contrebalancé sur les 9 passations.
2. <i>Restitution du plan du document</i>	« Redonnez le plan du texte en essayant d'être le plus précis et complet possible ; si vous vous ne rappelez pas des titres exact, vous pouvez rappeler les thèmes abordés successivement dans le document ».	Rappeler le plan, c'est-à-dire un équivalent de la table des matières : les titres, leurs niveaux hiérarchiques, leur imbrication
3. <i>Évaluation subjective de la charge cognitive</i>	« Sur une échelle de 1 à 7, pouvez-vous évaluer l'effort mental que vous a demandé la tâche que vous venez de réaliser ? »	La théorie de la charge cognitive stipule que nos ressources mentales permettant de traiter l'information sont limitées, et que cette quantité limitée de ressources peut être plus ou moins sollicitée par la tâche demandée. L'auto-évaluation de la charge cognitive est une méthode économique et ne perturbe pas la tâche principale comparativement aux autres méthodes existantes.
4. <i>Questions de compréhension</i>	Voir 4.6.	5 questions de compréhension étaient posées aux sujets. Les questions portaient sur l'ensemble du document et évaluaient la rétention et la mise en relation des concepts important.
5. <i>Écoute passive d'un document distracteur</i>	« Vous allez maintenant écouter un court texte traitant d'un sujet complètement différent. Vous devez l'écouter avec attention ».	(cette étape n'était répétée que 2 fois) Les sujets devaient écouter un document durant moins de 2 minutes traitant soit de données sur le Brésil soit d'espèces en voie de disparition.
6. <i>Questions sur le document distracteur</i>	Voir 4.6.	(cette étape n'était répétée que 2 fois). Deux questions était posées portant sur des points de détail du document distracteur.

**Tableau 19 – Déroulement de l'expérimentation sur les titres**

Enfin on demandait aux sujets quelle restitution ils avaient préféré, et pourquoi.

Les 3 conditions et les 3 textes utilisés étaient contrebalancés afin de contrôler l'ordre de présentation des conditions, de sorte que chaque condition soit passée autant de fois en 1<sup>re</sup>, 2<sup>ème</sup> et 3<sup>ème</sup> positions, et afin de contrôler la répartition des textes sur les conditions de sorte que chaque condition soit passée 3 fois avec chacun des 3 textes.

### 3.3. Résultats

Cette partie présente les résultats de l'évaluation des 3 méthodes de restitutions des titres auprès de 9 sujets non-voyants. Les différents indicateurs seront chacun présentés ci-après dans une section dédiée ; les données correspondant aux 3 conditions expérimentales y seront à chaque fois comparées ; le cas échéant, on présentera les éventuels biais de l'ordre de présentation et du document sur la mesure. La présentation des résultats dans chacune de ces sections se fera systématiquement de la manière suivante :

- § Caractéristiques de l'indicateur : calcul, étendue (valeurs prises par l'indicateur) et mesure (ce qu'il représente)
- § Normalité des données : on utilisera un test de Shapiro-Wilk pour tester la normalité des échantillons, dont les résultats seront présentés de la manière suivante :  $W = W\_value$ ,  $p >$  ou  $<$  *seuil de significativité* ; une valeur de P supérieure au seuil critique indique qu'on accepte l'hypothèse nulle que les données viennent d'une population normalement distribuée. Afin d'alléger la lecture des résultats, nous présenterons les résultats de ce test uniquement lorsque les données sont normales. Ainsi, si l'inverse n'est pas clairement énoncé, les données présentées seront non-normales. *À noter que nous avons testé la normalité de chaque indicateur, y compris les réponses des sujets qui se faisaient sur des échelles de Likert, car même si ces données peuvent être considérées comme ordinales, les items des échelles de Likert sont totalement ordonnés : ces échelles peuvent donc être vues comme continues.*
- § Présentation graphique des échantillons : effet de la condition sur l'indicateur
- Si les données ne suivent pas une loi normale, on présentera les médianes et quartiles des échantillons, qui sont dans ce cas de meilleurs descripteurs des distributions que la moyenne. *À noter que sur ces graphiques (dits « boîtes à moustaches »), les médianes sont représentées par un trait noir horizontal, et les premiers et troisièmes quartiles par les bornes inférieures et supérieures du rectangle. Un point rouge a été rajouté sur ces graphiques pour chaque échantillon, représentant leurs moyennes*
  - Si les données suivent une loi normale, on présentera sur les graphiques les moyennes et intervalles de confiance à 95% de ces moyennes
- § Commentaires sur les graphiques ; seront commentées notamment les moyennes ( $M$ ), les déviations standard ( $SD$ ), les médianes ( $ME$ ), les quartiles ( $Q_N$  où  $N=1$  ou  $N=3$  pour respectivement le 1<sup>er</sup> et le 3<sup>ème</sup> quartile), la valeur la plus basse des échantillons ( $MIN$ ) et la valeur maximale ( $MAX$ )
- § Tests statistiques et résultats de ces tests. Dans notre cas nous comparerons donc 3 groupes appariés pour chaque mesure, correspondants aux 3 conditions expérimentales. Nous vérifierons également pour chaque mesure la présence éventuelle d'effet de l'ordre de présentation des textes ou des textes eux-mêmes. *Ces tests ont été réalisés à l'aide du logiciel R ; les fonctions utilisées correspondant aux tests statistiques sont présentées en 4.3*
- Données normales : comparaison des moyennes avec un test d'analyse de la variance (ANOVA) à 1 facteur (ici la condition expérimentale) avec mesures répétées. Les résultats des ANOVA seront présentés de la manière suivante :  $F(\text{ddl}_{\text{inter-groupes}}, \text{ddl}_{\text{intra-groupes}}) = F\_value$ ,  $p <$  ou  $>$  au seuil de significativité (ddl désignent les degrés de liberté). Étant donné que la  $F\_value$  représente un ratio entre les moyennes, plus elle se rapproche de 1.0, plus il y a de chances d'accepter l'hypothèse nulle d'absence de différence entre les moyennes.
  - Données non-normales : comparaison des rangs des observations avec un test de Friedman pour échantillons appariés. Les résultats des tests seront présentés de la manière suivante :  $\chi^2(\text{ddl}) = \text{Chi\_value}$ ,  $p <$  ou  $>$  au seuil de significativité (ddl désignent les degrés de liberté). La  $\text{Chi\_value}$  représente ici l'amplitude de l'écart entre les différents groupes ; quand cette valeur dépasse un seuil critique on rejette l'hypothèse nulle. Ici, pour un seuil de significativité fixé à 5% en considérant que le test se fera ici avec 2 degrés de libertés (nombre de conditions moins 1), la valeur critique est de 5.99, reportée depuis une distribution standard du  $\chi^2$  de la densité de probabilité, pour  $y = 0.05$  et  $\text{ddl} = 2$ .



On utilisera les codes suivants pour désigner les 3 conditions expérimentales : CTRL = condition de contrôle ; 3DS = son spatialisé utilisé pour les titres ; 2V = seconde voix utilisée pour les titres.

### ***3.3.a. Effet de la stratégie de restitution sur la rétention du plan***

---

À la suite de l'écoute du document, il était demandé aux sujets de restituer le plan, c'est-à-dire l'ensemble des titres ou thèmes du texte, leur imbrication et leur ordonnancement. Cette restitution nous a permis de mesurer 3 indicateurs, déjà utilisés pour mesurer la rétention du plan [Lorch et al. 2012], dont les résultats sont reportés dans les sections ci-après :

1. Les thèmes des titres restitués relativement aux thèmes définis par les titres du document d'origine
2. La distance au plan, c'est-à-dire la différence de niveaux hiérarchiques entre les titres restitués et les titres correspondants du document
3. L'ordre des titres restitués, c'est-à-dire la différence entre l'ordonnancement des titres restitués et l'ordonnancement des titres du document

#### **3.3.a.1. Identification et rétention des thèmes du plan**

---

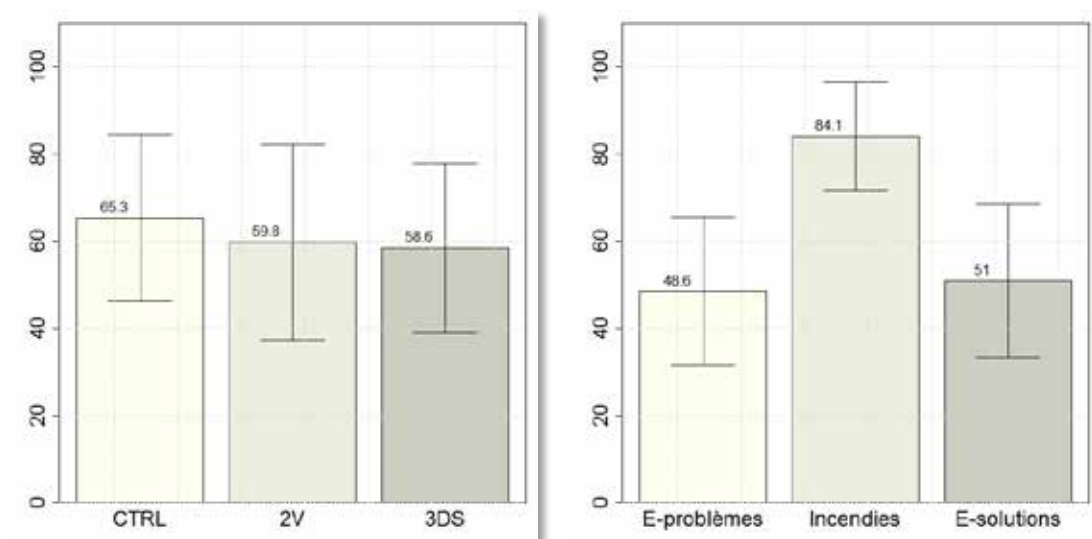
Caractéristiques de l'indicateur :

- § *Calcul* : correspondance entre les thèmes des titres rappelés par le sujet et les thèmes des titres du document. Chaque thème du plan restitué par les sujets était noté : 0 si le thème n'était pas présent dans le plan du document, 1 s'il était présent (sous une forme littérale ou équivalente sémantiquement), 0.5 si le thème a été mal restitué ou ne correspond pas totalement à un thème du plan. Le score de chaque sujet était ensuite ramené sur 100.
- § *Étendue* : [0,100], intervalle continu
- § *Mesure* : *identification et* rétention des thèmes ; le fait que le sujet rappelle le contenu d'un titre dans le plan rappelé indique qu'il a reconnu le contenu oralisé comme étant un titre durant l'écoute et qu'il a mémorisé correctement son contenu.

Normalité des données : Cet échantillon suit bien une loi normale (Shapiro-Wilk :  $W = 0.95$ ,  $p > 0.05$ ). On comparera donc les moyennes de cet échantillon avec un test d'analyse de la variance.

Tests statistiques : La stratégie de restitution des titres n'a pas eu d'effet significatif sur cet indicateur ( $F(2,24) = 0.16$ ,  $p > 0.05$ ). On notera également l'absence d'effet de biais de l'ordre de présentation ( $F(1,25) = 0.2$ ,  $p > 0.05$ ) sur cet indicateur. En revanche, on observe un effet significatif du texte présenté ( $F(2,24) = 7.99$ ,  $p < 0.01$ ). Le test post-hoc du *TukeyHSD* nous confirme que la moyenne du deuxième document est significativement plus importante que celle des deux autres documents. On en conclut que de par sa structure, le deuxième document a suscité un rappel des thèmes du plan meilleur que les 2 autres documents.

**Figure 40 - Rétention des thèmes du plan en fonction de la stratégie et du document : expérimentation sur la restitution des titres**



On remarque que quelle que soit la condition de restitution, le rappel des thèmes du plan a donné des résultats au-dessus de la moyenne (les sujets ont rappelé pour chaque condition environ 6 thèmes sur 10 présentés), avec cependant une variabilité très importante.

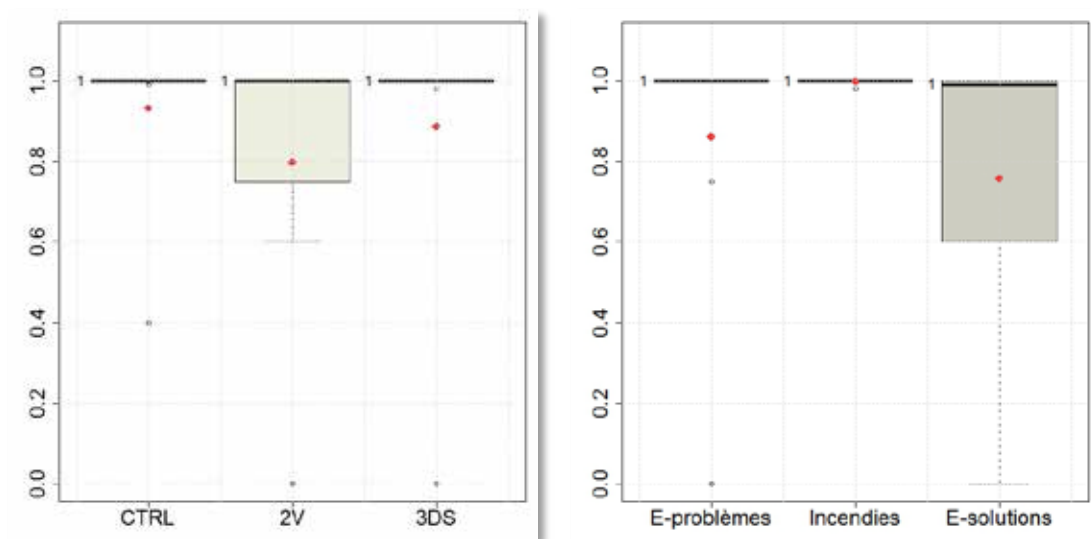
#### 3.3.a.2. Ordre des thèmes dans le plan rappelé

##### Caractéristiques de l'indicateur :

- § *Calcul* : coefficient de corrélation entre l'ordre des titres restitués et l'ordre d'apparition de ces mêmes titres dans le document, indépendamment de la quantité de titres rappelés. Un score de 1 signifie que les titres restitués l'ont été dans le bon ordre.
- § *Étendue* : [0;1], intervalle continu.
- § *Mesure* : cet indicateur mesure la représentation qu'à le sujet de l'ordre des thèmes dans le texte indépendamment de leur niveau hiérarchique et sa proximité à l'ordre dans le texte écouté.

Tests statistiques : La stratégie de restitution des titres n'a pas eu d'effet significatif sur cet indicateur ( $\chi^2(2) = 1.5, p > 0.05$ ). On notera également l'absence d'effet de biais de l'ordre de présentation ( $\chi^2(2) = 2.5, p > 0.05$ ). En revanche, on observe de nouveau un effet du texte ( $\chi^2(2) = 7.4, p < 0.05$ ). On remarque que pour les deux premiers textes, les scores plafonnent à un score maximal. Pour le dernier document, les scores présentent une variabilité importante, l'écart interquartile couvrant presque la moitié de l'étendue de l'indicateur. Un test de comparaisons multiple nous confirme que le troisième document, de par sa structure, a suscité un rappel de l'ordre des titres du plan moins bon que les 2 autres documents.

**Figure 41 - Ordre de restitution des titres en fonction de la stratégie et du document : expérimentation sur la restitution des titres**



Les scores d'ordre sont maximums pour les stratégies contrôle et son-3D ; la médiane et le troisième quartile sont à 1 pour la condition utilisant deux voix qui observe cependant une certaine variabilité. Ces distributions ne sont cependant pas considérées comme significativement différentes.

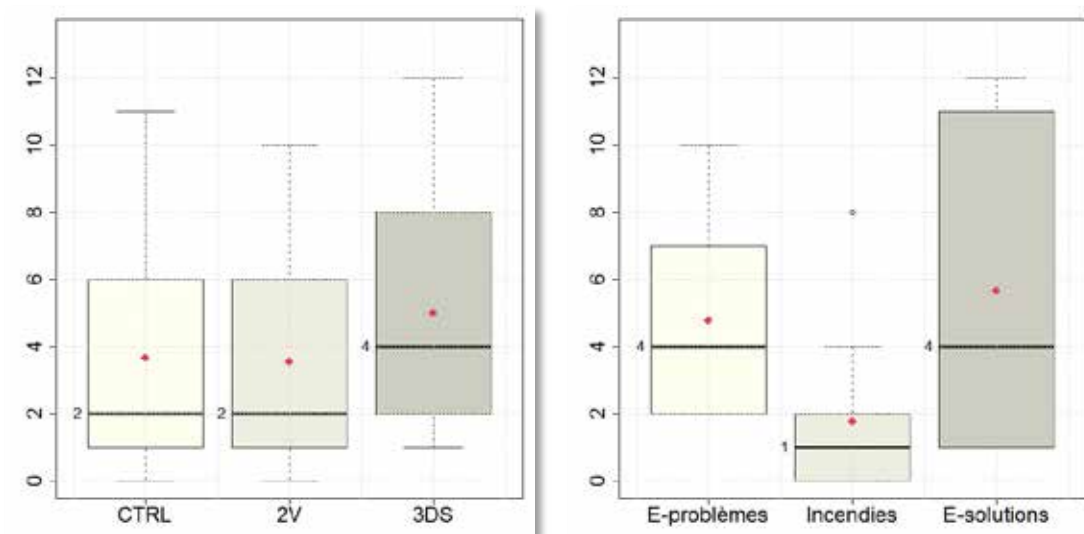
### 3.3.a.3. Distance entre le plan rappelé et le plan d'origine

#### Caractéristiques de l'indicateur :

- § *Calcul*: somme des différences en valeurs absolues entre les niveaux des titres restitués et les niveaux des titres correspondants du document. *Exemple* : si un sujet restitue un titre au niveau 1 alors qu'il se trouvait au niveau 2 (dans la hiérarchie du document), alors on ajoute 1 à son score de distance. Une distance de 0 montre que les titres restitués par les sujets l'ont été avec la bonne hiérarchie.
- § *Étendue* :  $\{0, N\}$ , intervalle d'entiers. Théoriquement il n'y a pas de limite à N d'après la façon dont est calculé l'indicateur, car un sujet est libre de choisir les niveaux des thèmes restitués
- § *Mesure* : cet indicateur mesure la représentation qu'a construite l'apprenant de la hiérarchie des thèmes et son adéquation avec la hiérarchie du plan du texte de départ.

Tests statistiques : La stratégie de restitution des titres n'a pas eu d'effet significatif sur cet indicateur ( $\chi^2(2) = 0.2, p > 0.05$ ). On notera également l'absence d'effet de l'ordre de présentation ( $\chi^2(2) = 0.4, p > 0.05$ ). En revanche, on observe un effet du texte ( $\chi^2(2) = 12.8, p < 0.01$ ). Un test de comparaisons multiples nous confirme que le second document, de par sa structure a suscité un rappel des niveaux des titres du plan plus précis que pour les 2 autres documents. On remarquera une variabilité importante pour les 2 autres documents.

Figure 42 - Distance au plan en fonction de la stratégie et du document : expérimentation sur la restitution des titres



A noter que, quelle que soit la condition, les scores d'ordre sont relativement bas mais très variables. On remarque que les scores pour la condition utilisant du son spatialisé tendent à être plus élevés (ordre rappelé moins bon), les écarts interquartiles étant équivalents entre les stratégies mais les médianes différentes.

### 3.3.b. Effet de la stratégie de restitution sur les scores de compréhension du contenu

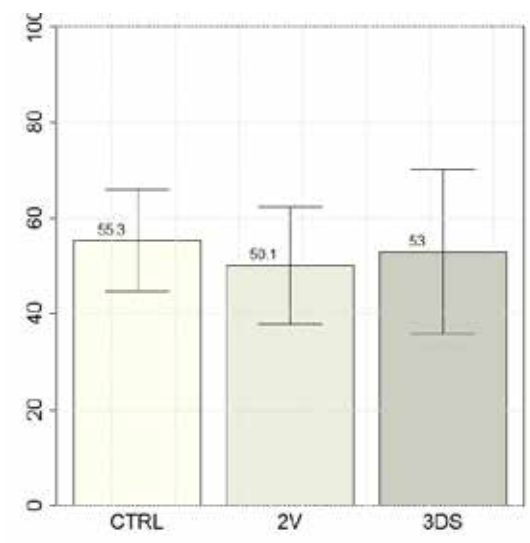
#### Caractéristiques de l'indicateur :

- § *Calcul* : évaluation des réponses des sujets aux 5 questions de compréhension sur chaque document ; le score était ensuite ramené sur 100. Pour chaque réponse, on attendait un certain nombre d'idées ou de concepts issus du document ; pour chaque idée on attribuait une note : 0 (non-sens ou non-restituée), 1 (réponse exact ou équivalent sémantique) ou 0.5 (réponse approximative ou incomplète faisant tout de même sens).
- § *Étendue* : [0,100], intervalle continu.
- § *Mesure* : compréhension du contenu du texte.

Normalité des données : Cet échantillon suit une loi normale (Shapiro-Wilk :  $W = 0.96$ ,  $p > 0.05$ ). On comparera donc les moyennes des échantillons dans les tests statistiques.

Tests statistiques : La stratégie de restitution des titres n'a pas eu d'effet significatif sur cet indicateur ( $F(2,24) = 0.2$ ,  $p > 0.05$ ). On notera également l'absence d'effet de l'ordre de présentation ( $F(1,25) = 0.3$ ,  $p > 0.05$ ) ainsi que du texte utilisé ( $F(2,24) = 0.9$ ,  $p > 0.05$ ).

Figure 43 - Scores de compréhension en fonction de la condition : expérimentation sur la restitution des titres



On remarque que les scores de compréhension sont moyens, quelle que soit la condition (le score maximal atteignable étant de 100), ainsi, l'absence d'effet des stratégies de restitution comparées ne peut être imputée à des effets plancher ou plafond au niveau des performances.

### 3.3.c. Effet de la stratégie de restitution sur les scores de charge cognitive globale

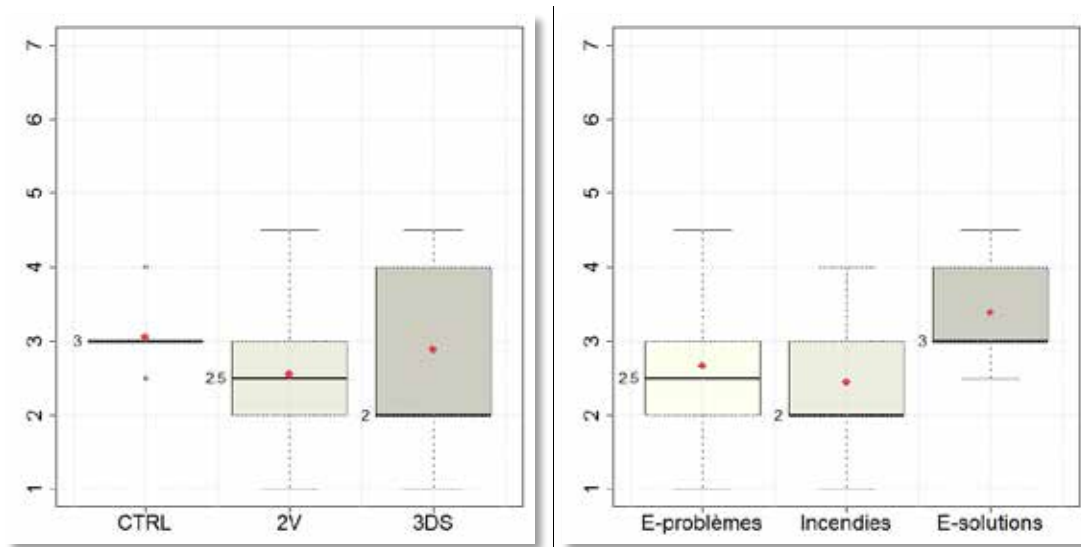
On analyse ici les résultats des réponses à la question : *De manière générale, La compréhension du document et de sa structure m'a demandé un effort mental ? (1 à 7, 1 étant un effort très peu important et 7 un effort très important)*. Cette question était posée immédiatement après l'écoute du texte oralisé.

#### Caractéristiques de l'indicateur :

- § *Calcul* : aucun ; report des réponses des sujets.
- § *Étendue* : {1, 7}, intervalle d'entiers.
- § *Mesure* : l'effort mental associé à la tâche de compréhension.

Normalité des données : Cet échantillon ne suit pas une loi normale (Shapiro-Wilk :  $W = 0.91$ ,  $p < 0.05$ ).

Tests statistiques : La stratégie de restitution des titres n'a pas eu d'effet significatif sur cet indicateur ( $\chi^2(2) = 0.8$ ,  $p > 0.05$ ). On notera également l'absence d'effet l'ordre de présentation ( $\chi^2(2) = 2.4$ ,  $p > 0.05$ ). En revanche, on observe un effet du texte ( $\chi^2(2) = 6.5$ ,  $p < 0.05$ ). Un test de comparaisons multiples montre cependant que le document traitant de solutions énergétiques a induit une charge mentale supérieure aux deux autres documents.

**Figure 44 - Charge cognitive globale en fonction de la stratégie et du document : expérimentation sur la restitution des titres**

On remarquera que quelle que soit la condition, les scores de charge cognitive étaient bas, ce qui traduit un effort mental peu important engagé pour effectuer la tâche demandée.

### 3.4. Discussion

Les stratégies de restitution des titres utilisées n'ont impacté aucun des indicateurs mesurés, que ce soit la compréhension, la rétention du plan ou la charge cognitive. Ces résultats montrent les stratégies avec du son spatialisé et un changement de voix ne diffèrent pas en termes d'efficacité. Plusieurs hypothèses peuvent expliquer l'absence de résultats significatifs dans nos mesures. La première et la principale était que la tâche réalisée n'était pas assez difficile pour observer un effet des indices non-verbaux, dont on sait qu'ils sont surtout efficace dans des situations où il y a une charge mentale importante [Klatzky et al. 2006] ; cela se vérifie à travers les scores observés de charge mentale peu élevés. Cette difficulté peu élevée pourrait provenir des connaissances antérieures des sujets, d'une quantité trop faible de contenu présenté, d'un trop faible nombre de titres et plus généralement d'une trop faible complexité du plan, ou encore de la vitesse d'énonciation de la synthèse vocale trop basse comparativement à ce dont les sujets avaient l'habitude [Asakawa and Takagi 2003]. Ce problème de calibrage de la difficulté pourrait également provenir des compétences développées par les non-voyants en compréhension orale de texte : on peut faire l'hypothèse de capacités supérieures dans ce domaine si l'on s'en réfère à la plasticité cérébrale permettant aux déficients de réallouer certaines zones corticales au traitement de l'audio ou du braille [Hugdahl et al. 2004; Burton et al. 2002]. Ainsi ce qui nous paraissait être une tâche d'une difficulté raisonnable a pu être beaucoup moins exigeante en réalité.

Cependant, on remarque que les scores liés à restitution du plan par les sujets sont au-dessus de la moyenne et les scores de compréhension moyens : il n'y aurait donc pas de plafonnement des performances, ce qui aurait pu sembler cohérent avec une tâche « facile ». En revanche, certaines études [McNamara et al. 1996] montrent que lorsqu'un texte est très cohérent localement et globalement le lecteur (voyant dans ces études) fait preuve d'une plus grande passivité et met en place des traitements plus superficiels ; un effet similaire a pu être mis en jeu dans notre cas. De plus, la littérature sur la charge cognitive montre qu'une charge cognitive trop faible (ou trop

importante) induit une dégradation de l'apprentissage [Teigen 1994; Paas et al. 2003]. Ainsi, au regard de ces deux derniers points et des scores de charge cognitive montrant des scores faibles, l'hypothèse d'un problème de calibrage de la difficulté de la tâche (et des contenus) semble plausible.

Ensuite, nous avons observé dans cette expérimentation d'important biais liés aux documents utilisés sur la rétention du plan ; malgré le contrebalancement des documents sur les conditions, cela a pu influencer l'ensemble des résultats du fait de la petite taille de notre échantillon ( $N = 9$ ). En effet, pour le document traitant des incendies, les sujets ont mieux rappelé les thèmes des titres du plan ainsi que la hiérarchie de ces titres que pour les 2 autres documents, et ce de manière significative. Cela pourrait être lié au fait ce document présentait un plan dont les items étaient fortement liés et déjà classés d'après leurs catégories sémantiques (ex. 1) les véhicules d'intervention, a) le camion de pompier). De plus, on remarque aussi que le troisième document, traitant de solutions énergétiques, observe de scores de restitution de l'ordre des titres et de distance au plan moins bons et plus variables que les 2 autres documents, ce qui pourrait indiquer que sa structure hiérarchique ait été plus complexe. Cela est confirmé par le fait que la charge cognitive était significativement plus élevée pour ce document que pour les 2 autres. Pour conclure sur ce biais, il faut noter que la difficulté dans l'évaluation de la restitution de la MFM est que celle-ci jouant un rôle de facilitateur de traitement de l'information, la mesure finale étant une performance globale, elle est fortement dépendante du contenu d'origine présenté. Les contenus des documents doivent donc être conçus avec plus de soins pour les futures expérimentations.

Enfin, il est possible que les non-voyants soient habitués à identifier les titres sur la base de l'information verbale fournie par les lecteurs d'écrans, et que les indices non-verbaux n'aient joué un rôle que de redondance, la tâche ayant à priori été plutôt facilement cognitivement parlant. Ces indices verbaux leur auraient donc suffi pour traiter et mémoriser la structure du texte. L'hypothèse de la moins grande efficacité de ces indices reposait en effet sur une *surcharge informationnelle* qui a pu ne pas être induite par notre tâche, pour les raisons évoquées précédemment. Une autre explication serait que l'habitude des sujets à traiter les indices verbaux ait pu compenser un potentiel effet positif des stratégies utilisant des indices non-verbaux ; pour évaluer cette hypothèse il faudrait faire passer la même expérimentation à des sujets non-déficients novices des synthèses vocales.

On observe cependant une tendance des préférences individuelles à influencer sur les performances : on remarque que les sujets ayant préféré le son spatialisé ou l'utilisation des deux voix tendent à observer de meilleures performances dans la condition qu'ils ont préférée. Ces résultats n'ont pu être traités statistiquement car les groupes n'étaient pas de même taille : ( $N = 3$ ) sujets ont préféré le son spatialisé tandis que ( $N = 6$ ) sujets ont préféré l'utilisation d'une seconde voix pour lire les titres.

## **4. Évaluation de 4 stratégies de restitution de la MFM auprès de sujets voyants**

---

---

Cette section présente une expérimentation menée sur l'évaluation de 4 stratégies de restitution de la MFM des documents auprès de 60 sujets non-déficients. Les résultats de cette étude n'ont pas encore été publiés.

Pour cette expérimentation, et au vu des résultats de la première expérimentation, nous avons tenté une approche plus globale de la restitution de la MFM, conformément à notre hypothèse générale.

## 4.1. Présentation des hypothèses

---

Sur la base de notre hypothèse globale stipulant que restitution de la MFM des textes permet une meilleure compréhension du contenu, et dans le cadre d'une tâche de compréhension d'un texte, nous avons formulé 3 catégories d'hypothèses portant sur : 1) l'impact du contenu et méta-informations à restituer au lecteur (le « *quoi* ») et les stratégies de restitution de l'information comprenant 2) « *comment* » restituer l'information et 3) « *quand* » restituer cette information. Ces différentes hypothèses, basées sur le rôle de la MFM décrit dans la littérature et sur les résultats de notre étude en situation (voir Chapitre II7), ont permis de guider la conception des 4 stratégies de restitution évaluées pour cette expérimentation (voir 4.2.c).

### 4.1.a. Hypothèses : contenu et méta-informations, le « *quoi* »

---

Hypothèse 1 : *Restituer les architectures textuelles du document (informations locales et transverses) permet d'améliorer la compréhension et de réduire la charge cognitive extrinsèque.* En effet, la MFM permet en lecture visuelle de (voir Chapitre I pour plus de détails sur le rôle de la mise en forme dans la compréhension de textes dans le cas de la lecture visuelle) :

- § Limiter les inférences structurales pendant la lecture (charge cognitive).
- § Désambiguïser certains contenus (compréhension).
- § Faciliter l'évaluation et la sélection des contenus, et ainsi permettre aux lecteurs d'activer les traitements adéquats [Lorch 1989] (charge cognitive et compréhension). Cela est rendu possible à travers les regroupements (initialement visuels) d'information, permettant à la fois de typer les contenus regroupés, de mettre en relation ces ensembles, mais aussi de mettre en avant ce qui est « important ».
- § Faciliter et améliorer la construction d'une représentation globale du document, en utilisant les informations structurales véhiculées par la mise en forme. Cela permet notamment de comprendre et hiérarchiser les thèmes, ce qui représente un point crucial pour la compréhension du texte [Lemarié et al. 2012] (charge cognitive et compréhension).

Hypothèse 2 : *Supprimer les informations non-pertinentes et/ou redondantes allège la charge cognitive.* Quelle que soit leur nature (contenu ou méta-informations), les informations redondantes et non pertinentes à la tâche augmentent massivement la charge cognitive extrinsèque (i.e., non pertinente pour apprendre) [Sweller 2010]. À l'oral, filtrer le contenu pour ne restituer que ce qui est pertinent (pour l'objectif fixé par la tâche) permettrait d'alléger la charge cognitive, d'autant que la modalité orale ne permet pas une sélection de l'information avant son traitement comme c'est le cas en lecture visuelle. Il a été démontré que filtrer de informations non-pertinentes à l'oral allège effectivement la charge cognitive dans le cas de tâches de navigation et recherche d'information sur un site web [Giraud 2014]. Nous mettrons à l'épreuve cette hypothèse dans le cadre d'une tâche de compréhension d'un texte expositif.



#### ***4.1.b. Hypothèses : stratégies de restitution de la MFM, le « comment »***

---

On parlera ici de restitution « *passive* » pour désigner toute restitution orale d'information qui se fait pendant la lecture, sans que l'utilisateur n'ait à effectuer aucune action particulière. À contrario, la restitution « *active* » fera référence aux commandes utilisables pour obtenir certaines informations.

Hypothèse 3 : *Restituer « passivement<sup>113</sup> » les informations locales de MFM en utilisant des sons non-verbaux et les propriétés des voix de synthèse permet d'alléger la charge cognitive et de diminuer la désorientation.*

La littérature montre en effet que les sons non-verbaux peuvent être plus efficaces que le verbal pour restituer de l'information en situation de charge cognitive [Klatzky et al. 2006].

Concernant la désorientation (i.e. la capacité à savoir où l'on est et où l'on va/peut aller dans un document [Ahuja and Webster 2001]), les sons non-verbaux ont déjà été utilisés dans le cas de tâches de navigation dans des structures hiérarchiques complexes, montrant leur plus grande efficacité [Smith et al. 2004; Maria and Anderson 2003] ; elles joueront ici le rôle de points de repères audio dans le document lors de la navigation. Les textes comprenant une structure hiérarchique, si la restitution des indices de structure se fait à l'aide de sons non-verbaux, la navigation au sein de cette structure devrait permettre aux utilisateurs de mieux s'orienter. À noter que [Kim and Hirtle 1995] pointe le fait que la complexité des structures hiérarchiques et la difficulté de leur accès, leur visualisation, est un facteur de désorientation.

Enfin, modifier les propriétés de la voix de synthèse (ex. modifier la prosodie, utiliser différentes voix, spatialiser la voix) permet de véhiculer une information supplémentaire en parallèle du message, afin de typer le contenu ou rajouter une information relative à ce dernier. On sait notamment que les pauses et la hauteur de la voix sont des marqueurs de la structure des documents utilisés spontanément par des lecteurs humains lors de la lecture visuelle d'un texte [Ouden et al. 2002]. Ces notions nous paraissent particulièrement intéressantes car elles font écho à une propriété d'un texte mis en forme qui est de véhiculer à la fois de l'information verbale et de l'information visuelle, qui sont traitées séparément dans en mémoire de travail [Baddeley and Hitch 1974], et donc de manière optimisée car le traitement est réparti sur deux unités distinctes (la boucle phonologique traitant le langage et le calepin visuo-spatial). Même si le modèle initial de Baddeley (le plus répandu en psychologie cognitive pour modéliser la mémoire de travail) ne considère pas l'audio au-delà du langage parlé, il existe dans la littérature des travaux démontrant l'existence d'unités de traitement séparées pour la musique et les sons non-verbaux [Berz 1995]. Cette propriété de la cognition humaine pourrait ainsi permettre d'optimiser la transmission de l'information à l'oral en permettant de typer le contenu verbal ou rajouter des méta-informations sur le texte lu à l'aide d'indices non-verbaux.

Hypothèse 4 : *Restituer activement la structure globale du texte (i.e. le plan) réduit la charge cognitive et la désorientation, et améliore la compréhension.* Cette hypothèse se base sur deux constats : comme mentionné précédemment, la structure joue un rôle central dans la compréhension ; de plus, la linéarité de la modalité audio ne permet pas d'avoir accès à cette structure en même temps que l'on accède au contenu. Ainsi nous faisons l'hypothèse qu'une

---

<sup>113</sup> La distinction entre restitution passive et restitution active se fait relativement aux actions utilisateur : la restitution qui se ferait à l'appui d'une commande est dite « active » et la restitution qui se fait pendant la lecture sans intervention de l'utilisateur est dite « passive »

restitution active (relevant de la dimension fonctionnelle de la MFM, car liée à des fonctions de lecture visuelle des titres) de la globalité de la structure permettrait une meilleure compréhension et une charge cognitive moins élevée.

### ***4.1.c. Hypothèse : stratégies de restitution de la mise en forme, le « quand » (non évaluée)***

---

Hypothèse / principe de conception 5 : *Prendre connaissance de la structure globale du texte avant d'accéder au contenu en lui-même améliore la compréhension et réduit la charge cognitive.* La présentation avant la lecture du texte d'une vue d'ensemble de la structure d'un texte (ex. une table des matières) permet d'explicitier l'organisation des thèmes et évite ainsi au lecteur d'avoir à inférer cette hiérarchie durant la lecture [Lorch 1989]. De plus, il a montré dans la littérature en psychologie des apprentissages que les *advance-organizers* (« informations présentées avant l'apprentissage permettant de mieux organiser et interpréter le contenu présenté en suivant [Mayer 2003]) ont un effet bénéfique sur l'apprentissage [Luiten et al. 1980]. Enfin, prendre connaissance de l'ensemble de la structure d'un document pour mieux se repérer avant de parcourir le contenu est une stratégie de navigation déjà utilisée par utilisateurs de lecteurs d'écran, notamment les non-voyants [Borodin et al. 2010].

Cette hypothèse n'a pas été évaluée directement avec les stratégies de restitution définies, dans le sens où l'on ne comparait pas la présentation de la structure globale avant ou après la lecture, par exemple. Cependant cette hypothèse a guidé la conception de la stratégie de restitution dite « optimale », qui sera comparée aux 3 autres dans les résultats. La prise en compte de cette hypothèse dans la conception relève de l'importance de centrer cette dernière autour des objectifs et caractéristiques de la tâche de compréhension, ici avec les *advance-organizers*.

## **4.2. Méthode**

---

Afin d'évaluer les hypothèses formulées ci-avant, 4 stratégies de restitution ont été évaluées. La population ayant pris part à cette étude, le matériel, le document utilisé, la procédure de test ainsi que ces stratégies de restitution sont présentées dans les sections ci-après.

### ***4.2.a. Participants***

---

Les 4 stratégies de restitution ont été évaluées auprès de 64 sujets non-déficients ; les données de 4 sujets ont été éliminées suite à des erreurs dans le décompte du temps imparti, ces sujets-ci ayant eu plus de temps que ce qui était prévu. Chacun des 60 sujets accédait au document à l'aide d'une des 4 stratégies de restitution, de sorte que le document ait été oralisé 15 fois avec chaque stratégie. Sur ces 60 sujets, 30 étaient de sexe masculin et 30 de sexe féminin, d'un âge moyen de 28 ans ( $SD = 10$  ans). Leur participation était volontaire mais gratifiée d'un chèque cadeau d'une grande enseigne.

### ***4.2.b. Matériel***

---

Les sujets pouvaient utiliser un clavier AZERTY standard qui était placé devant eux. L'écran de l'ordinateur sur lequel le logiciel READ était installé n'était pas visible, et servait à l'expérimentateur

à contrôler la bonne marche du test. Les sujets étaient équipés d'un casque audio pour la lecture du document et portaient un masque occultant sur les yeux.

Enfin, nous disposions pour ces tests de deux voix de synthèse respectant la norme SAPI5<sup>114</sup> : Virginie et Alain. L'ensemble des tests ont été réalisés avec l'application READ que nous avons développée (voir Chapitre IV5 pour plus de détails sur cette application)

Le document abordé par les sujets pendant les tests traitait d'économie, et plus précisément de l'inflation en France. Le choix du thème du document a été contraint par le fait qu'un maximum de sujets devaient comprendre le fond du texte sans pour autant être trop familiers avec les notions abordées. En effet, étant donné que nous mesurons la compréhension, l'expertise des sujets relative au domaine abordé dans le document aurait grandement influencé la compréhension [Chi et al. 1988]. De plus, en lecture visuelle, l'expertise des lecteurs influence également leur utilisation de la MFM [Wilhite 1989; Meyer et al. 1980]. Enfin, la MFM jouant un rôle de facilitateur de traitement de l'information [Lemarié 2006a], le contenu devait être assez complexe pour que l'on puisse observer un potentiel effet de la restitution de la MFM.

Le contenu du document était issu de divers sites de cours d'économie en ligne. Le document contenait environ 1500 mots, ce qui correspond à 10 minutes de lecture par synthèse vocale à 150 mots par minute ; cette vitesse d'énonciation correspond au rythme recommandé pour les livres audio grand public [Williams 1998], mais reste légèrement plus lent que le rythme par défaut de JAWS qui est d'environ 180 mots par minute<sup>115</sup>. Le même fichier de départ était utilisé pour chacune des 4 conditions expérimentales, et donc le même contenu était accessible pour chaque sujet. Cependant la durée totale de lecture différait de plus ou moins 2 minutes en fonction de la condition, cette durée correspond à l'ajout de métatexte dans le document.

La structure du document comprenait 3 niveaux hiérarchiques : 3 parties en tout, 5 sous-parties et 6 rubriques dans les sous-parties. L'imbrication des titres a été conçue de telle sorte qu'elle ne corresponde pas à une logique trop évidente (ex. 1. Animaux marins; a) Les dauphins ; b) les requins) mais qu'elle reste tout de même cohérente (ex. 1. Mesure et évolution de l'inflation ; a) la mesure ; b) l'évolution dans le temps).

Le document comprenait un certain nombre de phénomènes de MFM qui puissent être restitués par notre système, notamment : des titres, des énumérations (à plusieurs niveaux, incluses dans le discours), des notes de bas de page, des tableaux et des encadrés. Le document comprenait également certains phénomènes ambigus (i.e. s'ils ne sont pas signalés correctement l'interprétation du contenu correspondant est ambiguë), et un certain nombre de contenus périphériques à la compréhension du fond du document : des images, des remarques, des notes de bas de page. Douze blocs de contenus périphériques étaient inclus dans le document, pour un total d'environ 300 mots.

Chaque architecture textuelle véhiculée par la MFM a été formalisée à l'aide de la méthodologie définie au Chapitre III selon les étapes suivantes : découpage du document en objets textuels sur la base du contraste entre les segments de texte, typage des objets textuels et mise en relation de ces objets. Une copie du document utilisé pour les tests est disponible en 4.4 ci-dessous.

---

<sup>114</sup> Speech Application Programming Interface : interface de programmation permettant aux applications Windows (framework .NET) d'utiliser les voix de synthèse installées

<sup>115</sup> Source : <http://webaim.org/techniques/screenreader/>

#### ***4.2.c. Stratégies de restitution évaluées***

---

Afin de tester les hypothèses énoncées (voir partie 4.1), nous avons défini 4 configurations différentes à l'aide de READ (voir Chapitre IV5 pour plus de détails sur la plateforme développée). Chacune de ces 4 configurations correspond à une condition expérimentale testée auprès de 15 sujets avec le même document. On souhaitait ainsi évaluer 4 niveaux de restitution de la MFM ; une courte vidéo informelle présentant ces 4 stratégies de restitution est disponible en ligne ici<sup>116</sup>. Ci-dessous un aperçu de chacune des stratégies, qui seront détaillées dans les sections suivantes :

1. Condition « sans mise en forme » (codée<sup>117</sup> par la lettre **N**, en référence à la négation). Comme son nom l'indique, cette condition était une condition de contrôle utilisée comme point de repère du point de vue des mesures. Cette condition permettait de fournir un point de comparaison pour vérifier l'apport des 3 autres stratégies.
2. Condition « lecteur d'écran » (lettre **J**, en référence au lecteur d'écran JAWS). Cette condition reproduisait la restitution de la MFM produite par les lecteurs d'écrans lors d'un accès par synthèse vocale. Cette condition nous paraissait nécessaire d'un point de vue « écologique » au regard des technologies actuelles. Ici il s'agissait de comparer les stratégies de restitution que nous avons élaborée avec un autre point de référence : l'existant.
3. Condition « optimale » (lettre **O**). Cette condition s'attachait à restituer les informations de mise en forme d'une manière la plus optimale possible (définie lors de la conception et dont on a posé les bases précédemment dans le Chapitre IV4) et à filtrer les contenus non-pertinents avec la tâche.
4. Condition « exhaustive » (lettre **X**). Cette condition restituait l'ensemble des informations de MFM (de la même manière qu'en condition optimale) et l'ensemble du contenu (aucun filtrage) ; l'objectif était d'évaluer la pertinence d'une restitution exhaustive des informations de MFM.

À noter que pour chaque condition, une fois le test commencé la lecture démarrait du début du document et ne s'arrêtait que si l'utilisateur utilisait la commande « pause ». Les fonctions de navigation pour chacune des conditions changeaient l'emplacement du curseur de lecture dans le document ; la lecture continuait alors immédiatement après le déplacement. Ces déplacements se faisaient de *bloc en bloc*<sup>118</sup>, sauf pour la condition N ; on utilise le terme bloc pour désigner un regroupement de texte quel qu'il soit (titre, paragraphe, ligne). En fonction de la condition, la segmentation du texte en blocs se faisait selon différents critères, mais toujours basés sur les marques de MFM.

Les caractéristiques de chacune de ces 4 conditions (configurations) seront donc développées dans les sections ci-après ; seront notamment décrites pour chaque configuration les commandes disponibles ainsi que les modalités de restitution audio des informations de mise en forme.

---

<sup>116</sup> <https://youtu.be/5bz0GiO3Dmc>

<sup>117</sup> Afin de rendre la présentation des résultats plus lisible, on fait référence à chacune des 4 conditions expérimentales à l'aide d'une lettre unique. Les 4 lettres utilisées sont : N, J O et X

<sup>118</sup> « Bloc » au sens où on l'entend pour READ (la plateforme logicielle utilisée pour les expérimentations, voir Chapitre IV5), qui découpe le document en une séquence de bloc d'informations ; ici chaque bloc correspondait à un objet textuel.

#### 4.2.c.1. Condition « sans mise en forme » (N)

---

Pour cette condition expérimentale, le document était restitué à la manière d'un fichier audio auquel on accéderait depuis un lecteur multimédia, c'est-à-dire à la manière d'un flux audio continu. Tout phénomène de MFM était ignoré de sorte que le document soit lu comme un seul bloc de texte.

Description de la stratégie : Le document était lu par une seule voix de synthèse. La fin des blocs visuels (ex. paragraphes, titres) n'était pas signalée : à la fin de chaque bloc la synthèse vocale effectuait une pause d'une durée égale à la pause faite lors d'une phrase, de sorte qu'on ne puisse distinguer une fin de phrase d'une fin de bloc.

##### Touches et fonctions utilisables :

- § Touche ESPACE : pause / reprendre la lecture à l'emplacement où elle s'est arrêtée.
- § Touche FLECHE\_GAUCHE : saut de 3 secondes en arrière dans le fichier audio (la touche peut être maintenue pour répéter la commande rapidement).
- § Touche FLECHE\_DROITE : saut de 3 seconds en avant dans le fichier audio (la touche peut être maintenue pour répéter la commande rapidement).

#### 4.2.c.2. Condition « lecteur d'écran » (J)

---

Pour cette condition expérimentale, la lecture était similaire à une lecture vocale faite par un lecteur d'écran, avec des fonctions cependant limitées. Cette lecture « standard » correspond à la configuration par défaut d'un lecteur d'écran comme JAWS en termes de restitution orale. Trois entretiens avec des formateurs d'un institut pour jeunes aveugles nous ont confirmé que seuls les utilisateurs avancés modifiaient la configuration de leur lecteur d'écran sur le plan de la restitution audio, et que la plupart ne le faisaient que de façon limitée. De plus les enseignements qu'ils dispensent portant sur les lecteurs d'écrans sont réalisés avec la configuration par défaut.

Il est important de préciser que, malgré son nom, cette condition ne reproduit pas totalement l'accès au document par un lecteur d'écran, mais seulement la partie restitution orale (ou « lecture »). Les lecteurs d'écrans offrent en effet pléthore de fonctions pour se déplacer dans les documents et effectuer diverses opérations, qui ne sont ici pas implémentées. L'objectif est bien de comparer, non pas des systèmes, mais des stratégies de restitution de la mise en forme.

Description de la stratégie : Le document était lu par une seule voix de synthèse. Le document était découpé en blocs visuels (i.e. segmentés en fonction des espaces entre les blocs de texte) ; la fin des blocs était signalée pendant la lecture par une pause d'environ 0.5 secondes. Avant la lecture du contenu de chaque titre, la voix énonçait le niveau du titre (ex. « Titre de niveau 1 ») ; cela correspondant à la manière dont les lecteurs d'écrans signalent les titres par défaut.

##### Touches et fonctions utilisables :

- § Touche ESPACE : pause / reprendre la lecture au début du bloc où la lecture s'était arrêtée
- § Touche FLECHE\_GAUCHE : aller au bloc précédent. La touche pouvait être maintenue pour effectuer plusieurs déplacements rapidement.
- § Touche FLECHE\_DROITE : aller au bloc suivant. La touche pouvait être maintenue pour effectuer plusieurs déplacements rapidement.

### 4.2.c.3. Condition « optimale » (O)

---

Pour cette condition, les sujets accédaient au document avec une stratégie que nous avons essayé d'optimiser relativement à nos hypothèses, notamment sur le plan de la compréhension et de la charge cognitive associée.

La conception de cette stratégie de restitution s'est basée sur l'étude de la littérature des possibilités de restitution de l'information à l'oral (voir Chapitre IV), ainsi que sur 5 pré-tests réalisés avec des volontaires du laboratoire. De manière générale, la conception a été guidée par une double contrainte : rendre accessible un maximum d'informations de MFM (architectures textuelles décrites dans le métatexte), potentiellement utiles pour la compréhension, tout en évitant de donner trop d'informations (ou de les restituer d'une manière peu pertinente) car cela aurait l'effet inverse, en surchargeant la mémoire de travail des utilisateurs. Il fallait donc éviter que les modalités de restitution des informations de MFM (sons, texte et commandes principalement) ne présentent une trop grande complexité, et donc qu'elles ne soient ni trop nombreuses ni trop difficilement interprétables ou mémorisables. En effet, au lieu d'aider les sujets à mieux comprendre le contenu, une restitution trop complexe (que cette complexité provienne des modalités de restitution ou simplement de la quantité d'informations restituées) aurait contraint les sujets à utiliser leurs ressources cognitives pour comprendre la restitution en elle-même ; ces derniers auraient donc eu moins de ressources cognitives pour la compréhension du contenu. Cela nous a amené dans notre cas à choisir les informations et fonctions que nous pensions les plus pertinentes pour la tâche de compréhension.

Ce compromis était nécessaire dans un cadre expérimental comme celui-ci où les sujets n'utilisaient le système qu'une seule fois ; en revanche, il reste discutable si l'on souhaite développer un système de lecture à part entière, en considérant le fait que l'apprentissage puisse automatiser les traitements des informations de mises en forme. On pourrait alors restituer un plus grand nombre d'informations et de fonctions potentiellement utiles à un moment donné pour l'utilisateur, qui seraient utilisées une fois une certaine expertise du système acquise.

Cette stratégie restitue donc un certain nombre d'informations de mise en forme « localement », c'est-à-dire au fil de la lecture, les 3 mêmes fonctions de navigation que la condition précédente, ainsi que 2 fonctions supplémentaires proposant une restitution globale de l'information de mise en forme.

#### Description de la stratégie :

- § Pour cette stratégie les blocs étaient à la fois segmentés d'après les démarcations visuelles du document d'origine (le contraste entre objets textuels), mais aussi sur la base de marqueurs lexico-syntaxiques, quand par exemple certaines énumérations n'étaient pas marquées par des puces ou des retours à la ligne, mais des marqueurs lexicaux ou syntaxiques au sein des phrases (ex. « *il est important de comprendre que : 1).... ; 2)....* »).
- § Deux voix de synthèse étaient utilisées pour la lecture du document. Une voix féminine était dédiée à la lecture du contenu du document, tandis qu'une voix masculine lisait les blocs de métatexte, ajoutés au document en tant qu'équivalents informationnels de la MFM du texte d'origine. La voix masculine jouait le rôle de « voix-off » en quelque sorte. De cette façon, les voix étant très différentes et facilement discriminables, il était possible de connaître à tout instant la nature du contenu en cours de lecture (texte ou métatexte).

- § Concernant la restitution des titres, chaque titre était d'abord signalé par une icône audio très courte (environ 800 millisecondes) dont la hauteur moyenne en fréquence dépendait du niveau du titre (i.e. plus le titre avait un niveau élevé plus le son était aigu). L'objectif était de donner un repère de navigation pertinent, étant donné le rôle essentiel des titres dans la structuration des textes. Une fois l'icône audio jouée, le niveau du titre était lu par la voix masculine, puis le contenu du titre par la voix féminine.
- § Pour cette stratégie, nous avons utilisé les capacités du système auditif humain à localiser une source sonore dans l'espace pour restituer les informations sur la taille des blocs de texte pendant la lecture. Ainsi, la lecture des blocs était spatialisée et commençait toujours depuis une source sonore située sur la gauche des sujets pour se terminer sur leur droite ; ce choix a été motivé par le fait que la discrimination de sources sonores dans l'espace est la plus fine sur l'axe droite/gauche au niveau de la tête [Goose and Möller 1999; Rumsey 2001]. La source sonore se déplaçait donc en arc de cercle vers la droite une fois que la lecture arrivait au milieu du bloc de texte, et continuait à se déplacer jusqu'à la fin de la lecture du bloc. Ainsi, avec la localisation et la vitesse du déplacement du son, les utilisateurs pouvaient estimer la taille du paragraphe et la quantité de contenu restant avant la fin du paragraphe. L'implémentation de cette fonction a été motivée par l'importance avérée dans la littérature de ces informations en lecture visuelle : la lecture visuelle ralentit en fin de paragraphe pour que le lecteur puisse mettre en place des processus d'inférences et d'intégration du contenu [Just and Carpenter 1980], étant donné qu'une fin de paragraphe annonce la plupart du temps une rupture thématique.
- § Les énumérations du texte étaient précédées d'une icône audio (choisie pour sa grande différence en timbre avec l'icône audio utilisée pour les titres) et d'un bloc de métatexte de « prévisualisation » de l'énumération. Ce bloc de métatexte explicitait combien d'items comprenait l'énumération et quel était leur type (ex. « Trois figures de styles au théâtre »). L'icône audio devait servir de repère pour la navigation, et le bloc de prévisualisation devait permettre une première représentation du contenu qui allait être abordé, une sorte de préparation cognitive à l'intégration des informations à venir.
- § D'autres informations de mise en forme étaient incorporées au texte ; c'est à dire que l'on rajoutait du contenu textuel explicitant les phénomènes visuels suivants :
- Pour les items des énumérations, la lecture de l'item numéro N débutait par l'annonce de ce numéro ainsi que d'un type de l'item énoncé : N<sup>ième</sup> cause du réchauffement climatique)
  - Les phénomènes syntaxiques présentant une potentielle ambiguïté à l'oral étaient explicités par l'ajout de texte (ex. un exemple entre parenthèses dans une phrase était précédé du texte « exemple : »).
- § L'objectif était ici d'optimiser la restitution pour la tâche de compréhension. Or, on sait que les contenus non-pertinents pour la tâche augmentent la charge cognitive de façon inutile [Giraud et al. 2011]. En condition optimale, les contenus périphériques du document (ex. images illustratives) étaient donc supprimés de la lecture.
- § Enfin, avant d'accéder au document en lui-même, les sujets avaient en début de test 2 minutes et 10 secondes pour parcourir la table des matières du document, et ce afin de tester la validité de notre dernière hypothèse (voir 4.1.c)

Touches et fonctions utilisables :

- § Touche ESPACE : pause / reprendre la lecture au début du bloc où la lecture s'était arrêtée
- § Touche FLECHE\_GAUCHE : aller au bloc précédent. La touche pouvait être maintenue pour effectuer plusieurs déplacements rapidement.
- § Touche FLECHE\_DROITE : aller au bloc suivant. La touche pouvait être maintenue pour effectuer plusieurs déplacements rapidement.
- § Touche ECHAP : fonction « où suis-je » ; l'appui sur cette touche lance la lecture de la liste des titres englobant le bloc en cours de lecture (ex. pour ce paragraphe de la thèse, la liste des titres englobant est : Chap. 6 [...], Partie III [...], Point 2. [...] Section C. [...], Sous-section 3 [...]). Cette fonction s'appuie sur un résultat de la littérature qui montre son utilité dans le cas du parcours d'une structure hiérarchique constituée par une arborescence de fichiers [Smith et al. 2004].
- § Touche ENTREE : ouvre la table des matières comprenant l'ensemble des titres du document. L'utilisateur pouvait naviguer dans la table des matières comme il naviguait dans le document. La touche ENTREE servait également à revenir à la lecture du document ; le curseur de lecture était alors placé dans le document au niveau du dernier titre lu dans la table des matières (comme si les titres de la table des matières étaient des liens « hypertexte » vers les titres correspondants dans le document).

### 4.2.c.4. Condition exhaustive (X)

---

Cette condition est semblable à la condition optimale, excepté le fait que les contenus périphériques sont également lus (alors qu'ils étaient supprimés de la lecture pour la condition optimale), et que le lecteur n'a ici accès ni à la table des matières ni à la fonction « où suis-je ? ». Ainsi cette condition permet d'évaluer à la fois l'impact de ces deux fonctions (par différence avec la condition optimale), mais aussi l'impact des modalités de restitution non-fonctionnelles (restitution locale de l'information) en elles-mêmes, et notamment leur capacité à permettre au lecteur de filtrer des contenus non pertinents.

Description de la stratégie : mêmes chose que pour la condition optimale (voir section précédente).

Touches et fonctions utilisables :

- § Touche ESPACE : pause / reprendre la lecture au début du bloc où la lecture s'était arrêtée
- § Touche FLECHE\_GAUCHE : aller au bloc précédent. La touche pouvait être maintenue pour effectuer plusieurs déplacements rapidement.
- § Touche FLECHE\_DROITE : aller au bloc suivant. La touche pouvait être maintenue pour effectuer plusieurs déplacements rapidement.

### 4.2.d. Procédure de test

---

À noter que l'ensemble des questionnaires utilisés sont disponibles en 4.4 ci-dessous. Les tests se sont systématiquement déroulés de la manière suivante :

- § *Signature d'un formulaire de consentement éclairé*
- § *Évaluation des connaissances en économie du participant* : 4 questions dont 1 d'auto-évaluation et 3 autres évaluant la fréquence de la pratique de l'économie
- § *Lecture d'un document d'entraînement avec le système*, présentant les fonctions et modalités de lecture. Il y avait un document d'entraînement par condition.



- § *Lecture et navigation dans le document traitant d'économie*, avec l'une des 4 stratégies de lecture assignée aléatoirement.
- Consigne : « *Votre objectif pendant la lecture du document est d'essayer de comprendre au mieux le contenu du document ainsi que sa structure. Vous êtes libre de naviguer dans le document comme vous le souhaitez durant le temps qui vous est imparti.* »
  - (Uniquement en condition optimale) Parcours de la table des matières pendant 2 minutes maximum.
  - Temps imparti : durée totale de lecture passive du document plus 50% (ici environ 15min pour un document de 10minutes). Les sujets étaient libres d'arrêter avant la fin du temps imparti.
  - L'ensemble des actions des- sujets et des évènements système étaient enregistrés dans un fichier texte.
- § *Évaluation de la charge cognitive*
- 5 questions, détaillées dans la sous-partie des résultats correspondante
  - Les questions évaluaient la charge cognitive globale ainsi que plusieurs de ses composantes : la charge intrinsèque (liée au contenu), la charge extrinsèque (liée à la modalité orale, aux fonctions de lectures et aux informations métatextuelles), et la charge liée aux processus d'apprentissage [Paas et al. 2003].
- § *Évaluation de la désorientation*
- 6 questions mesurant le sentiment de désorientation des sujets. La validité de ces questions pour mesurer la désorientation a été vérifiée par [Ahuja and Webster 2001].
  - La désorientation est un concept qui a vu son émergence avec la navigation hypertexte [McDonald and Stevenson 1996] ; ce concept est analogue à l'orientation spatiale car il désigne le sentiment d'être perdu en termes de navigation (ne pas savoir où l'on se trouve, comment aller où l'on souhaite, ne pas savoir où aller).
- § *Question d'auto-évaluation des connaissances post-lecture*. Il était demandé aux sujets d'évaluer leur familiarité avec le contenu lu.
- § *Évaluation de la rétention du plan*. Il était demandé aux sujets de restituer le plan du texte (titres ou à défaut les thèmes abordés) aussi précisément que possible ; leur restitution était enregistrée à l'aide d'un dictaphone
- § *Évaluation de la compréhension du contenu*.
- 8 questions de compréhension ; les réponses étaient enregistrées à l'aide d'un dictaphone. 3 questions trouvaient leurs réponses dans la structure définie par le plan du document (ex. plusieurs sous-titres). 3 questions portaient sur des points plus précis du document (ex. une énumération). Les réponses à ces 6 questions étaient réparties dans toutes les parties du texte ; une attention particulière a été portée sur le fait de ne pas centrer les questions sur une partie du texte. Enfin 2 questions évaluaient les liens qu'ont pu faire les sujets entre leurs connaissances antérieures et le contenu appris.
- § *Tâches de recherche d'information* : deux emplacements à retrouver dans le document.
- Le document était rouvert ( curseur de lecture au début du document). Il était ensuite demandé au sujet : « *de naviguer le plus rapidement possible jusqu'à l'emplacement dans le document qui traite de...* ».

- Cette procédure était répétée 2 fois. On mesurait à chaque le temps mis pour atteindre la cible, et si l'emplacement pointé correspondait à l'emplacement demandé.

§ *Questionnaire SUS*<sup>119</sup> [Brooke 1996] mesurant l'utilisabilité du système.

§ *Question ouverte* : commentaires des sujets sur le système et le test.

Un test durait entre 35 et 50 minutes.

### 4.3. Résultats attendus

---

Nous reprenons dans cette partie chacune des hypothèses formulées, en donnant les résultats attendus par rapport aux stratégies de restitution définies.

Hypothèse 1 : *Restituer les informations de mise en forme du document permet d'améliorer la compréhension et d'alléger la charge cognitive.* La condition N ne restitue aucune information de mise en forme ; la condition J restitue la segmentation des blocs (à travers les pauses et les possibilités de navigation) et signale les titres ; les conditions O et X tentent de restituer un maximum d'informations de mise en forme localement, mais la condition O offre des restitutions globales supplémentaires. Au regard de cette hypothèse et de cette dernière uniquement, on s'attend à ce que la compréhension soit maximale et la charge cognitive minimale en condition O relativement aux 3 autres conditions, légèrement plus dégradées en condition X, encore plus en condition J, et finalement les plus dégradés en condition N.

Hypothèse 2 : *Supprimer les informations non-pertinentes et/ou redondantes allège la charge cognitive.* Ces informations étant filtrées uniquement en condition optimale « O », au regard de cette hypothèse et de cette dernière uniquement, on s'attend à ce que la charge cognitive soit plus faible en condition O que dans les autres conditions.

Hypothèse 3 : *Restituer pendant la lecture les informations locales de mise en forme en utilisant des sons non-verbaux et les propriétés des voix de synthèse permet d'alléger la charge cognitive et de diminuer la désorientation.* La condition N ne restitue aucune information de mise en forme ; la condition J restitue la segmentation des blocs et signale les titres ; les conditions O et X tentent de restituer un maximum d'informations de mise en forme. Au regard de cette hypothèse et de cette dernière uniquement, on s'attend à ce que la désorientation et la charge cognitive soient plus faibles en condition O et X que pour les conditions J et N.

Hypothèse 4 : *Restituer activement la structure globale du texte (i.e. le plan) réduit la charge cognitive et la désorientation, et améliore également la compréhension.* Au regard de cette hypothèse et de cette dernière uniquement, on s'attend à ce que les fonctions proposées en condition O permettent une meilleure compréhension, une charge cognitive et un sentiment de désorientation plus faible que pour les 3 autres conditions.

Hypothèse 5 : *Prendre connaissance de la structure globale du texte avant d'accéder au contenu en lui-même améliore la compréhension et réduit la charge cognitive.* Au regard de cette hypothèse et de cette dernière uniquement, on s'attend à ce que la lecture préalable de la table des matières en

---

<sup>119</sup> System Usability Scale

condition O permette une meilleure compréhension et une charge cognitive plus faible que pour les 3 autres conditions.

En reprenant pour chaque hypothèse les résultats attendus, la combinaison de l'ensemble des hypothèses devrait avoir les conséquences suivantes sur les principaux indicateurs :

§ Compréhension :  $O \gg X > J > N$

§ Charge cognitive :  $O \ll X < J < N$

§ Désorientation :  $O < X < J < N$

#### 4.4. Résultats

---

Cette partie présente les résultats de l'évaluation des 4 stratégies de restitutions définies avec 60 sujets non-déficients. On commencera par présenter les caractéristiques de notre population. Par la suite, les résultats des différents indicateurs seront chacun présentés dans une section dédiée ; les groupes indépendants correspondant aux 4 conditions expérimentales y seront à chaque fois comparés (*rappel des codes des conditions* :  $O = \text{condition optimale}$ ,  $J = \text{condition lecteur d'écran (JAWS)}$ ,  $N = \text{condition sans mise en forme (NoMFM)}$  et  $X = \text{condition exhaustive}$ ). La présentation des résultats dans chacune de ces sections se fera systématiquement de la manière suivante :

§ Caractéristiques de l'indicateur : voir 3.3.

§ Normalité des données : voir 3.3.

§ Présentation graphique des échantillons & commentaires : voir 3.3.

§ Tests statistiques et résultats de ces tests. Dans notre cas nous comparerons donc 4 échantillons indépendants pour chaque mesure, correspondants aux 4 conditions expérimentales. *Ces tests ont été réalisés à l'aide du logiciel R ; les fonctions utilisées correspondant aux tests statistiques sont présentées en Chapitre V4.3*

- Données normales : comparaison des moyennes avec un test d'analyse de la variance (ANOVA) à 1 facteur (ici la condition expérimentale). Les résultats des ANOVA seront présentés de la manière suivante :  $F(\text{ddl}_{\text{inter-groupes}}, \text{ddl}_{\text{intra-groupes}}) = F\_value, p < \text{ou} > \text{au seuil de significativité}$  (ddl désignent les degrés de liberté). Étant donné que la  $F\_value$  représente un ratio entre les moyennes, plus elle se rapproche de 1.0, plus il y a de chances d'accepter l'hypothèse nulle d'absence de différence entre les moyennes. On pourra être amené par la suite à effectuer un test post-hoc de l'ANOVA réalisés ici avec une implémentation du *Tukey-HSD*.
- Données non-normales : comparaison des rangs des observations avec un test de Kruskal-Wallis (données non normales). Les résultats des tests seront présentés de la manière suivante :  $H(\text{ddl}) = H\_value, p < \text{ou} > \text{au seuil de significativité}$  (ddl désignent les degrés de liberté). La  $H\_value$  représente ici l'écart entre les sommes des rangs des différents échantillons ; plus celle-ci est élevée, plus H est élevée, plus les distributions sont, à priori, différentes. On pourra être amené par la suite à effectuer un test post-hoc de comparaisons multiples comparant 2 à 2 les sommes des rangs de chaque échantillon.

Enfin, l'ensemble des résultats est synthétisé dans un tableau récapitulatif (voir 4.4.I).

#### 4.4.a. Caractéristiques de la population

L'expérimentation a été passée par 60 sujets plutôt jeunes ayant en moyenne 28 ans ( $SD = 10$  ans) ; le sujet le plus jeune avait 17 ans et le plus âgé 53. Tous les sujets avaient un niveau scolaire au moins équivalent à la classe de terminale. Aucun participant n'a rapporté de déficience particulière.

##### 4.4.a.1. Usage des technologies

Quatre questions étaient posées aux sujets au début des passations, afin d'évaluer l'usage qu'ils avaient des technologies utilisées durant cette expérimentation :

- § À quelle fréquence utilisez-vous un ordinateur ? (1 à 5)
- § À quelle fréquence utilisez-vous un smartphone ? (1 à 5)
- § À quelle fréquence écoutez-vous des voix de synthèse (GPS, traduction en ligne, etc.) ? (1 à 5)
- § À quelle fréquence écoutez-vous des livres audio ? (1 à 5)

Les résultats des réponses à ces questions sont présentés ci-dessous, par fréquence d'apparition de chacun des 5 niveaux de réponse. Pour chaque condition l'échantillon comprend ( $N=15$ ) données, soit ( $N=60$ ) sujets distribués sur 4 conditions.

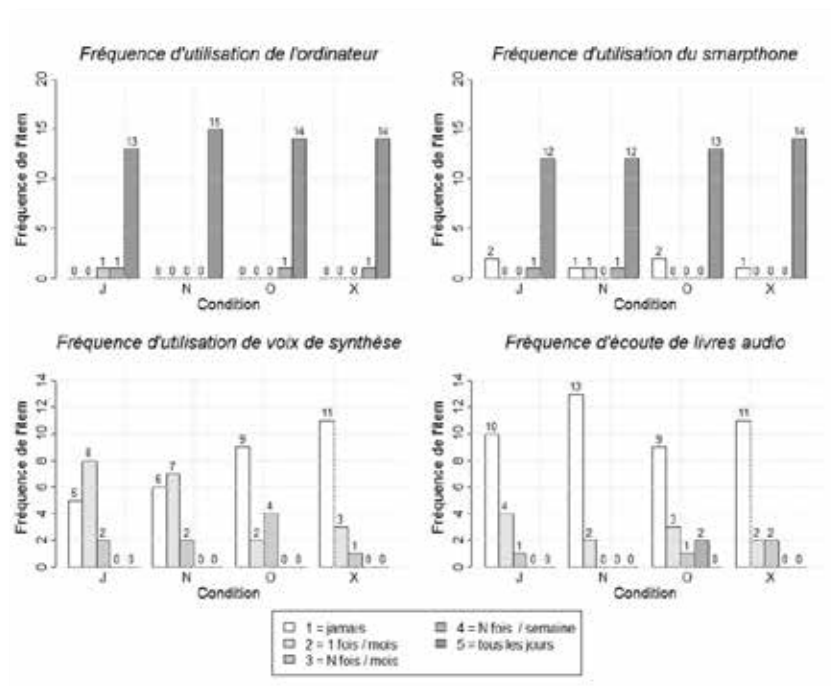


Figure 45 - Distributions des réponses aux questions qualifiant l'usage des technologies des sujets

Afin de tester l'homogénéité des échantillons pour chacune des 4 variables mesurées, correspondant aux 4 questions posées, nous comparons les échantillons de chaque condition. Cela est possible étant donné que la variable catégorielle, ici la condition, nous donne 4 groupes indépendants pour chaque mesure. Les groupes sont indépendants car les tests ont été menés en inter-sujets, c'est-à-dire que chaque sujet ne passait qu'une condition.

	Utilisation de l'ordinateur	Utilisation du smartphone	Écoute de voix de synthèse	Écoute de livres audio
Résultats - $\chi^2$	(6, N=60) = 4.14, p = 0.66	(9, N=60) = 5.88, p = 0.66	(6, N=60) = 10.24, p = 0.11	(9, N=60) = 9.81, p = 0.36

**Tableau 20- Résultats du test du  $\chi^2$  : questions sur l'usage des technologies**

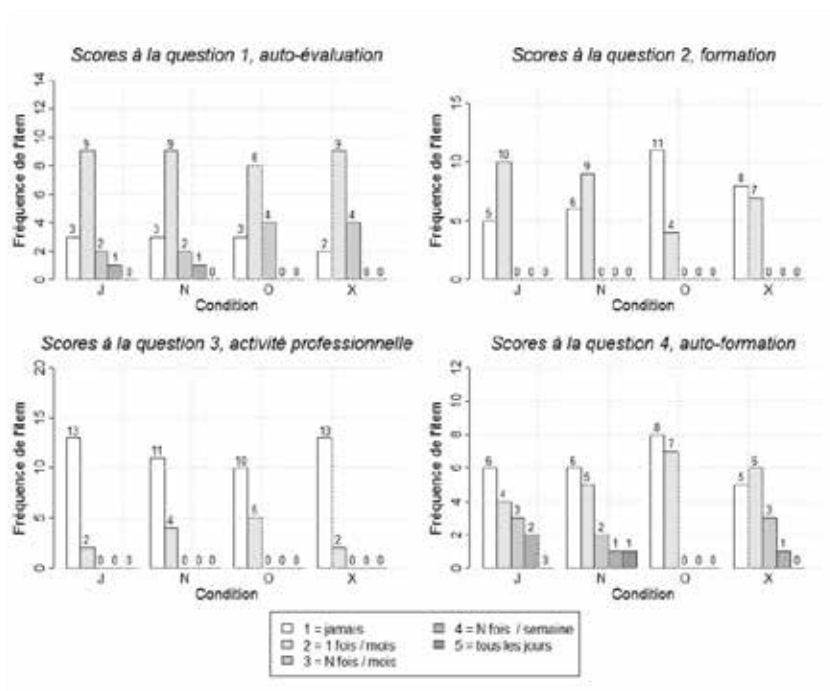
Les tests du  $\chi^2$  ne montrent aucune différence significative ( $p > .05$ ) entre les distributions de chaque variable entre les conditions, et ce pour chacune des 4 variables étudiées. On en conclut que notre population est homogène en termes de familiarité avec l'ordinateur (forte familiarité,  $M = 4.9$   $SD = 0.3$ ), avec le smartphone (forte familiarité,  $M = 4.5$   $SD = 1.3$ ), les voix de synthèse (faible familiarité,  $M = 1.6$   $SD = 0.7$ ) et les livres audio (faible familiarité,  $M = 1.4$   $SD = 0.8$ ).

#### 4.4.a.2. Connaissances antérieures

Pour évaluer les connaissances antérieures des sujets relatives au domaine abordé par le document, on leur posait 4 questions avant l'écoute du document, et une à postériori :

- § Auto-évaluation des connaissances en économie (1 à 5)
- § Fréquence de la pratique de l'économie pendant la formation (1 à 5)
- § Fréquence de la pratique de l'économie dans l'activité professionnelle (1 à 5)
- § Fréquence de l'auto-formation en économie (renseignements à titre personnel) (1 à 5)
- § Après l'écoute du document, on demandait ensuite au sujet d'évaluer de nouveau leurs connaissances par rapport à ce qu'ils avaient entendu (1 à 5)

Les résultats des réponses à ces questions sont présentés ci-dessous, par fréquence d'apparition de chacun des 5 niveaux de réponse.



**Figure 46 - Distributions des réponses aux questions de connaissances**

De la même façon que pour les variables de la section précédente, nous avons testé l'homogénéité des distributions de nos échantillons (les distributions des réponses à la question 1 sont-elles différentes d'une condition à l'autre ?). Pour cela nous avons effectué un test du  $\chi^2$ .

	Question 1	Question 2	Question 3	Question 4
Résultats - $\chi^2$	(9, N=60) = 4.14, p = 0.93	(3, N=60) = 5.88, p = 0.13	(3, N=60) = 10.24, p = 0.45	(45, N=60) = 9.81, p = 0.64

Tableau 21 - Résultats du test du  $\chi^2$  : questions de connaissances

Les tests du  $\chi^2$  ne montrent pas de différences significatives ( $p > .05$ ) entre les distributions des différentes variables entre les conditions, et ce pour chacune des 4 questions d'évaluation des connaissances. Concernant la question d'évaluation des connaissances posée après l'écoute du document, pas de différences non plus entre les distributions des 4 conditions :  $\chi^2 (12, N=60)=16.75$ ,  $p= 0.1592$ , avec un score dans la moyenne pour les participants ( $M = 2.6$ ,  $SD = 0.9$ ).

On peut donc en conclure que sur les 4 conditions expérimentales, les sujets avaient des connaissances antérieures homogènes. Ces connaissances antérieures étaient peu importantes, si l'on s'en réfère par exemple à la moyenne globale des réponses sur les 4 conditions pour les 4 questions présentées ( $M = 1.7$ ,  $SD = 0.4$ ).

#### 4.4.b. Effet de la stratégie de restitution sur la rétention du plan du texte

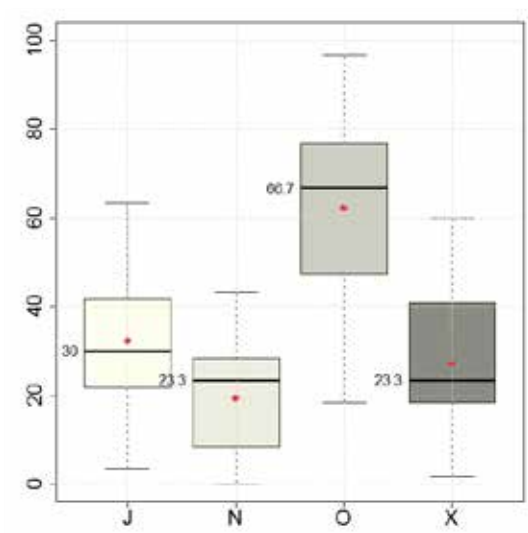
Le plan du texte restitué par les sujets nous a permis de mesurer 3 indicateurs, déjà présentés pour l'expérimentation précédente (voir 3.3.a). Un quatrième indicateur a été ici mesuré en plus, présenté ici dans la dernière sous-section : la structure restituée par les participants relativement à la structure du document, indépendamment du contenu (ex. « ce document est divisé en 4 parties, la première partie comprend 2 sections... »),.

##### 4.4.b.1. Identification et rétention des thèmes du plan

Caractéristiques de l'indicateur : voir 3.3.a.1.

Le graphique ci-dessous présente les médianes et quartiles des 4 sous-ensembles de l'échantillon correspondant aux 4 conditions expérimentales (les médianes  $ME$  sont représentées par un trait noir horizontal et les moyennes  $M$  sont représentées par un point rouge).

Figure 47 - Scores de rétention des thèmes du plan



On observe graphiquement une forte proximité (absence de différences) entre les distributions des échantillons J, N et X en ce qui concerne le nombre de thèmes correctement rappelés ; ces scores sont assez faibles (respectivement  $ME_J = 30$ ,  $ME_N = 23.3$  et  $ME_X = 23.3$ ), 100 correspondant au score maximal. L'échantillon de la condition O est lui distribué de manière largement supérieure avec un score médian dans le tiers supérieur des scores ( $ME_O = 66.7$ ).

Sur le plan inférentiel, le test de Kruskal-Wallis de comparaison des sommes des rangs des observations nous montre une différence significative (domination stochastique) entre ces 4 distributions,  $H(3) = 23.86$ ,  $p < 0.01$ . Pour savoir de manière plus précise quelles distributions sont considérées différentes entre elles, un test de comparaisons multiples a été utilisé (test non-paramétrique sur les rangs des observations<sup>120</sup>). Ce test détermine une valeur critique à partir d'un seuil de significativité demandé (ici  $p = 0.05$ ) et de la taille des échantillons, cette valeur correspondant au seuil de différence entre les moyennes de rangs des échantillons à partir duquel la différence est considérée comme significative, d'après la valeur de  $P$  spécifiée. Ici, le test donne un seuil de différence critique à 16.82, qui est dépassé pour les comparaisons entre (O, J), (O, N), et (O, X).

Comparaison	Différence observée	Seuil de différence significative	Résultat
J-N	10.83	16.82	non
J-O	18.57	16.82	oui
J-X	4.93	16.82	non
N-O	29.4	16.82	oui
N-X	5.9	16.82	non
O-X	23.5	16.82	oui

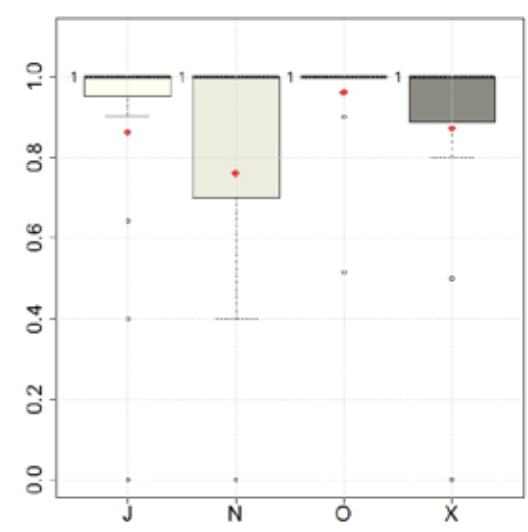
Tableau 22 – Résultats du test Post-Hoc de Kruskal-Wallis : thèmes restitués

On en conclut que dans la condition optimale « O », les thèmes ont été mieux restitués que dans chacune des 3 autres conditions, et que pour chacune des 3 autres conditions la restitution des thèmes était équivalente.

#### 4.4.b.2. Ordre des thèmes dans le plan rappelé

Caractéristiques de l'indicateur : voir 3.3.a.2.

Figure 48 - Ordre des thèmes restitués relativement à l'ordre d'origine



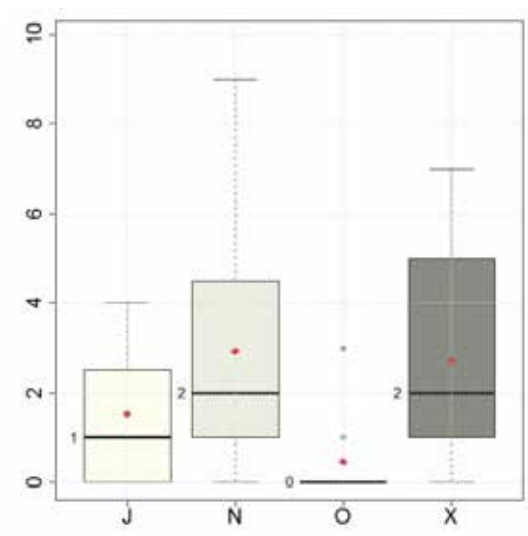
On observe que les distributions sont quasiment identiques entre les conditions, à l'exception d'un plus grand écart interquartiles pour la condition N. Cette absence de différence est en effet confirmée par un test de Kruskal-Wallis de comparaison des sommes des rangs des observations  $H(3) = 2.44$ ,  $p > 0.05$ . Cela confirme que la condition n'a pas influencé l'ordre des titres restitués.

<sup>120</sup> Il s'agit de la fonction [kruskalmc](#) utilisée avec le logiciel R

## 4.4.b.3. Distance entre le plan rappelé et le plan d'origine

Caractéristiques de l'indicateur : voir 3.3.a.3.

Figure 49 –Distributions des distances au plan en fonction de la condition expérimentale



On observe que globalement les distances au plan sont faibles, malgré une certaine variabilité. On notera que dans la condition O pour tous les sujets, à l'exception de 2, les titres restitués l'étaient avec les bons niveaux hiérarchiques.

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions,  $H(3) = 13.66$ ,  $p < 0.01$ . Un test de comparaisons multiples montre que l'échantillon O présente une distribution différente de celle de la condition N et celle de la condition X.

Comparaison	Différence observée	Seuil de différence significative	Résultat
J-N	8.5	16.82	non
J-O	11.3	16.82	non
J-X	8.0	16.82	non
N-O	19.9	16.82	oui
N-X	0.5	16.82	non
O-X	19.3	16.82	oui

Tableau 23 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée

On en conclut que les titres du plan restitués l'étaient avec une meilleure précision quant à leur niveau hiérarchique dans la condition O que dans les conditions N et X. La restitution des niveaux hiérarchiques des titres était équivalente pour ces deux dernières conditions. Concernant la condition J, même si elle se situe graphiquement entre O et les deux conditions N et X, aucune différence statistique n'a pu être trouvée entre celle-ci et les 3 autres.

## 4.4.b.4. Similitude entre la structure du plan restitué et la structure du plan d'origine

Caractéristiques de l'indicateur :

- § *Calcul* : Différence absolue et pondérée entre nombre de parties (facteur 2), sous parties (facteur 1) et rubriques (facteur 0.5) restituées par le sujet et celles du document ; le score est ensuite ramené sur 100 en considérant que la différence maximale correspond à une absence de restitution de la structure. Un score de 0 correspond donc à une absence de

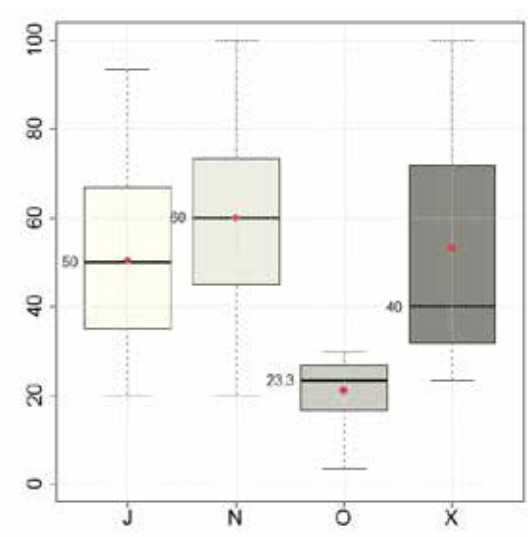


différence entre la structure restituée et la structure du document, et un score de 100 à une différence considérée comme maximale. Exemple : pour un document dont le plan ne contiendrait que 2 parties principale, si un sujet restitue 3 parties, son score de structure serait de :  $|2-3| * 2 = 2$  (ramené ici sur 100)

- § *Étendue* : [0;100], intervalle continu
- § *Mesure* : correspondance entre la structure hiérarchique restituée et la structure hiérarchique du document (indépendamment du contenu).
- § *Commentaire* : cet indicateur a été rajouté pour analyser le plan restitué étant donné que les sujets mentionnaient souvent une structure dépourvue de thèmes (ex. « il y a 3 parties ; la première partie comprend 2 sous parties »). Cet indicateur n'est pas mentionné dans la littérature mais permet de rendre compte de la similitude entre la structure « à vide » du plan restitué et celle du plan d'origine.

Normalité des données : Cet échantillon ne suit pas une loi normale (Shapiro-Wilk :  $W = 0.92$ ,  $p < 0.05$ ). Le graphique ci-dessous présente donc les médianes et quartiles des 4 sous-ensembles de l'échantillon correspondant aux 4 conditions expérimentales.

**Figure 50 - Structure hiérarchique restituée par les sujets**



On observe graphiquement que les distributions des échantillons J, N et X sont assez similaires avec de plus des médianes proches (respectivement 50, 60 et 40). La condition O se démarque clairement avec des valeurs bien inférieures aux 3 autres conditions (les écarts interquartiles de l'échantillon O ne recourent aucun des 3 autres,  $ME_O = 23.3$ ), et donc une structure restituée plus proche de celle du document. Dans les autres conditions, la restitution de la structure était globalement moyenne avec des scores médians autour de 50 ( $ME_J = 50$ ,  $ME_N = 60$  et  $ME_X = 40$ ).

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions,  $H(3) = 25.12$ ,  $p < 0.01$ . Un test de comparaisons multiples montre que l'échantillon O présente une distribution différente de celle des 3 autres conditions.

Comparaison	Différence observée	Seuil de différence significative	Résultat
J-N	5.9	16.82	non
J-O	23.3	16.82	oui
J-X	0.6	16.82	non

N-O	29.2	16.82	oui
N-X	5.3	16.82	non
O-X	23.9	16.82	oui

Tableau 24 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée

On en conclut que la structure hiérarchique du plan a été retenue de manière plus précise en condition O que dans les 3 autres conditions, et assez proche d'une restitution parfaite. Dans les autres cas, la restitution de la structure était moyenne pour les 3 autres conditions.

#### 4.4.c. Effet de la stratégie de restitution sur les scores de compréhension du texte

##### Caractéristiques de l'indicateur :

- § *Calcul* : évaluation des réponses des sujets aux 8 questions de compréhension ; le score était ensuite ramené sur 100. Pour chaque réponse, on attendait un certain nombre d'idées ou de concepts issus du document ; pour chaque idée on attribuait une note : 0 (non-sens ou non-restituée), 1 (réponse exacte littérale ou équivalente sémantiquement) ou 0.5 (réponse approximative ou incomplète faisant tout de même sens).
- § *Étendue* : [0,100], intervalle continu.
- § *Mesure* : compréhension du contenu du texte.

Normalité des données : Cet échantillon suit une loi normale (Shapiro-Wilk :  $W = 0.97$ ,  $p > 0.05$ ), ce qui nous est confirmé par une inspection graphique de l'homogénéité de la distribution des quantiles sur la droite de Henry (quantiles de la loi normale centrée réduite, figure de droite).

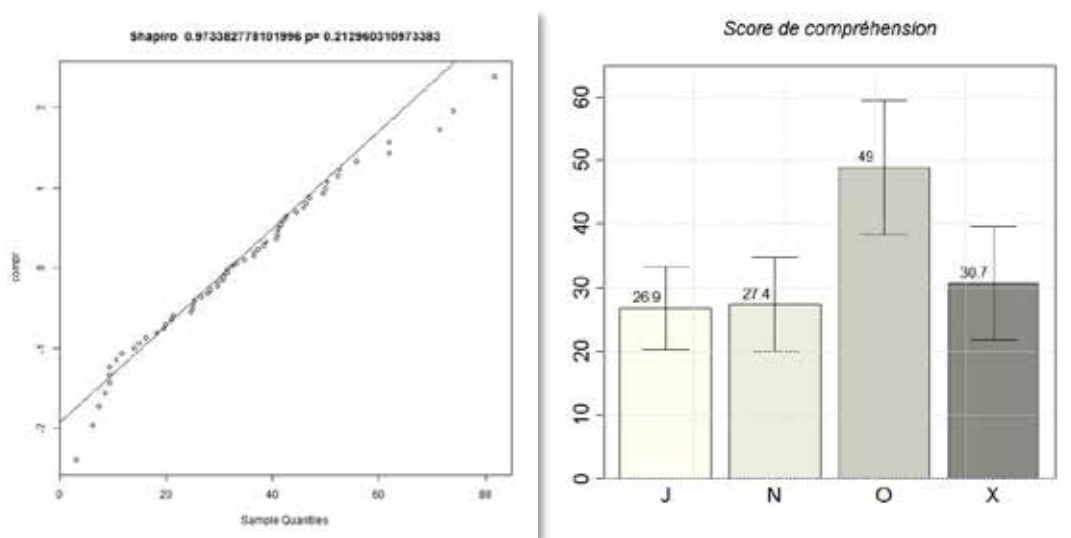


Figure 51 – Scores de compréhension ; (gauche) droite de Henry vs Quantiles ; (droite) moyennes et intervalles de confiance

On observe graphiquement une forte proximité entre les moyennes des échantillons J, N et X et un important recouvrement des intervalles de confiance à 95% des moyennes ; ces scores sont assez faibles (respectivement  $M_J = 26.9$  et  $SD_J = 11.9$ ,  $M_N = 27.4$  et  $SD_N = 13.4$ ,  $M_X = 30.7$  et  $SD_X = 16.2$ ). La moyenne de l'échantillon de la condition O est largement supérieure aux autres ( $M_O = 49$  et  $SD_O = 19.2$ ), tout en restant moyenne par rapport à un score maximal de 100.

Une ANOVA à 1 facteur nous confirme l'existence de différences entre les moyennes des 4 échantillons :  $F(3,56) = 6.88$ ,  $p < 0.01$ ,  $\eta^2 = 0.27$ . Une analyse post-hoc à l'aide du test Tukey-HSD nous

montre que la moyenne de l'échantillon O et des 3 autres (paire à paire) diffère de manière significative.

Comparaison	Différences entre moyennes	Borne inférieure (intervalle de confiance)	Borne supérieure (intervalle de confiance)	P
N-J	0.52	-14.41	15.46	0.99
O-J	22.1	7.16	37.03	0.001
X-J	3.8	-11.10	18.78	0.90
O-N	21.58	6.64	36.42	0.002
X-N	3.3	-11.62	18.26	0.94
X-O	-18.26	-33.2	-3.3	0.01

Tableau 25 – Résultats du test Tukey-HSD (ANOVA post-hoc) : scores de compréhension

On en conclut que dans la condition optimale « O », le contenu du document est mieux compris que dans chacune des 3 autres conditions, et que pour chacune des 3 autres conditions la compréhension ne différait pas significativement.

#### 4.4.d. Effet de la stratégie de restitution sur la charge cognitive

Cette partie présente les résultats des mesures de charge cognitive ; étaient mesurés la charge cognitive globale et ses 3 composantes : la charge cognitive intrinsèque, extrinsèque et celle liée aux processus d'apprentissage (voir 2 pour plus de détails).

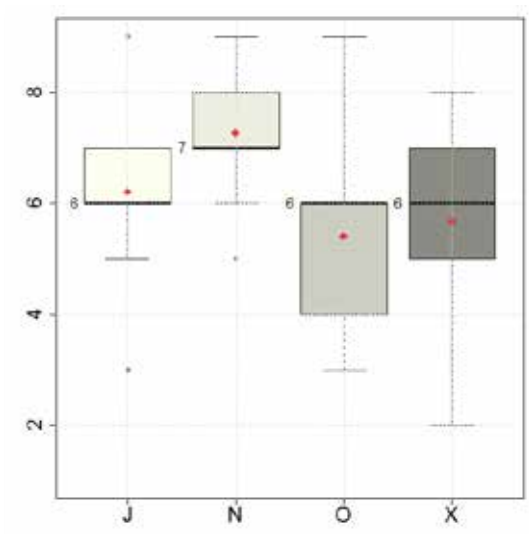
##### 4.4.d.1. Charge cognitive globale

On analyse ici les résultats des réponses à la question : *De manière générale, La compréhension du document et de sa structure m'a demandé un effort mental ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important)*. Cette question était posée immédiatement après la réalisation de la tâche d'écoute par le sujet.

Caractéristiques de l'indicateur :

- § Calcul : aucun ; report des réponses des sujets.
- § Étendue : {1, 9}, intervalle d'entiers.
- § Mesure : charge cognitive globale.

Figure 52 - Scores des réponses à la question 1 sur la charge mentale globale



On observe graphiquement que les médianes des échantillons sont quasiment identiques ( $ME_J = 6$ ,  $ME_N = 7$ ,  $ME_O = 6$  et  $ME_X = 6$ ). La tâche demandée était donc globalement « moyennement exigeante » cognitivement parlant, sans être trop difficile (le milieu de l'échelle est à 5 et correspond à une difficulté moyenne). Même si les médianes sont proches, les distributions des échantillons ne sont pas identiques.

En effet, le test de Kruskal-Wallis de comparaison des sommes des rangs des observations nous montre une différence significative (domination stochastique) entre ces 4 distributions,  $H(3) = 12.51$ ,  $p < 0.01$ . Un test post-hoc de comparaisons multiples a été utilisé pour observer les différences entre chaque paire de conditions. Ici le test donne un seuil de différence critique à 16.82, qui est dépassé pour les comparaisons entre (O, N) et (N, X).

Comparaison	Différence observée	Seuil de différence significative	Résultat
J-N	12.7	16.82	non
J-O	8.0	16.82	non
J-X	4.4	16.82	non
N-O	20.7	16.82	oui
N-X	17.1	16.82	oui
O-X	3.5	16.82	non

**Tableau 26 - Résultats du test Post-Hoc de Kruskal-Wallis : charge cognitive globale**

On en conclut que la tâche a été globalement jugée plus exigeante cognitivement parlant dans la condition N que dans la condition O et la condition X. Dans les autres cas il n'y a pas de différence significative entre la charge cognitive globale induite par les différentes conditions expérimentales.

#### 4.4.d.2. Charge intrinsèque liée au contenu

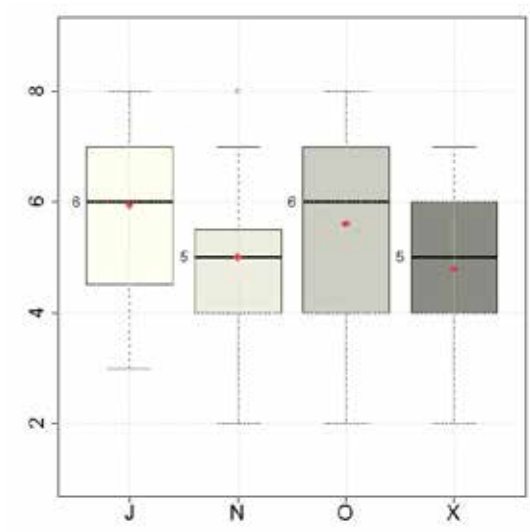
On analyse ici les résultats des réponses à la question : *Comment évalueriez-vous la difficulté de compréhension du contenu ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important)*.

##### Caractéristiques de l'indicateur :

- § *Calcul* : aucun ; report des réponses des sujets.
- § *Étendue* : {1, 9}, intervalle d'entiers.
- § *Mesure* : on mesure l'effort mental induit par la difficulté de compréhension du contenu uniquement. Dans les termes de la théorie de la charge cognitive, ce qu'on mesure ici est donc une partie de la *charge intrinsèque* car elle est propre à la tâche (ici une tâche de compréhension), c'est-à-dire qu'elle est censée ne dépendre que des caractéristiques du contenu à apprendre et des connaissances et capacités des sujets.

Normalité des données : Cet échantillon ne suit pas une loi normale (Shapiro-Wilk :  $W = 0.93$ ,  $p < 0.05$ ). Le graphique ci-dessous présente donc les médianes et quartiles des 4 sous-ensembles de l'échantillon correspondant aux 4 conditions expérimentales.

**Figure 53 - Charge cognitive liée au contenu**



On observe graphiquement que les médianes de tous les échantillons sont proches ( $ME_J = 6$ ,  $ME_N = 5$ ,  $ME_O = 6$  et  $ME_X = 5$ ) et que leurs distributions sont très semblables. Ces scores montrent que le contenu était moyennement difficile (5 correspondant à une difficulté de compréhension du contenu dite « moyenne »).

Cette absence de différence est en effet confirmée par un test de Kruskal-Wallis de comparaison des sommes des rangs des observations ( $H(3) = 4.19$ ,  $p > 0.05$ ). Cela confirme que la modalité de présentation du contenu (la condition) n'a pas influencé la difficulté ressentie par rapport au contenu, ce qui est d'ailleurs en cohérence avec la théorie de la charge cognitive qui stipule que la présentation du contenu influe sur la charge extrinsèque. Cela est également cohérent avec le fait que la distribution de sujets sur les conditions donne des répartitions équivalentes en termes de connaissances antérieures (cf. 4.4.a.2).

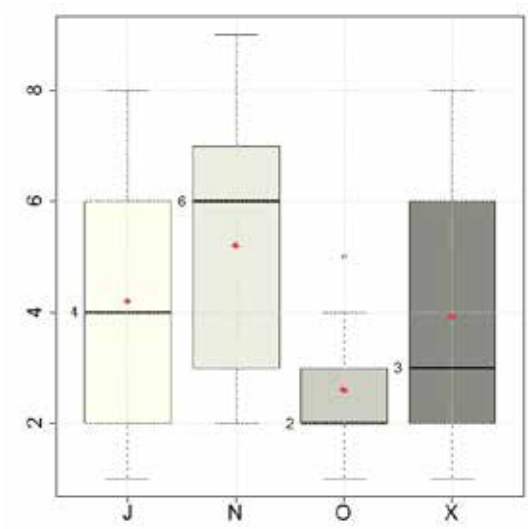
#### 4.4.d.3. Charge cognitive liée à l'intelligibilité de la synthèse vocale

On analyse ici les résultats des réponses à la question : *Comprendre la façon dont la synthèse vocale prononçait les mots du texte (son intelligibilité) m'a demandé un effort mental ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important).*

##### Caractéristiques de l'indicateur :

- § *Calcul* : aucun ; report des réponses des sujets.
- § *Étendue* : {1, 9}, intervalle d'entiers.
- § *Mesure* : on mesure donc l'effort mental engagé pour comprendre la voix de synthèse : son intelligibilité, et la difficulté que pouvait induire la compréhension d'un texte lu. Dans les termes de la théorie de la charge cognitive, ce qu'on mesure ici est donc une partie de la *charge extrinsèque* car elle dépend de la modalité d'accès au contenu.

Figure 54 - Charge cognitive induite par la compréhension de la voix de synthèse



On observe graphiquement que les distributions des échantillons sont très hétérogènes, avec une grande variabilité visible à travers les différences entre les 1<sup>er</sup> et 3<sup>ème</sup> quartiles des distributions. Les médianes des échantillons varient de ( $ME_O = 2$ ) à ( $ME_N = 6$ ). La grande dispersion des données ne permet pas graphiquement plus de précisions, mise à part le fait que la condition « O » ait une répartition sensiblement plus regoupée et faible en valeurs que les autres.

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions,  $H(3) = 8.70$ ,  $p < 0.05$ . Un test de comparaisons multiples montre que seuls les échantillons O et N présentent des distributions différentes (différence observée de 18.47 pour un seuil critique à 16.82,  $p < 0.05$ ). On peut en conclure qu'en condition optimale (O), les sujets estimaient la voix de synthèse plus intelligible qu'en condition sans restitution de la MFM (N).

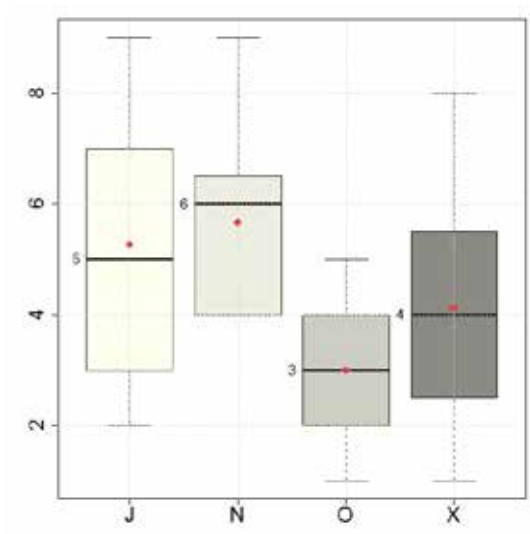
#### 4.4.d.4. Charge extrinsèque liée à la présentation du contenu

On analyse ici les résultats des réponses à la question : *Comprendre la façon dont le texte était présenté à l'oral m'a demandé un effort mental ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important).*

##### Caractéristiques de l'indicateur :

- § *Calcul* : aucun ; report des réponses des sujets.
- § *Étendue* : {1, 9}, intervalle d'entiers.
- § *Mesure* : on mesure donc l'effort mental engagé pour comprendre et interpréter les modalités de présentation de l'information à l'oral. Dans les termes de la théorie de la charge cognitive, ce qu'on mesure ici est donc une partie de la *charge extrinsèque* car elle dépend de facteurs extérieurs à la tâche, qui sont ici les modalités de présentation orales des stratégies de restitution évaluées.

Figure 55 - Charge cognitive liée à la présentation du contenu



On observe graphiquement que les distributions des échantillons O et X se placent légèrement en dessous des deux autres conditions ( $ME_O = 3$  et  $ME_X = 4$  contre  $ME_J = 5$  et  $ME_N = 6$ ). Les écarts interquartiles importants et le fait qu'ils se recoupent visuellement ne permettent pas d'autres conclusions sur la simple observation des distributions.

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions,  $H(3) = 15.59$ ,  $p < 0.01$ . Un test de comparaisons multiples montre que l'échantillon O présente une distribution différente de celle de J et de celle de N.

Comparaison	Différence observée	Seuil de différence significative	Résultat
J-N	4.3	16.82	non
J-O	18.7	16.82	oui
J-X	8.4	16.82	non
N-O	23.0	16.82	oui
N-X	12.8	16.82	non
O-X	10.3	16.82	non

Tableau 27 - Résultats du test Post-Hoc de Kruskal-Wallis : charge cognitive liée à la présentation du contenu

On en conclut que les conditions de présentation du contenu à l'oral ont demandé moins de ressources mentales en condition O que pour les conditions N et J, dans le cadre de l'évaluation subjective des sujets. Dans les autres cas les ressources mises en jeu étaient équivalentes.

#### 4.4.d.5. Charge cognitive liée à l'apprentissage

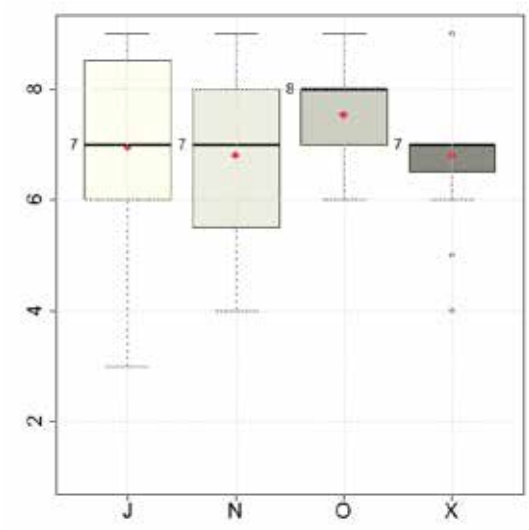
On analyse ici les résultats des réponses à la question : *Jusqu'à quel point étiez-vous concentré pendant la lecture ? (1 à 9, 1 étant très peu concentré et 9 très concentré).*

##### Caractéristiques de l'indicateur :

- § *Calcul* : aucun ; report des réponses des sujets.
- § *Étendue* : {1, 9}, intervalle d'entiers.
- § *Mesure* : charge cognitive liée à l'apprentissage ; dans les termes de la théorie de la charge cognitive on mesure ici la charge « germane », représentant les ressources mentales allouées au traitement et à la génération de schémas mentaux.

Normalité des données : Cet échantillon ne suit pas une loi normale (Shapiro-Wilk :  $W = 0.91$ ,  $p < 0.05$ ). Le graphique ci-dessous présente donc les médianes et quartiles des 4 sous-ensembles de l'échantillon correspondant aux 4 conditions expérimentales.

**Figure 56 - Concentration des sujets pendant la tâche**



On observe graphiquement que les médianes de tous les échantillons sont proches ( $ME_J = 7$ ,  $ME_N = 7$ ,  $ME_O = 8$  et  $ME_X = 7$ ) et que leurs distributions sont très semblables, avec cependant une plus grande homogénéité pour X et O contre une plus grande dispersion pour J et N, relativement aux écarts interquartiles.

Cette absence de différence est en effet confirmée par un test de Kruskal-Wallis de comparaison des sommes des rangs des observations  $H(3) = 2.88$ ,  $p > 0.5$ . Cela confirme que la modalité de présentation du contenu (la condition) n'a pas influencé la concentration des sujets qui devaient effectuer une tâche de compréhension.

#### **4.4.e. Effet de la stratégie de restitution sur l'efficacité**

L'efficacité représente un ratio entre la performance et l'effort mental engagé (les ressources cognitives) pour obtenir cette performance. De nombreuses études montrent l'intérêt de distinguer l'efficacité de l'efficacité (la performance) [Paas et al. 2003]. Ici la charge cognitive globale est mesurée à travers la question présentée ci-dessus. Nous avons fait le choix d'utiliser les scores de compréhension (cf. 4.4.c) dans le calcul de l'indice d'efficacité, car c'est le score reflétant le mieux la performance à notre tâche, étant donné que c'était ce qui était demandé aux sujets, de « comprendre au mieux le contenu du document et sa structure ».

##### Caractéristiques de l'indicateur :

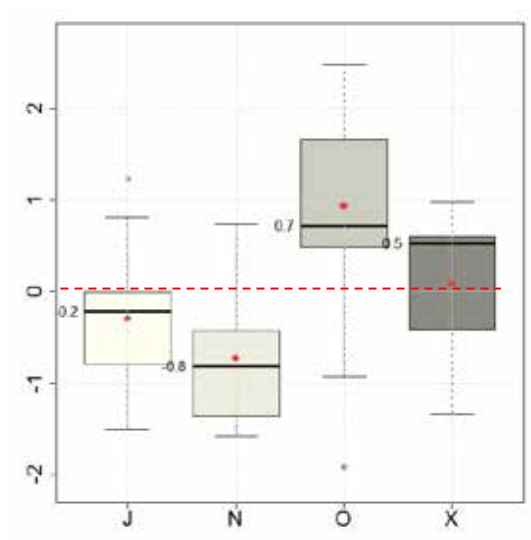
- § *Calcul* : différence entre les scores normalisés (*z-score*) de compréhension (représentant ici la performance à la tâche demandées) et celui de charge cognitive globale, le tout divisé par  $\sqrt{2}$  pour représenter la distance avec la droite d'efficacité nulle (où la performance est équivalente à la charge cognitive et l'efficacité est donc nulle) [Paas et al. 2003].
- § *Étendue* :  $[-X; Y]$ , intervalle continu. Les bornes dépendent de la distribution de l'échantillon étant donné qu'elles sont calculées à partir de sa moyenne et de sa déviation standard (pour les *z-scores*). Ici la distribution des scores de compréhension et de charge cognitive globale



nous donnent une borne inférieure théorique à -2.57 (pour un score de compréhension à 0 (min) et une charge cognitive globale de 9 (max)), et une borne supérieure théorique à 4.86 (compréhension à 100 (max) et charge cognitive à 9 (min)).

§ *Mesure* : distance de l'efficacité par rapport à un score de 0, correspondant à la conjonction d'une charge cognitive et d'un indice de performance tous deux moyens au regard des échantillons étudiés. Pour les valeurs négatives du score, l'efficacité est donc faible (ex. charge plus forte que la performance), et pour les valeurs positives l'efficacité est au contraire forte, toujours relativement à l'échantillon.

Figure 57 - Scores d'efficacités calculés sur la base des scores de compréhension



On observe que l'efficacité est négative entre le 1<sup>er</sup> et le 3<sup>ème</sup> quartile pour les conditions J et N, qu'elle positive pour la condition O entre le 1<sup>er</sup> et le 3<sup>ème</sup> quartile ( $Q1 = 0.48$ ,  $Q3 = 1.66$ ), et que pour la condition X la médiane est positive ( $ME_X = 0.5$ ), mais le 1<sup>er</sup> quartile négatif ( $Q1_X = -0.42$ ). Les scores sont assez proches de zéro globalement (efficacité globale moyenne), et les scores positifs sont assez éloignés du maximum théorique de 4.86. Cela s'explique par le fait que la charge cognitive était au-dessus de la moyenne dans toutes les conditions (cf. 4.4.d.1), montrant ainsi que la tâche peut être qualifiée de moyennement difficile.

Le test Kruskal-Wallis montre cependant l'existence de différences significatives entre la moyenne des rangs de ces distributions  $H(3) = 26.38$ ,  $p < 0.01$ . Un test de comparaisons multiples montre que l'échantillon O présente une distribution différente de celle de la condition N et celle de la condition J.

Comparaison	Différence observée	Seuil de différence significative	Résultat
J-N	14.9	16.82	non
J-O	17.8	16.82	oui
J-X	1.43	16.82	non
N-O	32.7	16.82	oui
N-X	16.3	16.82	non
O-X	16.4	16.82	non

Tableau 28 - Résultats du test Post-Hoc de Kruskal-Wallis : scores d'efficacité basées sur la compréhension

On observe que les comparaisons entre N et X ainsi qu'entre O et X donnent des différences observées très proches du seuil critique (comparativement à la différence observée entre J et X par exemple).

D'après ces résultats, l'efficacité est donc meilleure en condition O qu'en condition N et J, et on observe une forte tendance entre O et X. On observe également une tendance à la domination stochastique de X sur N, et donc une tendance à une efficacité plus importante en condition X qu'en condition N. Les distributions ne présentent aucune différence pour l'efficacité entre les conditions J et X, ni entre les conditions J et N. On en conclut qu'en condition O, une moindre charge mentale a été engagée pour des performances supérieures aux conditions J et N.

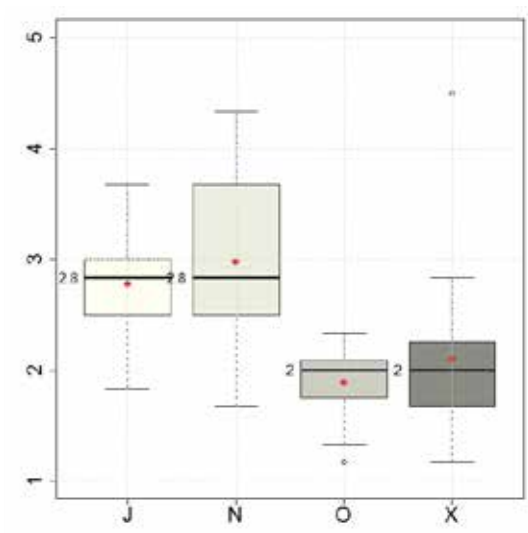
#### 4.4.f. Effet de la stratégie de restitution sur la désorientation

À la suite de la tâche réalisée par les sujets, on leur posait 6 questions mesurant leur sentiment de désorientation durant la tâche. La fiabilité des différentes questions quant à leur capacité à mesurer la même chose (la désorientation), a été évaluée en calculant l'alpha de Cronbach sur l'ensemble des réponses aux 6 questions de désorientation ; ici  $\alpha = 0.73$  ( $\alpha > 0.7$ ) ce qui dénote d'une bonne fiabilité admise [Nunnally 1978].

##### Caractéristiques de l'indicateur :

- § *Calcul* : moyenne des réponses aux 6 questions de désorientation. Un score de 1 correspond à une absence de désorientation des sujets.
- § *Étendue* : [1, 5], intervalle continu.
- § *Mesure* : désorientation des sujets pendant l'accès au document.

Figure 58 - Scores de désorientation



On distingue graphiquement deux groupes de 2 conditions aux distributions similaires. À noter que les médianes au sein de chaque groupe sont égales ( $ME_J = ME_N = 2.8$  et  $ME_O = ME_X = 2$ ). Le premier groupe présente un score moyen de désorientation (l'échelle allant de 1 à 5 le score médian est à 3) avec presque un point de différence avec le second groupe, qui lui présente des scores plus faibles.

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions,  $H(3) = 25.80$ ,  $p < 0.01$ . Un test de comparaisons multiples montre que les 2 groupes de condition (J, N) et (O, X)

sont bien différents quelles que soient les conditions de chacun des groupes comparées deux à deux : les comparaisons entre J et O, et J et X, N et O, et N et X présentent des résultats significatifs.

<i>Comparaison</i>	<i>Différence observée</i>	<i>Seuil de différence significative</i>	<i>Résultat</i>
J-N	1.6	16.82	non
<i>J-O</i>	<i>23.9</i>	<i>16.82</i>	<i>oui</i>
<i>J-X</i>	<i>19.7</i>	<i>16.82</i>	<i>oui</i>
<i>N-O</i>	<i>25.5</i>	<i>16.82</i>	<i>oui</i>
<i>N-X</i>	<i>21.3</i>	<i>16.82</i>	<i>oui</i>
O-X	4.3	16.82	non

**Tableau 29 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée**

On en conclut que la désorientation des sujets était la même en condition J qu'en condition N, et de la même façon elles étaient équivalentes en condition O et en condition X. Ces deux groupes présentent une différence significative de distributions, ce qui se traduit ici par le fait qu'en condition J comme en condition N, la désorientation est plus importante qu'en condition O ou X.

#### ***4.4.g. Effet de la stratégie de restitution sur les tâches de recherche d'information***

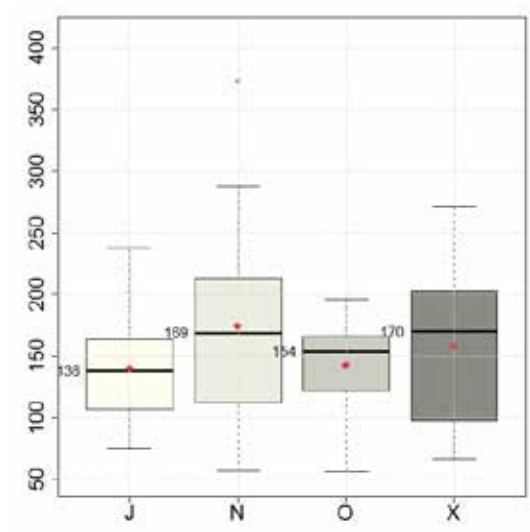
À la suite de la tâche d'écoute réalisée par les sujets, on leur demandait d'effectuer 2 tâches de recherche d'information qui consistaient à retourner le plus rapidement possible à un point précis dans le document, et ce afin d'évaluer leur représentation de la structure du document ainsi que leur capacité à utiliser les repères structuraux du texte et les fonctions du système. Cette mesure de performance permet de compléter la mesure subjective de désorientation, en ce sens qu'elle permet d'évaluer la capacité des sujets à s'orienter dans le document, dans le cadre d'une tâche de recherche d'information.

##### **4.4.g.1. Temps moyens mis pour trouver les emplacements cibles**

###### Caractéristiques de l'indicateur:

- § *Calcul* : somme des 2 durées des 2 tâches de navigation pour chaque sujet. Exemple : si un sujet a mis 100 secondes pour la première tâche puis 43 pour la seconde, on lui affecte alors la valeur 143.
- § *Unité* : en secondes.
- § *Étendue* : {0, N}, intervalle continu.
- § *Mesure* : performance aux tâches de navigation en termes de temps

Figure 59 – Temps moyens cumulés des deux tâches de recherche d'information



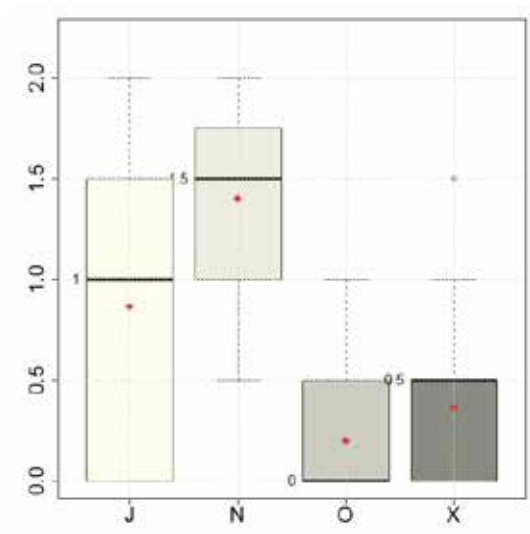
On observe que les distributions sont quasiment identiques entre les conditions, et que les médianes sont proches. Cette absence de différence est en effet confirmée par un test de Kruskal-Wallis de comparaison des sommes des rangs des observations,  $H(3) = 1.70$ ,  $p > 0.5$ . Cela confirme que la condition n'a pas influencé la rapidité avec laquelle les sujets retrouvaient un emplacement dans le document. À noter qu'il n'y a pas non plus de différences significatives entre les temps si l'on compare séparément les conditions pour la tâche 1 et la tâche 2,  $H(3) = 4.03$ ,  $p > 0.5$  et  $H(3) = 2.61$ ,  $p > 0.5$ , respectivement.

#### 4.4.g.2. Réussites de la recherche d'information

##### Caractéristiques de l'indicateur :

- § *Calcul* : Somme de l'analyse des réussites à retrouver l'emplacement pour chacune des deux tâches de navigation, pour chaque sujet. Une note de 0 attribuée à une tâche (pour un sujet donné) correspond à un succès quant à retrouver l'emplacement demandé dans le document, 0.5 à un échec où l'emplacement pointé se trouvait dans la section suivant ou précédant immédiatement l'emplacement cible, et 1 dans les autres cas (échec). Étant donné que cet indicateur est calculé avec la somme des échecs/réussites, un sujet avec un score de 1 signifie qu'il a obtenu 2 échecs « proches » ( $0.5 \times 2$ ) pour les 2 tâches de navigation, ou qu'il a obtenu une réussite et un échec (0 et 1). Plus le score est faible plus les tâches de navigation ont été réussies en termes de proximité des cibles.
- § *Étendue* :  $\{0, 0.5, 1, 1.5, 2\}$ , ensemble discret de valeurs.
- § *Mesure* : Réussite à retrouver les emplacements recherchés en termes de proximité des cibles pointées.

Figure 60 – Réussites aux tâches de recherche d'information



En conditions O et X il y a eu assez peu d'échecs ( $ME_O = 0$  et  $ME_X = 0.5$ ). En condition N, 75% des scores sont au-dessus de 1.0 ( $Q1_N = 1.0$ ), il y a donc eu beaucoup d'échecs. En condition J la distribution est très étendue ( $Q1_J = 0$  et  $Q3_J = 1.5$ ), les résultats sont donc hétérogènes.

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions,  $H(3) = 26.34$ ,  $p < 0.01$ . Un test de comparaisons multiples montre qu'il y a moins d'échecs en condition O qu'en condition N, et moins en condition X qu'en condition N.

Comparaison	Différence observée	Seuil de différence significative	Résultat
J-N	13.2	16.82	non
J-O	15.8	16.82	non
J-X	11.0	16.82	non
N-O	29.0	16.82	oui
N-X	24.2	16.82	oui
O-X	4.9	16.82	non

Tableau 30 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée

On note également que la différence observée entre les échecs à la condition O et J est proche du seuil de significativité. La différence observée entre N et J est également importante, même si en dessous du seuil critique. On en conclut donc que les tâches de recherche en condition N ont donné lieu à plus d'échecs que pour les conditions O et X de manière significative. La condition J tend à montrer plus d'échecs que la condition O (non-significatif).

À noter que les scores d'échecs aux tâches de navigation sont moyennement corrélés avec les scores de désorientation des sujets ( $Pearson\ r(58) = 0.44$ ,  $p < 0.01$ , corrélation linéaire), ce qui semble indiquer que les sujets ayant ici observé un échec aux tâches de navigation se sont également sentis désorientés.

#### 4.4.h. Effet de la stratégie de restitution sur l'utilisabilité subjective

Caractéristiques de l'indicateur :

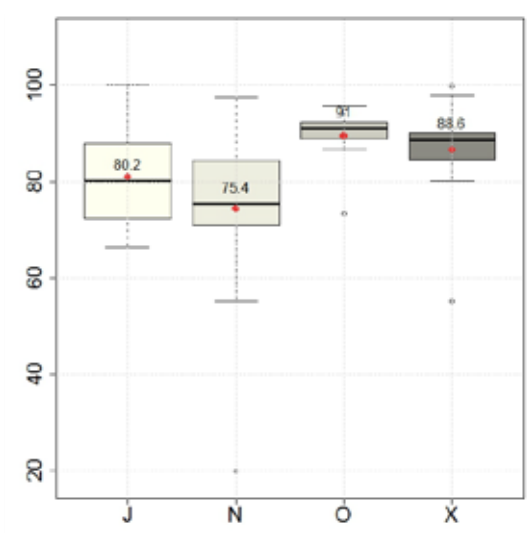
- § Calcul : somme des réponses aux 10 questions d'utilisabilité (chacune sur une échelle de 1 à 5), puis ramené sur 100 comme habituellement pour un SUS [Brooke 1996]. À noter que la

fiabilité des différentes questions quant à mesurer la même chose (l'utilisabilité) est bonne (Cronbach  $\alpha = 0.83$ ,  $\alpha > 0.7$ ).

§ *Étendue* : [0 ; 100], intervalle continu.

§ *Mesure* : utilisabilité subjective du système ; cela couvre l'efficacité, l'efficience et la satisfaction de l'utilisateur quant au système (norme ISO 9241-11). Plus généralement, l'utilisabilité représente la pertinence d'un outil relativement à une tâche et un contexte donnés.

Figure 61 - Scores d'utilisabilité du système en fonction de la condition



L'observation graphique montre des distributions homogènes pour les conditions O et X, et plus dispersées pour les conditions J et N. Globalement les scores sont hauts en valeur, avec un score médian minimum pour la condition N de ( $ME_N = 75.4$ ) et un score médian maximum de ( $ME_O = 91$ ) pour la condition O.

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions  $H(3) = 13.16$ ,  $p < 0.01$ . Un test de comparaisons multiples montre que seuls les échantillons O et N présentent des distributions différentes (différence observée de 20.57 pour un seuil critique à 16.82,  $p < 0.05$ ). On peut donc en conclure que le système de lecture était perçu plus utilisable en condition O qu'en condition N.

La littérature indique [Ahuja and Webster 2001] que l'utilisabilité est habituellement corrélée avec les mesures de désorientation. Ici la corrélation basée sur les rangs des observations nous donne un coefficient de corrélation négatif moyen ( $Pearson r(58) = -0.66$ ,  $p < 0.01$ , corrélation linéaire) entre la désorientation et l'utilisabilité, ce qui est en cohérence avec la littérature. Avoir mesuré l'utilisabilité conjointement avec la désorientation semble donc avoir été pertinent dans notre cas au regard des différences dans les résultats de ces deux indicateurs.

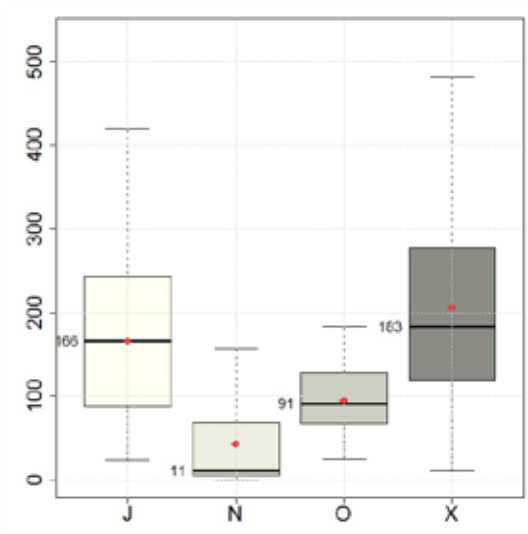
#### 4.4.i. Effet de la stratégie de restitution sur l'utilisation des commandes

Même si l'étude n'avait pas pour but d'étudier les comportements de navigation des utilisateurs, les fichiers « trace » nous ont permis de recueillir des données quantitatives sur l'utilisation des commandes et fonctions mises à disposition pour la navigation dans le document.

## 4.4.i.1. Commandes de navigation

Quelle que soit la condition, les sujets pouvaient se déplacer dans le document à l'aide des touches « flèche » du clavier. On reporte ici le nombre moyen d'utilisation de la commande par participant durant la lecture en fonction de la condition expérimentale. À noter que le déplacement avec ces touches se faisait de bloc en bloc dans les conditions J, X et O et se faisait à la manière d'une navigation dans un flux audio (« retour / avance rapide ») dans la condition sans MFM (N). On ne peut donc pas comparer directement l'utilisation des commandes dans la condition N avec celle des autres conditions, puisque les commandes n'avaient pas la même fonction. La condition N a quand même été incluse sur les graphiques, mais pas dans les tests statistiques.

**Figure 62 – Report quantitatif de l'utilisation des commandes de navigation**



Concernant le nombre total de commandes utilisées pendant la lecture, on observe une grande variabilité dans chacune des conditions. Pour les échantillons les plus hétérogènes (J et X) on observe un écart interquartiles de plus de 150 commandes utilisées : ( $Q3_J - Q1_J = 155$ ) et ( $Q3_X - Q1_X = 159$ ). Les conditions O et N semblent plus homogènes avec globalement moins de commandes de navigation utilisées que dans les deux autres conditions ( $ME_O = 91$  et  $ME_N = 11$  contre  $ME_J = 166$  et  $ME_X = 183$ ) et un écart interquartile plus réduit : ( $Q3_N - Q1_N = 64$ ) et ( $Q3_O - Q1_O = 60$ ).

Quantitativement parlant on observe que l'utilisation des commandes est proche en conditions J et X, qu'il est plus faible en condition O mais toujours important, et finalement beaucoup plus faible en condition N. Concernant ce dernier point, il révèle des déplacements quasi-inexistants en mode N, car en termes de déplacement, il faut utiliser « K » fois une commande de déplacement (ou appuyer pendant « Y » secondes sur la touche, ce qui aurait été inscrit dans le fichier trace comme Y\*5 utilisations de la commande de navigation) pour aller d'un paragraphe à l'autre (K et Y dépendant de la taille du paragraphe). À contrario il suffit d'un seul appui sur la même commande dans les autres conditions pour passer au paragraphe suivant ou précédent. Cela montre donc bien que les sujets ne se déplaçaient presque pas dans le document en condition sans MFM(N).

On exclut donc la condition N pour les tests inférentiels. Le test de Kruskal-Wallis montre une différence significative entre ces trois groupes,  $H(2) = 2.88$ ,  $p < 0.05$ , avec une valeur du  $\chi^2$  cependant faible confirmant l'hétérogénéité des distributions, observée graphiquement. Un test de

comparaisons multiples confirme une différence significative entre les distributions de la condition O et de la condition X (différence observée de 12.70 avec un seuil critique à 12.65). Les commandes sont donc moins utilisées en condition O qu'en condition X, et de manière équivalente entre les conditions X et J. Les résultats ne permettent pas de conclure quant à l'utilisation des commandes entre les conditions O et J.

Durant les tests, nous avons observé une stratégie de lecture fréquente chez les sujets, qui était d'effectuer une première écoute exhaustive pour ensuite naviguer. Ci-dessous nous avons donc étudié l'utilisation des commandes avant d'avoir atteint une première fois la fin du document, et après.

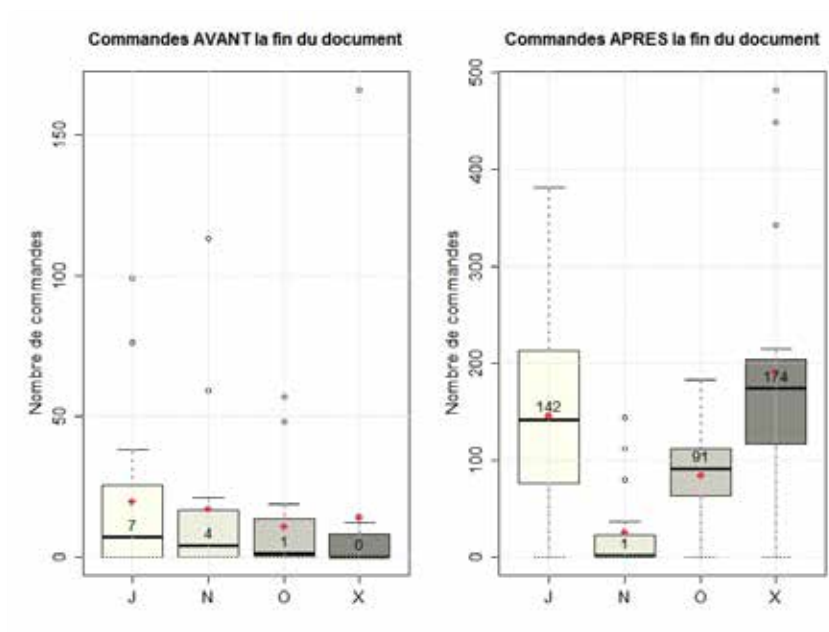


Figure 63 - Utilisation des commandes de navigation avant (à gauche) et après (à droite) la fin du document

L'examen graphique des données confirme que la navigation s'effectue principalement après que les sujets soient arrivés à la fin du document une première fois. En effet la majeure partie des commandes utilisées l'est bien une fois la fin du document atteinte. Cela est confirmé à travers les résultats des tests de Kruskal-Wallis ( $H(1) = 11.15$ ,  $H(1) = 16.67$  et  $H(1) = 16.12$ ,  $p < 0.01$ ), comparant respectivement les échantillons pour les conditions J, O et X, avant et après la fin du document (ex. J avant VS. J après).

Concernant les commandes utilisées avant la fin du document, les distributions ne présentent pas de différence significative. En effet même si un premier test de Kruskal-Wallis donne ( $H(2) = 1.68$ ,  $p < 0.05$ ), la valeur du  $\chi^2$  est faible, et un test de comparaisons multiples ne met en avant aucune différences entre les conditions J, O et X deux à deux. Cela signifie que l'hypothèse *nulle* a été rejetée par le premier test, mais qu'un examen des groupes deux à deux ne permet pas de mettre en avant de différences significatives.

Concernant les commandes utilisées après la fin du document, des différences existent entre les distributions,  $H(2) = 8.46$ ,  $p < 0.05$ . Un test post-hoc de comparaisons multiples montre qu'effectivement X domine O (différence observée de 13.67 pour seuil critique à 12.65).

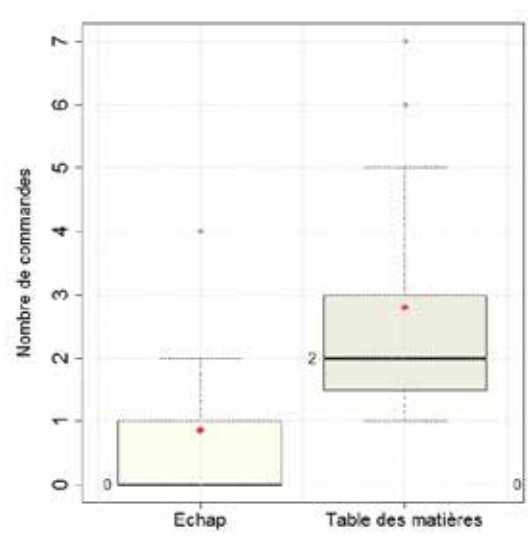


On peut donc conclure que les commandes de navigation étaient en très grande majorité utilisées seulement une fois la fin du document atteinte, et que globalement la condition X a suscité une plus grande utilisation des commandes que la condition O, et une utilisation équivalente à la condition J.

#### 4.4.i.2. Commandes spécifiques au mode optimal

Dans le mode de lecture optimal, les sujets avaient accès à deux fonctions supplémentaires : ECHAP leur permettait de lire un message les situant dans le document, et ENTREE leur permettait d'accéder à la table des matières et de revenir au document (ici seules les appuis sur ENTREE pour passer dans la table des matières ont été comptabilisés).

Figure 64 - Report quantitatif des commandes du mode optimal utilisées



On observe graphiquement que la touche ECHAP a été délaissée par la majorité des sujets ( $ME_{Echap} = 0$ ). La moitié des sujets passaient au moins 2 fois dans la table des matières ( $ME_{Tdm} = 2$ ), et 75% passaient entre 1.5 et 3 fois dans la table des matières pendant la lecture. On observe cependant certaines valeurs extrêmes au regard de la distribution : 1 sujet a utilisé 4 fois la touche ECHAP et 2 autres ont ouvert respectivement 6 et 7 fois la table des matières pendant la lecture.

Concernant la table des matières, ces résultats n'indiquent cependant pas combien de temps les sujets ont passé à parcourir la table des matières. Nous avons récupéré ces temps à l'aide des fichiers traces. Cet échantillon de durées suit une loi normale (Shapiro-Wilk :  $W = 0.94$ ,  $p > 0.05$ ) ; les sujets ont en moyenne parcouru la table des matières pendant ( $M = 146$ ) secondes ( $SD = 79$ ), soit 2min26secs, ce qui représente 13% du temps total imparti pour le parcours du document en condition optimale, et équivaut à un parcours total de la table des matières, durant 2min10. À noter que l'écart-type est très important en comparaison de la moyenne :  $\pm 1\text{min}19\text{secs}$ .

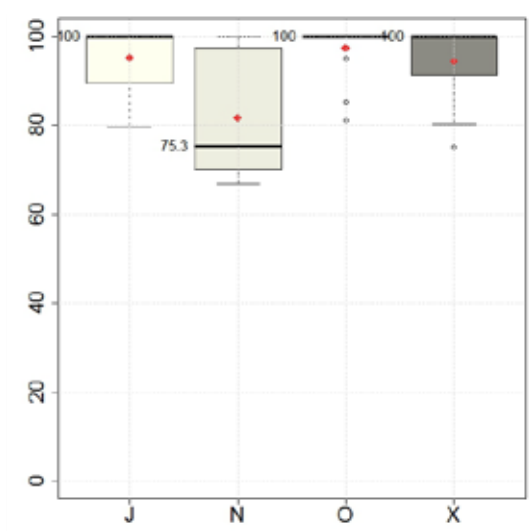
#### 4.4.j. Effet de la stratégie de restitution sur les temps de parcours du document

Caractéristiques de l'indicateur :

- § *Calcul* : pourcentage du temps imparti pour la tâche qui a été utilisé par les sujets. Un score de 80% pour un sujet signifie qu'il a arrêté la lecture en ayant encore 20% du temps imparti disponible.
- § *Étendue* : [0 ; 100], intervalle continu.

§ *Unité* : pourcentage.

**Figure 65 – Pourcentage du temps total imparti utilisé pour le parcours du document**



L'observation graphique montre que pour les conditions J, O et X la majorité des sujets ont utilisé la totalité du temps imparti. En condition O seuls 3 sujets ne l'ont pas fait. La condition N a une distribution plus étendue et une médiane à ( $ME_N = 75.3$ ), contre ( $ME_{J, O, X} = 100$ ) pour les 3 autres conditions, ce qui signifie que 50% des sujets en condition N ont arrêté la lecture une fois 75% du temps imparti écoulé. À noter qu'une lecture exhaustive du document représentait environ 70% du temps total imparti.

Le test Kruskal-Wallis montre bien une différence significative entre ces distributions,  $H(3) = 14.69$ ,  $p < 0.01$ . Un test de comparaisons multiples montre que seuls les groupes O et N présentent des distributions différentes (différence observée de 20.93 pour un seuil critique à 16.82,  $p < 0.05$ ). On peut en conclure qu'en condition sans MFM (N), les sujets ont arrêté la lecture plus rapidement qu'en condition optimale (O).

#### **4.4.k. Retours qualitatifs**

À la fin des tests, il était demandé aux sujets d'exprimer les remarques et commentaires qu'ils pouvaient avoir sur la tâche réalisée et le système utilisé. Les remarques ainsi que leurs fréquences d'apparition dans les retours de sujets sont synthétisées dans le tableau ci-dessous (seules les remarques ayant été faites par au moins 2 sujets apparaissent dans le tableau). Les retours qualitatifs correspondants des sujets sont en Chapitre V4.1.

Le tableau ci-dessous présente les différents thèmes (ou catégories) des remarques des sujets, avec le nombre d'occurrences du thème dans les retours qualitatifs et à chaque fois les conditions passées par les sujets ayant fait ces remarques.

Thème	Nombre d'occurrences	Conditions correspondantes	Exemples
Concentration fluctuante pendant la tâche	N = 56	O, J, N, X	- j'étais moins concentré vers la fin - j'ai décroché au bout d'un moment ça a été dur de rattraper
Problèmes de diction de la voix de synthèse	N = 12	O, J, N, X	- Parfois la synthèse avale des mots. - Moi c'est la voix robotique que j'ai du mal à bien comprendre

Il manquait des fonctions pour parcourir la structure	N = 11	J, N	- Il faudrait une table des matières, surtout pour les gros documents - Le seul truc qui manque c'est l'arbre, ça serait plus simple de se balader
Pertinence de la restitution de la structure et des fonctions associées	N = 11	O, X	- Je trouve la table des matières vraiment bien - Les sons des titres pour me repérer ça m'a aide
Difficulté d'apprendre et de retenir un texte lu	N = 8	O, J, X	- Je n'ai pas l'habitude d'apprendre comme ça. - J'avais du mal à me concentrer, je suis plutôt visuel
Difficulté liée à la lecture de deux numéros consécutifs : le niveau du titre puis sa numérotation (ex. titre de niveau 1 ; 2. [...])	N = 8	O, J, N, X	- l'appellation titre de niveau 1 / niveau 2 suivie de la numérotation des titres est perturbante - Peut-être qu'il serait plus simple d'utiliser le terme « Grand 1 petit A » plutôt que « titre de niveau 1 ; 2 »
Difficulté d'appréhender à la fois la structure et le contenu	N = 6	O, J, X	- Moi je le sentais c'était soit la forme soit le fond (pour la compréhension). - Je me concentrais plus sur la structure que sur le fond...
Icônes audio et son spatialisé occultés	N = 6	O, X	- Les sons des titres je ne les ai pas forcément utilisés - J'ai zappé l'histoire du son qui passe de gauche à droite
La présence d'autres fonctions de navigation aurait été appréciable	N = 6	O, J, N, X	- j'aurais bien aimé d'autres touches [fonctions] pour sauter les pages et des touches pour sauter les chapitres / sous chapitres - j'aurais aimé pouvoir aller directement aux notes de bas de page
Difficulté d'interprétation de certains contenus périphériques et confusion	N = 4	J, N	- Les 1/2/3 qui arrivent en bas de page je ne sais pas à quoi ça correspond. - les notes de bas de page au milieu c'est perturbant
La prise de notes aurait facilité l'apprentissage	N = 3	O, J, X	- j'aurais besoin de pouvoir prendre des notes en parallèle - ça aurait été plus simple si j'avais pu prendre des notes

**Tableau 31 - Synthèse des remarques qualitatives des sujets voyants**

#### 4.4.1. Synthèse des résultats

Ci-dessous un tableau reprenant l'ensemble des résultats présentés précédemment. Les résultats présentent les différences significatives et absences de différences entre les différentes conditions expérimentales pour les variables étudiées.

Quatre condition étaient étudiées : J (Jaws, lecteurs d'écran), O (optimale), X (exhaustive) et N (sans MFM, *NoMFM*). Pour présenter les résultats de manière synthétique, on présente les différences significatives avec des opérateurs mathématiques d'inégalité (« > » et « < »). Par exemple si les résultats de la condition optimale (O) ont été supérieurs à ceux de la condition sans MFM (N), on reportera dans le tableau « O > N » dans la colonne des résultats significatifs. De la même façon, on présente l'absence de différences significative à l'aide de l'opérateur égal, signifiant ici qu'il n'y a pas de différence paire à paire entre les membres de l'égalité, par exemple : O = J = N signifie qu'aucune des 3 paires (O et J, O et N, et J et N) ne diffèrent des autres dans les résultats obtenus.

À noter que pour chaque variable du tableau, un lien vers la partie correspondante des résultats est inséré.

Indicateur étudiée	Différences significatives	Résultats vérifiant les hypothèses ?
<a href="#">Plan : thèmes</a>	O > (J, N, X)	Partiellement
<a href="#">Plan : distance</a>	O < (N, X)	Partiellement
<a href="#">Plan : ordre</a>	-	Non
<a href="#">Plan : différence de structure</a>	O < (J, N, X)	Partiellement
<a href="#">Compréhension</a>	O > (J, N, X)	Partiellement

<u>Charge cognitive globale</u>	$O < N, X < N$	Partiellement
<u>Charge intrinsèque (contenu)</u>	-	Non
<u>Charge intrinsèque (TTS)</u>	$O < N$	Partiellement
<u>Charge extrinsèque (présentation du contenu)</u>	$O < (J, N)$	Partiellement
<u>Charge d'apprentissage</u>	-	-
<u>Efficience</u>	$O > (N, J)$ ; tendances : $O > X$ et $X > N$	Oui
<u>Désorientation</u>	$O < (N, J)$ ; $X < (N, J)$	Oui
<u>Recherche d'informations : temps</u>	-	-
<u>Recherche d'informations : échecs</u>	$O < N$ et $X < N$ ; tendance : $O < J$	-
<u>Utilisabilité</u>	$O > N$	-
<u>Commandes de navigation utilisées</u>	$X > O$	-
<u>Durée de parcours du document</u>	$O > N$	-

Tableau 32 - Synthèse des résultats : expérimentation sur sujets voyants

## 4.5. Discussion

Nous avons testé 4 stratégies différentes de restitution de la mise en forme, chacune présentant un degré de restitution plus ou moins important, une restitution plus ou moins complète. Une stratégie ne restituait aucune mise en forme et les participants accédaient au texte à la manière d'un fichier audio, une autre reproduisait la restitution orale proposée traditionnellement par les lecteurs d'écran, une autre restituait exhaustivement tous les indices de mise en forme du texte et une dernière stratégie (dite « optimale ») reprenait la même oralisation que la stratégie exhaustive, en proposant en plus un accès global à la structure du texte avec table des matières, et un filtrage du contenu non-pertinent. Chaque sujet abordait le même document avec l'une des 4 stratégies mentionnées, avec pour objectif d'essayer de comprendre au mieux le contenu du document et sa structure ; ils devaient ensuite répondre à un questionnaire de compréhension, évaluer la charge cognitive induite par la tâche ainsi que la désorientation ressentie pendant cette dernière.

Nos résultats montrent globalement la supériorité de la condition optimale sur les 3 autres, quel que soit l'indicateur mesuré : cette stratégie est donc significativement plus efficace à presque tous les niveaux dans le cadre d'une tâche de compréhension. En revanche les 3 autres stratégies ont montré des performances équivalentes, ce qui indique que la restitution orale type « lecteur d'écran » n'était pas plus efficace que la condition « sans mise en forme », la plus dépouillée, et que restituer l'ensemble des phénomènes locaux de mise en forme (condition « exhaustive ») ne donne de meilleurs résultats sauf sur certains indicateurs (la charge cognitive et désorientation, discutés ci-après). Ce qui caractérise la stratégie optimale était la présence de 2 fonctions de lecture

supplémentaires associées à des raccourcis clavier. Cependant, sur ces deux fonctions seule la table des matières semble avoir été utilisée par les sujets (voir 4.4.i.2), et ainsi semble être le facteur auquel les résultats peuvent être imputés. Ce besoin de pouvoir accéder à la structure du texte s'est aussi exprimé dans les retours qualitatifs des sujets, qui lorsque la stratégie de restitution était dépourvue de ce type de fonction, les sujets suggéraient d'en rajouter (N = 11). Certains sujets en condition optimale, ayant accès à la table des matières, suggéraient même d'implémenter d'autres fonctions de navigation dans la structure (N = 6). Ces résultats soulignent donc l'importance de prendre en compte les objectifs et caractéristiques de la tâche, qui ici pour une tâche de compréhension impliquaient la construction d'une représentation globale de l'organisation thématique du texte, et au niveau de la restitution de la MFM montrent l'intérêt d'une approche fonctionnelle de sa restitution, c'est-à-dire la restitution des fonctions de la MFM en lecture visuelle.

Cependant, même si la restitution de la structure globale semble jouer un rôle prépondérant, restituer la MFM localement semble alléger la charge cognitive pour des performances de compréhension équivalentes à une absence de restitution. Ces résultats seront développés dans les paragraphes suivants.

Ainsi, lorsque le contenu en lui-même est cohérent comme c'était le cas ici, restituer les phénomènes locaux de MFM (condition exhaustive) n'améliore pas les performances comparativement à une restitution minimale de la MFM (stratégie « sans mise en forme ») ; on parle bien ici de restitution « minimale » (et pas d'absence de restitution) malgré le nom de la stratégie correspondante, car on pourrait arguer que la numérotation des titres dans le texte est un indice de MFM qui n'a pas été supprimé de la restitution pour cette stratégie. [Lorch et al. 2012] ont d'ailleurs montré que la numérotation était un indice suffisant pour retenir et organiser les thèmes du plan à l'oral. On peut donc en conclure que lorsque le texte est cohérent, des indices minimaux de MFM (ici, des indices sur un objet toutefois essentiel : les titres) suffisent au lecteur pour mettre en œuvre des inférences sur la structure thématique aussi bonnes que lorsque l'ensemble des indices de MFM est restitué (condition exhaustive).

Cependant, ne pas restituer la MFM a un coût cognitif puisque la nature et la structure des composantes du texte doivent être inférées car elles ne sont pas rendues explicites ; c'est d'ailleurs ce que l'on peut observer pour les résultats de charge cognitive globale et extrinsèque qui augmentent lorsque la MFM n'est pas restituée (conditions « lecteur d'écran » et « sans mise en forme »). On peut donc conclure que la restitution locale de la MFM permet de limiter les inférences et que la restitution de la structure globale a le même rôle mais permet en plus d'améliorer la compréhension : d'où les scores d'efficacité obtenus. On peut d'ailleurs formuler l'hypothèse qu'il existe à l'oral un « seuil » (au sens conceptuel) de volume d'informations de MFM restituées au-delà duquel restituer la MFM diminue les performances, car on a plus de chance de restituer des informations non-pertinentes ou redondantes dont on sait qu'elles sont un facteur de surcharge cognitive [Giraud et al. 2011]. De plus, la linéarité de la modalité orale implique que chaque information restituée devra être traitée par l'auditeur, d'où une hausse naturelle de la charge cognitive dans ces conditions. Ainsi, pour optimiser à la fois les performances et les ressources cognitives de manière générale lors de la restitution orale de la MFM nous pensons qu'il est important d'adopter une approche globale, « holistique » de restitution, en prenant en compte la tâche, ses objectifs, la modalité et les objets textuels mis en forme à restituer.

À propos de la prise en compte la tâche dans la restitution, il est bon de noter que la tâche de compréhension proposée aux sujets était d'une difficulté moyenne ; pour une tâche plus simple, une trop grande facilitation du traitement du contenu (à travers par exemple une restitution exhaustive de la MFM) aurait pu conduire à un désengagement des apprenants de la tâche et donc pénaliser les performances [McNamara et al. 1996]. Cette difficulté de la tâche se vérifie dans notre cas avec les résultats de la charge cognitive liée à l'apprentissage avec un effort important quelle que soit la condition, les scores de compréhensions globalement moyens par rapport au score maximal et les faibles connaissances antérieures des sujets.

Un autre aspect de la restitution de la mise en forme est que l'apprentissage pourrait jouer un rôle dans leur impact sur la compréhension. Même si on ne restituait ici que les phénomènes de MFM pertinents pour la tâche (conditions optimale et exhaustive), en utilisant des indices non-verbaux on met en place une certaine grammaire que les sujets devaient interpréter en parallèle de leur activité de compréhension. En clair, on rajoutait donc des traitements supplémentaires, même si ces traitements donnaient des informations pertinentes. Ainsi l'automatisation de ces traitements cognitifs développée avec l'expertise [Rasmussen 1983] permettrait peut-être d'amplifier les différences observées entre les conditions ; cette hypothèse serait à vérifier en réalisant plusieurs sessions de test avec les mêmes participants sur le protocole utilisé. Plus généralement, tout comme la lecture visuelle nécessite un important apprentissage pour utiliser efficacement les marques de MFM, il semble raisonnable qu'un équivalent oral demande également une certaine expertise pour être utilisé efficacement.

Le résultat principal concernant la restitution des phénomènes locaux de MFM (conditions optimale et exhaustive) réside dans leur rôle dans la diminution de la désorientation (*hypothèse 3*) : les indices associés permettraient donc d'offrir des repères de navigation aux sujets, qui ressentent moins de désorientation et tendent à obtenir plus de réussites aux tâches de recherche d'information. On observe d'ailleurs que les commandes de navigation ont été autant utilisées avec la stratégie exhaustive qu'avec la stratégie lecteur d'écran, ce qui confirme qu'à déplacements équivalents, ce sont les indices locaux de mise en forme qui ont permis aux sujets de se sentir moins désorientés pendant la navigation. Nous pensons que le son spatialisé a pu jouer un rôle de repère local pendant la lecture en indiquant la taille des paragraphes et que les icônes audio ont pu jouer un rôle de repères pour la navigation, à la manière de ce qui est appelé les *landmarks*<sup>121</sup> dans l'étude de la cognition spatiale [Werner et al. 1997]. Cependant de nouvelles études devront être réalisées afin de préciser le rôle de chacune des modalités de restitution utilisées pour ces tests. Enfin, il sera intéressant de voir si ces modalités de restitution pourront être utilisées efficacement à des vitesses de lecture bien supérieures comme celles utilisées par certains déficients visuels [Asakawa and Takagi 2003] ; là-encore de nouvelles études devront être menées. Cependant l'aspect non-verbal pourrait jouer en faveur de ces modalités de restitution à des vitesses élevées.

Un résultat plus général et utile pour la recherche est que la mesure de la désorientation semble être un indicateur important dans le cadre de l'accès non-visuel aux documents. Premièrement car il

---

<sup>121</sup> Cette théorie stipule que l'orientation spatiale se base sur 3 niveaux de représentations : les repères visuels (*landmarks*), les chemins reliant ces repères (*route*) et des représentations intégrant les deux à d'autres informations de différentes natures (*surveys*). Il sera d'ailleurs intéressant pour de futures études d'étudier l'apport de travaux existants sur la cognition spatiale dans la navigation hypertexte [Brouwers and others 2015] à la cas de l'accès non-visuel aux documents

mesure un ensemble de paramètres inhérents à la navigation, et particulièrement pertinents dans ce cas précis de l'oral, où la construction de la représentation de l'espace de navigation se fait au fil de la lecture (pour plus détails voir 2). De plus, cet indicateur a donné ici des résultats différents de l'utilisabilité (qui place uniquement la condition optimale au-dessus des autres), en faisant un indicateur complémentaire. On peut également noter le fait que les scores de désorientation étaient moyennement corrélés avec les réussites des tâches de recherche d'information, mais pas avec les temps mesurés pour ces mêmes tâches ; ainsi pour des tâches de recherche d'information, la désorientation peut être encore une fois un indicateur complémentaire à la performance, voir explicatif.

Enfin, un dernier résultat important concerne l'étude des stratégies de navigation dans le document. Ces stratégies n'ont jamais été étudiées à notre connaissance dans le cas d'une tâche de compréhension, contrairement au cas de la recherche d'information sur le web [Borodin et al. 2010; Vigo and Harper 2013]. Or ici la grande majorité des sujets ont adopté une même stratégie (ce qui est confirmé par les résultats sur l'utilisation des commandes de navigation) qui consistait à faire une première lecture exhaustive avec très peu de déplacements, et une fois la fin du document atteinte débutait une phase de navigation, souvent après un retour au début du document. De nouvelles études devraient être menées sur ces stratégies, et ce afin de proposer une lecture adaptée du document dans le cadre d'une tâche de compréhension. Il est à noter que cette stratégie a également été observée chez des sujets non-voyants explorant des pages web inconnues [Francisco-Revilla and Crow 2010] ; on peut donc en conclure que certains résultats de la littérature concernant les stratégies de navigation pourraient être applicables à une tâche de compréhension qui est en effet par nature une activité complexe pouvant comporter des phases exploratoires ou de recherche d'information.

Revenons en maintenant aux hypothèses formulées pour cette expérimentation (voir section 4.1). Concernant les hypothèses portant sur le « quoi », l'ensemble des informations restituée, la première hypothèse a été partiellement vérifiée comme évoqué précédemment : restituer les informations de mise en forme permet de tendre vers un allègement de la charge cognitive en limitant les inférences, mais n'a pas eu ici d'effet sur la compréhension. L'autre hypothèse (*hypothèse 2*) qui portait sur la suppression des contenus non-pertinents pour l'apprentissage n'a pas été vérifiée ; on observe en effet que quelle que soit la stratégie de restitution utilisée, filtrant ou non ces contenus, la charge cognitive intrinsèque liée au contenu reste inchangée. Or la littérature montre que dans le cas d'une activité d'utilisation d'un site web (basé sur des tâches de recherche d'information), filtrer les contenus non-pertinents et/ou redondants allège la charge cognitive [Giraud 2014]. Même si ce dernier auteur ne précise pas le volume de contenu filtré dans ses expérimentations ni sa répartition dans la page, étant donné que ces travaux portaient sur des pages web le filtrage de regroupements comme les menus, les bandeaux publicitaires et les encarts divers en dehors du contenu principal, on peut supposer que le volume de contenu filtré était important. Il est donc possible que la quantité de contenus filtrés ici ait été trop peu importante pour observer un effet sur la charge cognitive, ou qu'ils aient été trop dispersés dans le texte.

Concernant les hypothèses portant sur le « comment » restituer l'information de MFM (*hypothèse 3* et *hypothèse 4*), les modalités non-verbales utilisées ont effectivement permis de diminuer la désorientation et tendent à diminuer la charge cognitive (stratégies optimale et exhaustive) ; l'hypothèse 3 a donc été partiellement validée, et montre que les modalités de restitution éprouvées

étaient ici pertinentes. La restitution locale de la MFM a également permis aux sujets de montrer de meilleures performances dans les tâches de recherche d'information. Ensuite, la restitution de la structure globale dans la stratégie optimale a elle permis de nettement améliorer la compréhension et de diminuer la charge cognitive, de la même façon qu'en lecture visuelle [Meyer et al. 1980; Lemarié et al. 2012]. Cependant elle n'a pas joué un rôle de diminution de la désorientation des sujets ; l'hypothèse 4 donc été partiellement validée également. À propos ce dernier point, il est possible que les déplacements non-linéaires (des liens) induits par la table des matières ait pu être désorientant. Ainsi, si la table des matières a permis aux de sujets de pouvoir plus facilement aller là où ils voulaient, un potentiel bénéfique en termes d'orientation a pu être compensé par le fait que la navigation non-linéaire rendait plus difficile la localisation absolue du curseur de lecture dans le document (le « où suis-je »), les deux se seraient alors compensés dans les scores de désorientation, ne montrant pas de différence entre les stratégies optimale et exhaustive.

Enfin, la dernière hypothèse portant sur le « quand » restituer l'information de mise en forme, qui s'est traduite ici par une lecture de la table des matières avant l'accès au document en lui-même, ne peut être confirmée ou infirmée. En effet le fait que la condition optimale ait amélioré la compréhension et diminué la charge cognitive peut être imputé seulement en partie au fait que la table des matières ait été parcourue avant la lecture du contenu, étant donné que l'utilisation de la table des matières pendant la lecture représentant 13% du temps total de lecture avec un écart type de 7%. Là encore de nouvelles études devront être menées pour confirmer le rôle de cette présentation préliminaire de la structure globale ; cependant les travaux en psycholinguistique ayant montré l'utilité pour l'apprentissage des *advance-organizers* (voir 4.1.c), nous pensons qu'il est important de prendre en compte ces résultats dans le cadre d'une approche holistique de la restitution de la MFM.

L'ensemble de ces résultats ont une implication directe sur la conception de systèmes de lecture non-visuelle de documents. La restitution locale de la MFM permet d'explicitier le contenu du texte, et ainsi contribuer à réduire la charge cognitive ; cependant dans le cas d'un texte cohérent cette restitution n'améliore pas les performances de compréhension. Elle a en revanche permis aux utilisateurs de mieux s'orienter lors de la navigation, probablement grâce aux indices non-verbaux utilisés pour restituer la MFM, tels que des icônes audio ou du son spatialisé. Quant à la restitution (fonctionnelle) de la structure globale, elle est capitale et doit être accessible à travers plusieurs fonctions pour permettre une bonne compréhension du contenu et pour réduire la charge cognitive. De plus, en ce qui concerne les technologies actuelles, nos résultats montrent que la stratégie de restitution correspondant à l'oralisation faite par les lecteurs d'écrans observe des performances équivalentes à la lecture d'un fichier audio, ce qui montre que ces solutions peuvent et doivent être améliorées. Plus généralement, ces résultats montrent l'importance d'adopter une démarche globale prenant en compte à la fois les utilisateurs, la tâche et la nature et les propriétés de l'information à restituer. L'expérimentation présentée dans la partie suivante permettra de voir dans quelle mesure ces résultats peuvent être validés auprès d'une population de sujets déficients visuels.



## 5. Évaluation de 2 stratégies de restitution de la MFM auprès sujets non-voyants

---

Cette section présente l'expérimentation ayant évalué 2 stratégies de restitution de la mise en forme des documents auprès de 12 sujets déficients visuels. Les résultats n'ont pas encore été publiés. Les 2 stratégies évaluées ici correspondent aux stratégies « optimale » (O) et lecteur d'écran (J) déjà évaluées avec des sujets non-déficients (voir 4).

### 5.1. Présentation

---

De la même façon que nous avons évalué différentes stratégies de restitution de la mise en forme avec des sujets non-déficients, nous souhaitons évaluer l'ensemble de ces mêmes stratégies avec des volontaires déficients visuels, afin de pouvoir recueillir des données avec les utilisateurs finaux des interfaces non-visuelles, mais aussi pour comparer les résultats des deux expérimentations. Ce souhait initial s'est heurté à des considérations pratiques ; en effet il est extrêmement compliqué de trouver 60 volontaires déficients visuels pour passer ces tests : nous avons donc adapté le protocole expérimental. Initialement, nous avons testé 4 stratégies de restitution en inter-sujets auprès de 60 volontaires non-déficients, soit 15 volontaires par condition de restitution. Ici nous avons testé 2 des 4 conditions initiales, en intra-sujets pour compenser le faible nombre de sujets, avec 12 volontaires déficients visuels ; nous avons limité les tests à 2 conditions par sujet étant donné que cela portait déjà la durée du test à 1h30. Ces différences entre les deux protocoles expérimentaux empêcheront une comparaison statistique directe des résultats ; nous le ferons cependant sur le plan descriptif.

Les deux conditions expérimentales choisies sont la condition optimale (O) et la condition lecteur d'écran (J). Nous donc avons choisi de comparer la condition qui à priori offrait la meilleure restitution de la mise en forme avec la condition qui était la plus représentative des outils utilisés par les déficients pour accéder aux documents. La conception des stratégies de restitution est décrite auparavant (voir 4.2.c).

Les hypothèses expérimentales sont les mêmes que pour l'expérimentation avec les sujets non-déficients (voir 4.1). Elles sont succinctement reprises et commentées ci-dessous :

1. *Restituer les informations de MFM du document permet d'améliorer la compréhension et d'alléger la charge cognitive.* On comparera ici deux degrés de restitution à travers les conditions J et O.
2. *Supprimer les informations non-pertinentes et/ou redondantes allège la charge cognitive.* Les contenus périphériques sont présents en condition J et supprimés de la lecture en condition O.
3. *Restituer « passivement » les informations locales de MFM, en utilisant des sons non-verbaux et les propriétés des voix de synthèse, permet d'alléger la charge cognitive et de diminuer la désorientation.* Cette restitution est incluse dans la condition optimale. Cependant il ne sera pas possible de dissocier les effets de cette restitution passive des effets de la restitution de la structure globale (voir hypothèses 4 et 5 ci-dessous)
4. *Restituer activement la structure globale du texte (i.e. le plan) réduit la charge cognitive et la désorientation, et améliore également la compréhension.* Testée à travers la condition optimale.

5. *Prendre connaissance de la structure globale du texte avant d'accéder au contenu en lui-même améliore la compréhension et réduit la charge cognitive.* Testée à travers la condition optimale.

Le protocole expérimental était identique à celui utilisé pour les non-déficients (voir 4.2.d), à l'exception près qu'étant donné que les passations se faisaient en intra-sujets (chaque participant est exposé successivement aux 2 stratégies de restitution), le protocole était répété deux fois pour chaque sujet, pour que chacun passe les deux conditions expérimentales. Chaque condition était testée avec un document différent. L'ordre des conditions était contrebalancé entre les sujets, et les deux documents étaient assignés à chaque condition de sorte que chaque document soit accédé 12 fois avec chaque stratégie, et que chaque sujet passe les deux conditions avec ces deux documents. Le tableau ci-dessous montre comment ont été contrebalancés l'ordre de passation des conditions, des documents, et la répartition des deux conditions sur les deux documents (les paires « document + condition »).

Sujet	Texte lu en premier	Condition de restitution	Texte lu en second	Condition de restitution
1	Économie	Optimale	Europe	Jaws
2		Jaws		Optimale
3	Europe	Optimale	Économie	Jaws
4		Jaws		Optimale

Tableau 33 – Contrebalancement des conditions et documents : expérimentation avec sujets déficients visuels

## 5.2. Méthode

### 5.2.a. Participants

Les 2 stratégies de restitution ont été évaluées auprès de 12 sujets déficients visuels. Sur ces 12 sujets, 7 étaient de sexe masculin et 5 de sexe féminin, d'un âge moyen de 45 ans ( $SD = 15$  ans). Leur participation était volontaire mais gratifiée d'un chèque cadeau d'une grande enseigne.

La très grande majorité ( $N = 10$ ) utilisait un ordinateur avec un lecteur d'écran et des voix de synthèse tous les jours ; 1 sujet préférait utiliser sa tablette tactile, et 1 autre utilisait un bloc note braille<sup>122</sup>. Ces deux derniers sujets utilisaient ces appareils également avec des lecteurs d'écran incluant des voix de synthèse, et ce sur des bases quotidiennes.

Presque tous les participants ( $N = 11$ ) présentaient une cécité totale. Parmi ces 11 sujets, 9 étaient déficients de naissance ou précoce (avant 6 ans), et 2 autres étaient déficients visuels depuis l'adolescence. Un sujet présentait une cécité partielle (1/50<sup>ème</sup>) de naissance (ce niveau de déficience est considéré comme une *cécité* par l'OMS<sup>123</sup>).

### 5.2.b. Matériel

Le matériel était identique à celui utilisé pour l'expérimentation avec les sujets non-déficients (voir 4.2.b). La différence résidait ici dans le fait qu'on utilisait un second document en plus du document traitant d'économie utilisé pour l'expérimentation précédente, qui traitait de l'Union Européenne dans la mondialisation.

<sup>122</sup> Pour plus de détails sur ce qu'est un bloc-notes braille, voir Chapitre II.5.3.

<sup>123</sup> <http://apps.who.int/classifications/icd10/browse/2010/en#/H54>

Pour éviter un éventuel biais des documents sur les résultats, il fallait que les deux documents traitent de sujets différents mais soient équivalents en termes de complexité, de mise en forme et de structure. Le second document traitant de l'Europe dans la mondialisation a donc été conçu selon les mêmes critères de conception que le document traitant d'économie (voir 4.2.b). Il présentait le même nombre de parties et sous-parties, la même quantité d'informations périphériques, le même nombre de phénomènes de mises en forme et la même taille totale en nombre de mots (à 92 mots près).

### **5.2.c. Stratégies de restitution évaluées**

---

Les deux stratégies de restitution testées étaient la stratégie optimale (O) et la stratégie lecteur d'écran (J) telles qu'elles ont été présentées pour l'expérimentation précédente (voir 4.2.c).

### **5.2.d. Procédure de test**

---

La procédure de test était identique à celle de l'expérimentation précédente (voir 4.2.d) excepté que l'ensemble des étapes de la procédure étaient répétées deux fois avec les deux documents présentés précédemment. Un nouveau questionnaire de compréhension a cependant été défini pour le document traitant de la mondialisation et de l'Union Européenne (voir 4.4 ci-dessous). Chaque test durait entre 65 et 95 minutes.

## **5.3. Résultats**

---

Cette partie présente les résultats obtenus avec 12 sujets déficients visuels. On commencera par présenter les caractéristiques de notre population. Par la suite, les résultats des différents indicateurs seront chacun présentés dans une section dédiée ; les échantillons des 2 conditions expérimentales y seront à chaque fois comparés (*rappel des codes des conditions* : O = condition optimale, J = condition lecteur d'écran (JAWS)). Seront également présentés les éventuels effets de l'ordre et du document sur l'indicateur en question, c'est-à-dire les éventuels biais d'ordre de la mesure (couple [document, condition] passé en premier ou en second), et les éventuels biais liés au document (*rappel des codes des documents* : ECO = document traitant d'économie, UE = document traitant de l'Union-Européenne dans la mondialisation). On rappelle que l'on utilisait 2 documents différents pour ces tests, chaque sujet ayant dû passer les 2 conditions successivement avec à chaque fois un des deux documents, d'où les variables catégorielles : document (UE ou ECO), ordre (1 ou 2) et condition (O ou J).

La présentation des résultats pour chaque indicateur se fera systématiquement de la manière suivante :

- § Rappels concernant l'indicateur : calcul, étendue (valeurs prises par l'indicateur), mesure (ce qu'il représente)
- § Normalité des données : voir 3.3.
- § Recherche d'un éventuel effet de l'ordre et/ou du document sur l'indicateur. Pour chacune des deux variables catégorielles (ordre et document) : test statistique de différence des moyennes (si données normales) ou des rangs (si données non normales)
- § Présentation graphique de l'effet de la stratégie sur l'indicateur : voir 3.3.
- § Commentaires sur les graphiques : voir 3.3..

- § Tests statistiques et résultats de ces tests. À noter que les échantillons sont ici appariés, car chaque sujet passait successivement les 2 conditions. *Ces tests ont été réalisés à l'aide du logiciel R ; les fonctions utilisées correspondant aux tests statistiques sont présentées en Chapitre V4.3.*
- Données normales : test de *Student* sur données appariées pour comparer les moyennes des échantillons. Les résultats sont reportés de la manière suivante :  $t(N_{\text{degrés de liberté}}) = T\_value, p < \text{ou} > \text{seuil de significativité}$ . À noter que les résultats sont significatif quand la *T\_value* dépasse un seuil critique (en valeur absolue) de 1.96 pour un seuil de significativité à 5% (valeur reportée depuis une distribution normale standard pour  $y = 0.05$ ).
  - Données non-normales : test de Wilcoxon sur données appariées pour comparer les rangs des échantillons. Les résultats sont reportés de la manière suivante :  $Z = Z\_value, p < \text{ou} > \text{seuil de significativité}$ . *Z* est la valeur normalisée (*z-value*) de la statistique donnée par le test, habituellement appelée *V* pour les échantillons appariés (et *W* pour les non-appariés). À noter que *V* donne ici la somme des rangs des paires pour lesquelles la valeur du premier échantillon est supérieure à la valeur du second (les graphiques reprennent les échantillons dans cet ordre-là). *Z* est la valeur normalisée de *V*. Pour un seuil de significativité à 5%, le seuil critique de *Z* en valeur absolue est de 1.96 (valeur reportée depuis une distribution normale standard pour  $y = 0.05$ ).

Enfin, l'ensemble des résultats est synthétisé dans un tableau récapitulatif (voir 5.3.I).

### ***5.3.a. Connaissances antérieures des sujets***

---

Pour évaluer les connaissances antérieures des sujets relatives aux domaines abordés par les documents, on leur posait 4 questions avant l'écoute de chaque document, et une à posteriori :

- § Auto-évaluation des connaissances du domaine (1 à 5)
- § Fréquence de la pratique du domaine pendant la formation (1 à 5)
- § Fréquence de la pratique du domaine dans l'activité professionnelle (1 à 5)
- § Fréquence de l'auto-formation du domaine (renseignements à titre personnel) (1 à 5)
- § Après l'écoute du document, on demandait ensuite au sujet d'évaluer de nouveau leurs connaissances par rapport à ce qu'ils avaient entendu (1 à 5)

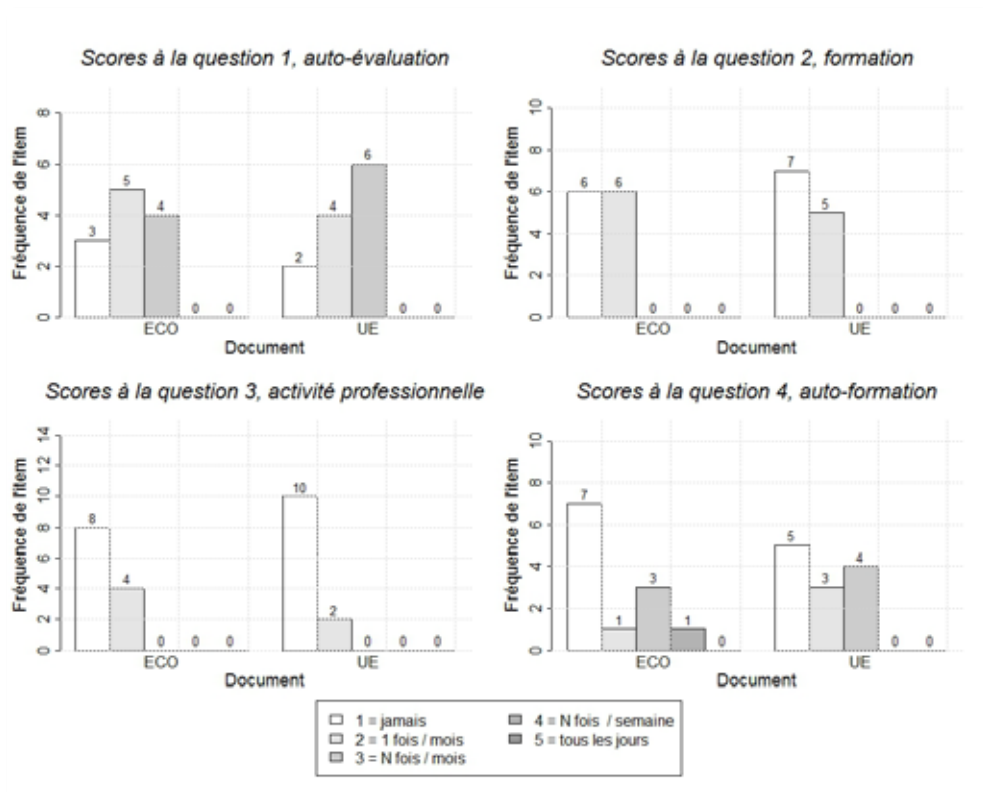


Figure 66 - Connaissances antérieures des sujets déficients visuels

On remarque que globalement les sujets présentaient des connaissances faibles à moyennes quel que soit le document (sur l'ensemble des questions  $ME = 1.7$ ,  $SD = 0.8$ ). On teste l'homogénéité de ces distributions entre les documents (entre ECO et UE) pour chacune des 4 questions à l'aide d'un test du  $\chi^2$ .

	Question 1	Question 2	Question 3	Question 4
Résultats - $\chi^2$	(2, N=24) = 0.71, p = 0.7	(1, N=24) = 0, p = 1	(1, N=24) = 0.22, p = 0.64	(3, N=24) = 2.48, p = 0.48
Effet du document				

Tableau 34- Résultats du test du  $\chi^2$  : connaissances antérieures des sujets déficients visuels, effet du document

Les tests du  $\chi^2$  ne montrent aucune différence significative ( $p > .05$ ) entre les distributions de chaque variable entre les 2 documents, et ce pour chacune des 4 questions. On en conclut que les sujets avaient un niveau de connaissances équivalent d'un document à l'autre : aucun des deux textes n'était plus familier aux sujets que l'autre.

### 5.3.b. Effet de la stratégie de restitution sur la rétention du plan du texte

À la suite de l'écoute de chaque document, il était demandé aux sujets de restituer le plan, c'est-à-dire l'ensemble des titres ou thèmes du texte, leur imbrication et leur ordonnancement. Cette restitution nous a permis de mesurer 4 indicateurs dont les résultats sont reportés dans 4 les sections ci-après :

1. Les thèmes des titres restitués relativement aux thèmes définis par les titres du document d'origine
2. La distance au plan, c'est-à-dire la différence de niveaux hiérarchiques entre les titres restitués et les titres correspondants du document

3. L'ordre des titres restitués, c'est-à-dire la différence entre l'ordonnancement des titres restitués et l'ordonnancement des titres du document
4. La structure restituée par les participants relativement à la structure du document, indépendamment du contenu (ex. « ce document est divisé en 4 parties, la première partie comprend 2 sections... »)

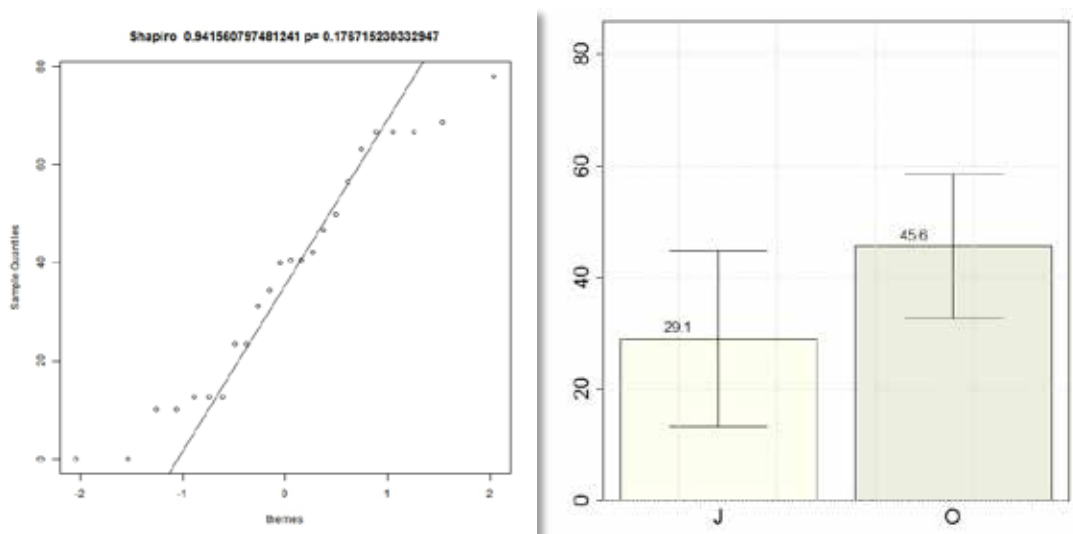
Les 3 premiers indicateurs ont déjà utilisé en psycholinguistique pour mesurer la rétention du plan [Lorch et al. 2012].

### 5.3.b.1. Identification et rétention des thèmes du plan

Caractéristiques de l'indicateur : voir 3.3.a.1.

Normalité des données : La distribution de cet échantillon suit une loi normale (Shapiro-Wilk :  $W = 0.93$ ,  $p > 0.05$ ) ; l'inspection graphique confirme la proximité des quantiles avec la droite de Henry, bien que les scores extrêmes s'en éloignent quelque peu (voir Figure 67 ci-dessous).

Effets de l'ordre et du document : Ni l'ordre (*Student paired t-test* :  $t(11) = 0.09$ ,  $p > 0.05$ ) ni le document (*Student paired t-test* :  $t(11) = 0.57$ ,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les moyennes et intervalles de confiance des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.



**Figure 67 - Rétention des thèmes du plan chez les sujets déficients visuels; (gauche) quantiles et droit de Henry ; (droite) moyennes et intervalles de confiance**

En condition J la rétention des thèmes se trouve en moyenne dans le tiers inférieur des scores ( $M_J = 29.1$ ), tandis qu'en condition O les scores s'approche d'un score moyen ( $M_O = 45.6$ ) relativement à un score maximal de 100. Sur le plan inférentiel, un test de Student à mesures répétées montre une différence significative entre ces 2 échantillons ( $t(11) = -3.01$ ,  $p < 0.05$ ).

Dans la condition optimale « O », les thèmes ont été mieux restitués qu'en condition J.

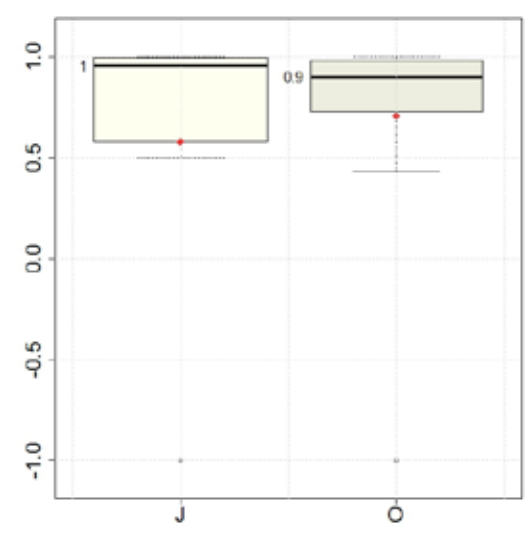
### 5.3.b.2. Ordre des thèmes dans le plan rappelé

---

Caractéristiques de l'indicateur : voir 3.3.a.2.

Effets de l'ordre et du document : Ni l'ordre (*Wilcoxon paired*  $Z = -0.09$ ,  $p > 0.05$ ) ni le document (*Wilcoxon paired*  $Z = 0.44$ ,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les médianes et distributions des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.

**Figure 68 - Ordre des thèmes restitués (sujets déficients visuels)**



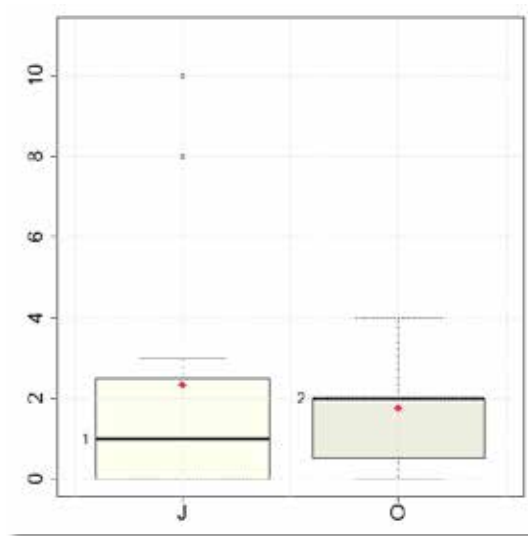
On observe que les 2 distributions sont très similaires ( $ME_J = 1$  et  $ME_O = 0.9$ ) et que les scores sont proches d'une restitution parfaite de l'ordre. Un test de *Wilcoxon* sur données appariées nous confirme l'absence de différence entre ces échantillons ( $Z = 0.36$ ,  $p > 0.05$ ).

### 5.3.b.3. Distance entre le plan rappelé et le plan d'origine

---

Caractéristiques de l'indicateur : voir 3.3.a.3.

Effets de l'ordre et du document : Ni l'ordre (*Wilcoxon paired*  $Z = 0.31$ ,  $p > 0.05$ ) ni le document (*Wilcoxon paired*  $Z = 0.71$ ,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les médianes et distributions des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.

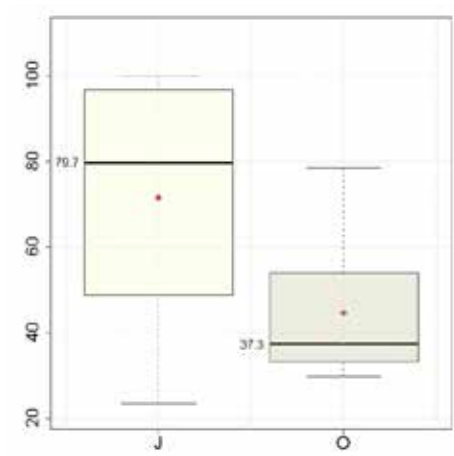
**Figure 69 - Distance au plan (sujets déficients visuels)**

On observe graphiquement que dans les deux conditions, la distance au plan est faible ( $ME_J = 1$  et  $ME_O = 2$ ), et donc que les titres restitués l'ont été globalement avec le bon niveau hiérarchique. Les distributions des échantillons sont assez similaires. Sur le plan inférentiel, un test de *Wilcoxon* (comparant les rangs moyens) sur données appariées nous confirme l'absence de différence entre ces échantillons ( $Z = -0.1$ ,  $p > 0.05$ ).

#### 5.3.b.1. Similitude entre la structure du plan restitué et la structure du plan d'origine

Caractéristiques de l'indicateur : voir 4.4.b.4.

Effets de l'ordre et du document : Ni l'ordre (*Wilcoxon*  $Z = -1.33$ ,  $p > 0.05$ ) ni le document (*Wilcoxon*  $Z = -0.47$ ,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les médianes et distributions des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.

**Figure 70 - Structure restituée (sujets déficients visuels)**

On observe que même si les distributions se recoupent légèrement, la condition O montre des valeurs bien inférieures à la condition J ( $ME_J = 79.7$  et  $ME_O = 37.3$ ). En condition J la structure différait



fortement de la structure d'origine du document. Un test de *Wilcoxon* sur données appariées nous confirme que ces échantillons ont des distributions différentes ( $Z = 2.59, p < 0.01$ ).

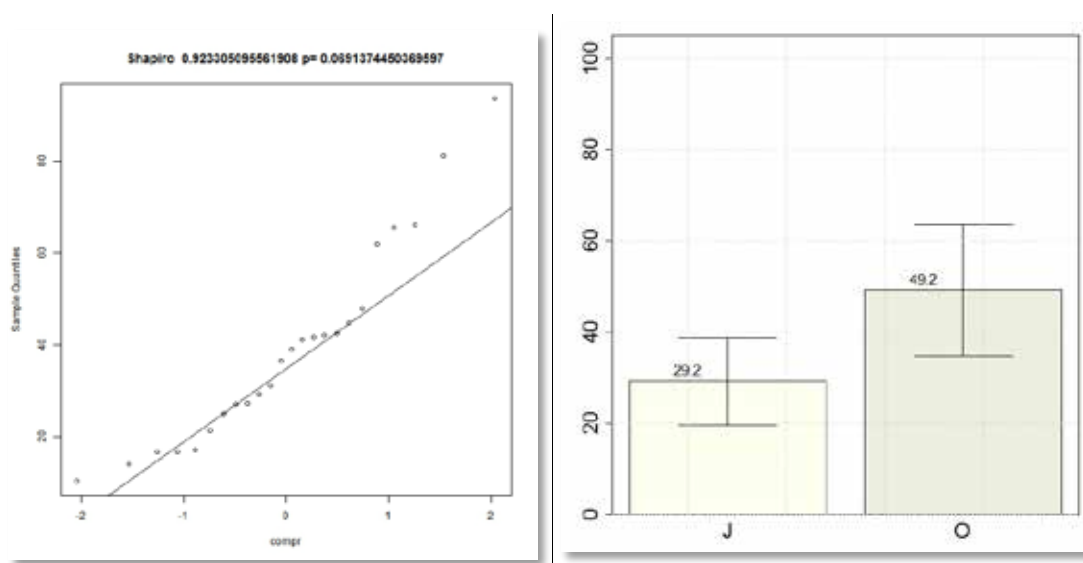
La structure est donc mieux restituée en condition O qu'en condition J, avec environ un facteur 2. Étant donné que nous avons écarté les effets de l'ordre et du document sur l'indicateur, on en conclut que la condition O améliore l'exactitude de la structure retenue par les sujets.

### 5.3.c. Effet de la stratégie de restitution sur les scores de compréhension du texte

Caractéristiques de l'indicateur : voir 4.4.c.

Normalité des données : Cet échantillon suit bien une loi normale (Shapiro-Wilk :  $W = 0.92, p > 0.05$ ) ; l'inspection graphique confirme la proximité des quantiles avec la droite de Henry, bien que les scores extrêmes s'en éloignent quelque peu (voir ci-dessous).

Effets de l'ordre et du document : Ni l'ordre (*Student paired*  $t(11) = -0.51, p > 0.05$ ) ni le document (*Student paired*  $t(11) = -0.3, p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les moyennes et intervalles de confiance des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.



**Figure 71 – Scores de compréhension (sujets déficients visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance des scores**

En condition J, le score moyen de compréhension est dans le tiers inférieur des scores ( $M_J = 29.2$ ), tandis qu'en condition O les scores s'approchent d'un score moyen ( $M_O = 49.2$ ) relativement à un score maximal de 100. De plus, on observe que les intervalles de confiance ne se recoupent presque pas, ce qui est généralement un bon indicateur graphique, d'autant plus que les échantillons sont appariés. Un test de *Student* sur données appariées confirme une différence significative entre ces 2 échantillons ( $t(11) = -3.18, p < 0.01$ ).

On en conclut que dans la condition O, le contenu a été mieux compris qu'en condition J. Étant donné que nous avons écarté les effets de l'ordre et du document sur l'indicateur, on en conclut que la condition O améliore la compréhension du texte.

### 5.3.d. Effet de la stratégie de restitution sur la charge cognitive

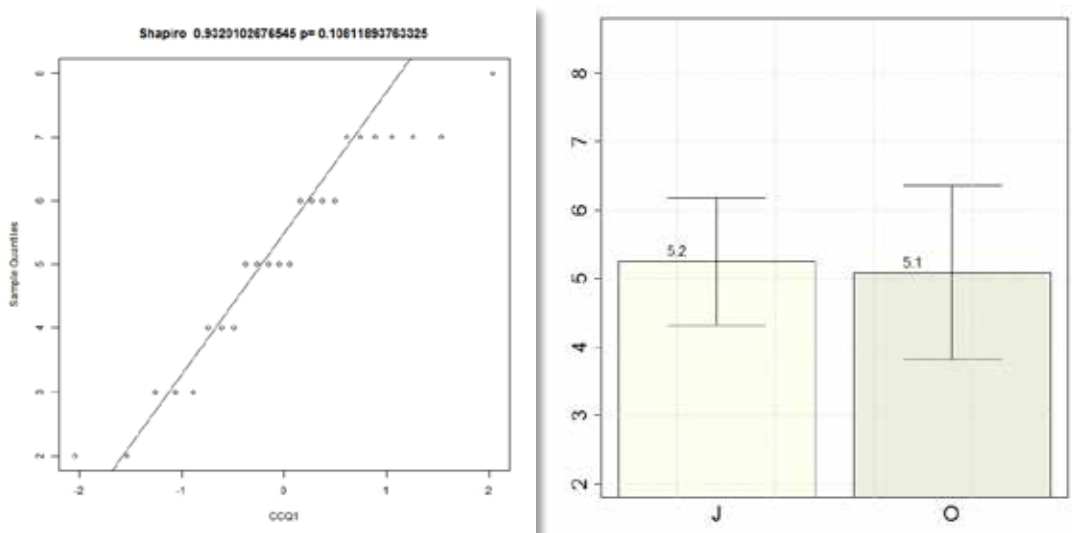
#### 5.3.d.1. Charge cognitive globale

On analyse ici les résultats des réponses à la question : *De manière générale, la compréhension du document et de sa structure m'a demandé un effort mental ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important).*

Caractéristiques de l'indicateur : voir 4.4.d.1.

Normalité des données : Cet échantillon suit bien une loi normale (Shapiro-Wilk :  $W = 0.93$ ,  $p > 0.05$ ) ; l'inspection graphique confirme la proximité des quantiles avec la droite de Henry.

Effets de l'ordre et du document : Ni l'ordre (*Student paired t*(11) = -0.52,  $p > 0.05$ ) ni le document (*Student paired t*(11) = -1.08,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les moyennes et intervalles de confiance des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.



On observe que les moyennes des deux échantillons sont presque égales ( $M_J = 5.2$  et  $M_O = 5.1$ ). Les intervalles de confiance sont également similaires. L'ensemble montre une charge cognitive globale moyenne (5 étant la médiane sur cette échelle de 1 à 9) pour les deux conditions. Un test de *Student* sur données appariées nous confirme l'absence de différence entre ces échantillons ( $t(11) = 0.26$ ,  $p > 0.05$ ).

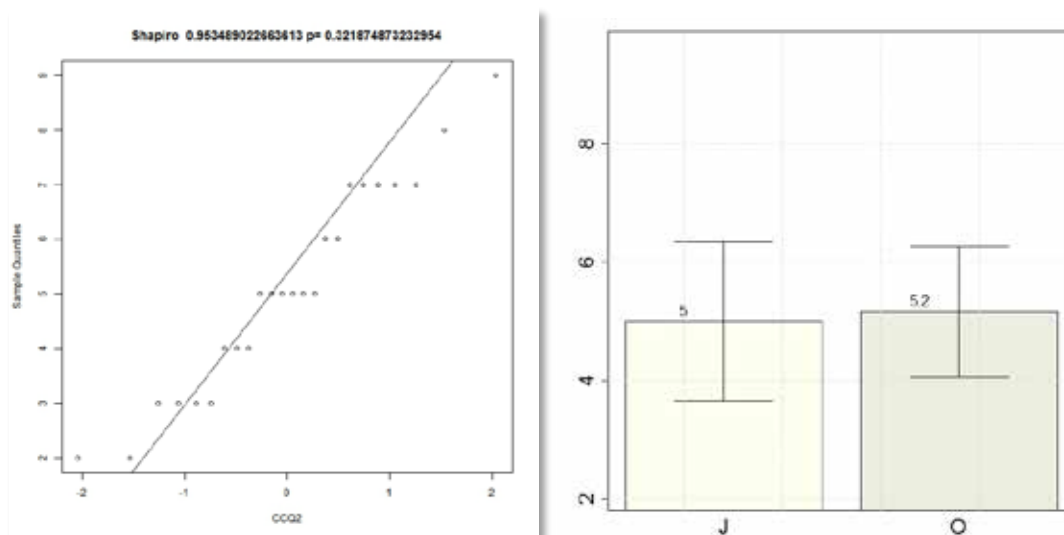
#### 5.3.d.2. Charge intrinsèque liée au contenu

On analyse ici les résultats des réponses à la question : *Comment évalueriez-vous la difficulté de compréhension du contenu ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important).*

Caractéristiques de l'indicateur : voir 4.4.d.2.

Normalité des données : Cet échantillon suit bien une loi normale (Shapiro-Wilk :  $W = 0.95$ ,  $p > 0.05$ ) ; l'inspection graphique confirme la proximité des quantiles avec la droite de Henry.

Effets de l'ordre et du document : Ni l'ordre (*Student paired*  $t(11) = -0.21$ ,  $p > 0.05$ ) ni le document (*Student paired*  $t(11) = -0.21$ ,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les moyennes et intervalles de confiance des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.



On observe que les moyennes des deux échantillons sont presque égales ( $M_J = 5$  et  $M_O = 5.2$ ). Les intervalles de confiance sont également similaires. L'ensemble montre une charge cognitive intrinsèque moyenne pour les deux conditions. Un test de *Student* sur données appariées nous confirme l'absence de différence entre ces échantillons ( $t(11) = -0.21$ ,  $p > 0.05$ ).

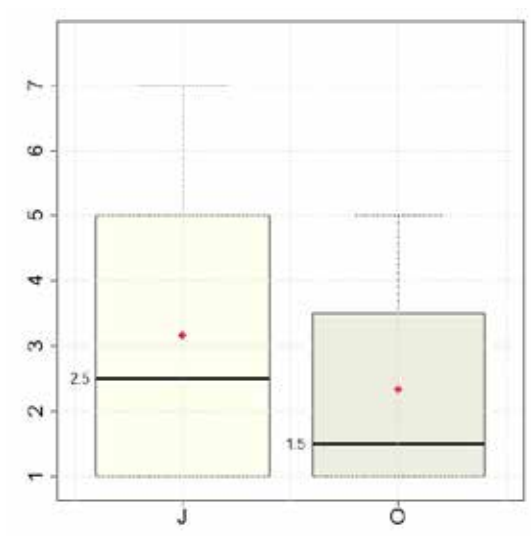
### 5.3.d.3. Charge intrinsèque liée à l'intelligibilité de la synthèse vocale

On analyse ici les résultats des réponses à la question : *Comprendre la façon dont la synthèse vocale prononçait les mots du texte (son intelligibilité) m'a demandé un effort mental ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important)*.

Caractéristiques de l'indicateur : voir 4.4.d.3.

Effets de l'ordre et du document : Ni l'ordre (*Wilcoxon paired*  $Z = 0.94$ ,  $p > 0.05$ ) ni le document (*Wilcoxon paired*  $Z = -1.92$ ,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur, avec cependant une valeur de  $Z$  très proche du seuil critique de 1.96 pour l'effet du document. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les médianes et distributions des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.

Figure 74 - Charge intrinsèque liée à la modalité orale (sujets déficients visuels)



On observe que les médianes des deux échantillons sont proches ( $ME_J = 2.5$  et  $ME_O = 1.5$ ), et faibles en valeur, ce qui correspondant donc à une charge liée à la modalité orale globalement faible, quelle que soit la condition. Même si la médiane de l'échantillon O est inférieure à celle de la condition J, les écarts interquartiles importants ne permettent pas de conclure sur la seule base de l'inspection graphique. Un test de *Wilcoxon* sur données appariées nous confirme l'absence de différence entre ces échantillons ( $Z = 1.94$ ,  $p > 0.05$ ) ; on notera cependant que  $Z$  est très proche du seuil critique de 1.96.

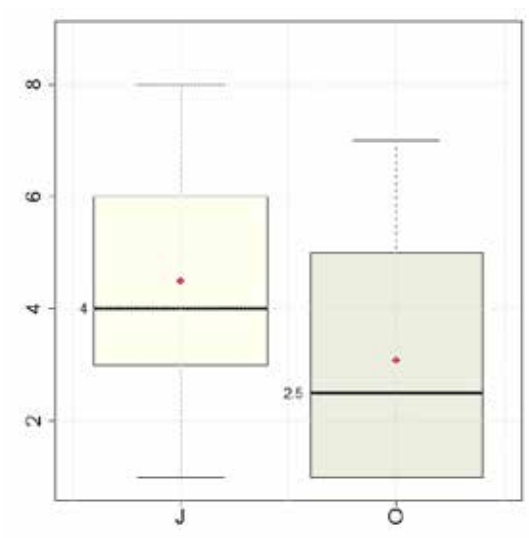
#### 5.3.d.4. Charge extrinsèque liée à la présentation du contenu

On analyse ici les résultats des réponses à la question : *Comprendre la façon dont le texte était présenté à l'oral m'a demandé un effort mental ? (1 à 9, 1 étant un effort très peu important et 9 un effort très important).*

Caractéristiques de l'indicateur : voir 4.4.d.4.

Effets de l'ordre et du document : Ni l'ordre (*Wilcoxon paired*  $Z = -0.82$ ,  $p > 0.05$ ) ni le document (*Wilcoxon paired*  $Z = -0.82$ ,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les médianes et distributions des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.

Figure 75 - Charge extrinsèque (sujets déficients visuels)



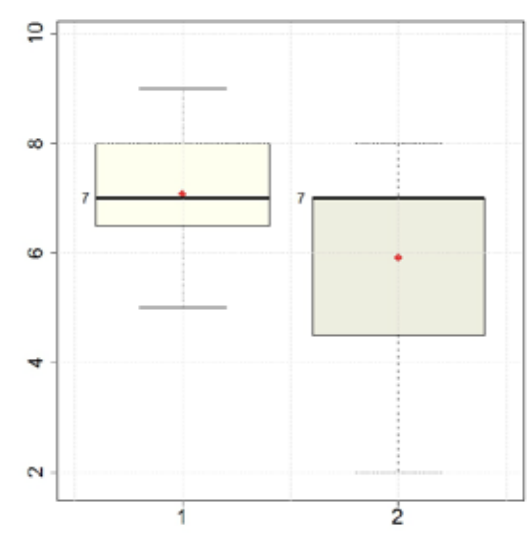
On observe une certaine différence graphiquement entre ces deux échantillons ( $ME_J = 4$  et  $ME_O = 2.5$ ) ; globalement les données indiquent une charge moyenne à faible. L'échantillon O présente cependant un écart interquartile étendu ( $Q3_O - Q1_O = 5$ ), et un fort recouvrement avec l'échantillon de la condition J. Un test de *Wilcoxon* sur données appariées nous confirme l'absence de différence entre ces échantillons ( $Z = 1.33$ ,  $p > 0.05$ ).

#### 5.3.d.1. Charge cognitive liée à l'apprentissage

On analyse ici les résultats des réponses à la question : *Jusqu'à quel point étiez-vous concentré pendant la lecture ? (1 à 9, 1 étant très peu concentré et 9 très concentré)*.

Caractéristiques de l'indicateur : voir 4.4.d.5.

Effets de l'ordre et du document : Ni la **condition** (*Wilcoxon paired*  $Z = 0.06$ ,  $p > 0.05$ ) ni le **document** (*Wilcoxon paired*  $Z = 0.71$ ,  $p > 0.05$ ) n'ont d'effet significatif sur cet indicateur. Cependant l'**ordre** de présentation du document présente un effet significatif (*Wilcoxon paired*  $Z = 2.07$ ,  $p < 0.05$ ). On présente donc les deux sous-ensembles de l'échantillon correspondant aux documents lu en premier et en second durant les tests.

**Figure 76 - Effet de l'ordre de présentation sur la charge cognitive liée à l'apprentissage**

On observe effectivement que l'échantillon 2 montre des valeurs plus faibles que l'échantillon 1 (75% des données de l'échantillon 2 sont inférieures ou égales à la valeur de la médiane de l'échantillon 1). On en conclut donc que la répétition de la tâche a engendré un effet de fatigue/lassitude chez les sujets qui ont engagé moins de ressources cognitives pour l'apprentissage lors de la seconde répétition de la tâche de compréhension.

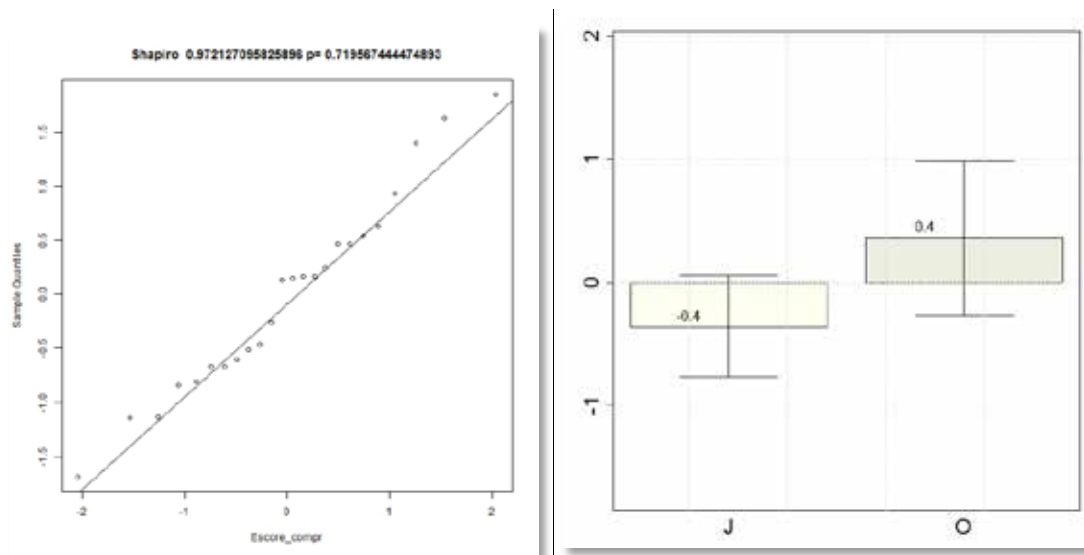
### ***5.3.e. Effet de la stratégie de restitution sur l'efficacité***

L'efficacité représente un ratio entre la performance et l'effort mental engagé (les ressources cognitives) pour obtenir cette performance. Ici la performance est mesurée à travers les scores de compréhension et l'effort mental à travers la charge cognitive globale.

Caractéristiques de l'indicateur : voir 4.4.e.

Normalité des données : Cet échantillon suit bien une loi normale (Shapiro-Wilk :  $W = 0.97$ ,  $p > 0.05$ ) ; l'inspection graphique confirme la proximité des quantiles avec la droite de Henry.

Effets de l'ordre et du document : Ni l'ordre (*Student paired t*(11) = -0.02,  $p > 0.05$ ) ni le document (*Student paired t*(11) = 0.51,  $p > 0.05$ ) n'ont un effet significatif sur cet indicateur. On s'intéresse donc à l'effet de la condition sur l'indicateur. Le graphique ci-dessous présente donc les moyennes et intervalles de confiance des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.



**Figure 77 – Efficience (sujets déficients visuels); (gauche) quantiles et droit de Henry ; (droite) moyennes et intervalles de confiance**

On observe que l'efficience est en moyenne négative en condition J et positive en condition O ( $M_J = -0.4$  et  $M_O = 0.4$ ). Un test de Student sur données appariés montre une différence significative entre ces 2 échantillons ( $t(11) = -2.42$ ,  $p < 0.05$ ). On en conclut qu'en moyenne l'efficience est meilleure en condition O qu'en condition J ; de plus l'efficience est positive en moyenne en condition O, c'est-à-dire que les sujets ont engagé moins de charge cognitive pour une performance égale, ou que les sujets ont obtenu de meilleures performance à charge égale avec la condition J.

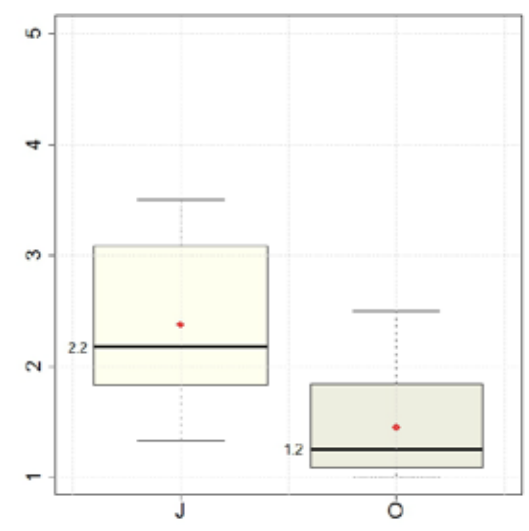
### **5.3.f. Effet de la stratégie de restitution sur la désorientation**

À la suite de la tâche réalisée par les sujets, on leur posait 6 questions mesurant leur sentiment de désorientation durant la tâche. La fiabilité des différentes questions quant à leur capacité à mesurer la même chose (la désorientation), a été évaluée en calculant l'alpha de Cronbach sur l'ensemble des réponses aux 6 questions de désorientation ; ici  $\alpha = 0.85$  ( $\alpha > 0.7$ ) ce qui dénote d'une bonne fiabilité admise [Nunnally 1978].

Caractéristiques de l'indicateur : voir 4.4.f.

Effets de l'ordre et du document : Ni l'ordre (*Wilcoxon paired*  $Z = -0.8$ ,  $p > 0.05$ ) ni le document (*Wilcoxon paired*  $Z = 0.04$ ,  $p > 0.05$ ) n'ont d'effet significatif sur cet indicateur. Le graphique ci-dessous présente donc les médianes et distributions des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.

Figure 78 - Scores de désorientation (sujets déficients visuels)



On observe que le troisième quartile de la condition O est inférieur au premier quartile de la condition J ( $Q3_O = 1.13 < Q1_J = 1.83$ ), ce qui indique que 75% des données de l'échantillon O ont des valeurs inférieures à 75% des données de l'échantillon J. Globalement on observe que la désorientation était faible pour la condition O ( $ME_O = 1.2$ ) et faible à moyenne pour la condition J ( $ME_J = 2.2$ ). Un test de *Wilcoxon* sur données appariées nous confirme que ces deux distributions sont différentes ( $Z = 2.85, p < 0.01$ ).

On en conclut que les sujets étaient moins désorientés en condition O qu'en condition J. Étant donné que nous avons écarté les effets de l'ordre et du document sur l'indicateur, on en conclut que la désorientation est diminuée en condition O.

### 5.3.g. Effet de la stratégie de restitution sur les tâches de recherche d'information

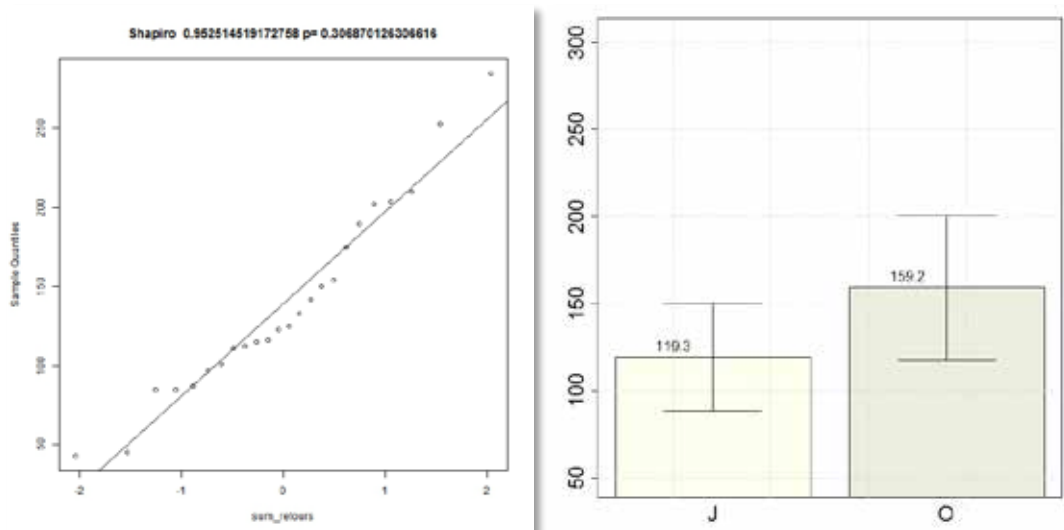
#### 5.3.g.1. Temps moyens mis pour trouver les emplacements cibles

Caractéristiques de l'indicateur : voir 4.4.g.1.

Normalité des données : Cet échantillon suit bien une loi normale (Shapiro-Wilk :  $W = 0.95, p > 0.05$ ) ; l'inspection graphique confirme la proximité des quantiles avec la droite de Henry.

Effets de l'ordre et du document : Ni l'ordre (*Student paired*  $t(11) = -0.25, p > 0.05$ ) ni le document (*Student paired*  $t(11) = 1.44, p > 0.05$ ) n'ont d'effet significatif sur cet indicateur. Le graphique ci-dessous présente donc les moyennes et intervalles de confiance des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.





**Figure 79 - Durées cumulées des tâches de navigation (sujets déficients visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance**

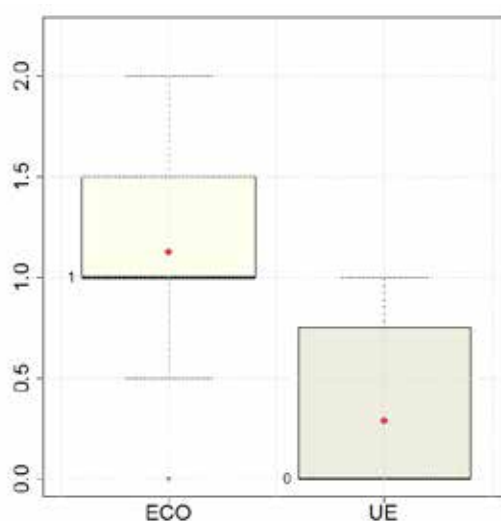
On observe que la moyenne de la condition O est supérieure à celle de la condition J ( $M_O = 159$  et  $M_J = 119$ ). On observe cependant que les intervalles de confiance se recoupent fortement. Un test de *Student* sur données appariées confirme l'absence de différence entre les moyennes de ces deux échantillons ( $t(11) = -1.66, p > 0.05$ ).

### 5.3.g.2. Réussites de la recherche d'informations

Caractéristiques de l'indicateur : voir 4.4.g.2.

Effets de l'ordre et du document : Ni l'**ordre** (*Wilcoxon paired*  $Z = -0.1, p > 0.05$ ) ni la **condition** (*Wilcoxon paired*  $Z = 1.12, p > 0.05$ ) n'ont d'effet significatif sur cet indicateur. Cependant le **document** présente un effet significatif (*Wilcoxon paired*  $Z = 2.65, p < 0.01$ ). On présente donc les deux sous-ensembles de l'échantillon correspondant aux deux documents lus.

**Figure 80 - Échecs lors des tâches de navigation (sujets déficients visuels)**



On observe que pour le document UE, 50% des sujets n'ont eu aucun échec ( $ME_{UE} = 0$ ), tandis que pour le document ECO, 50% des sujets ont fait au moins 1 échec (ou deux échecs proches) sur les

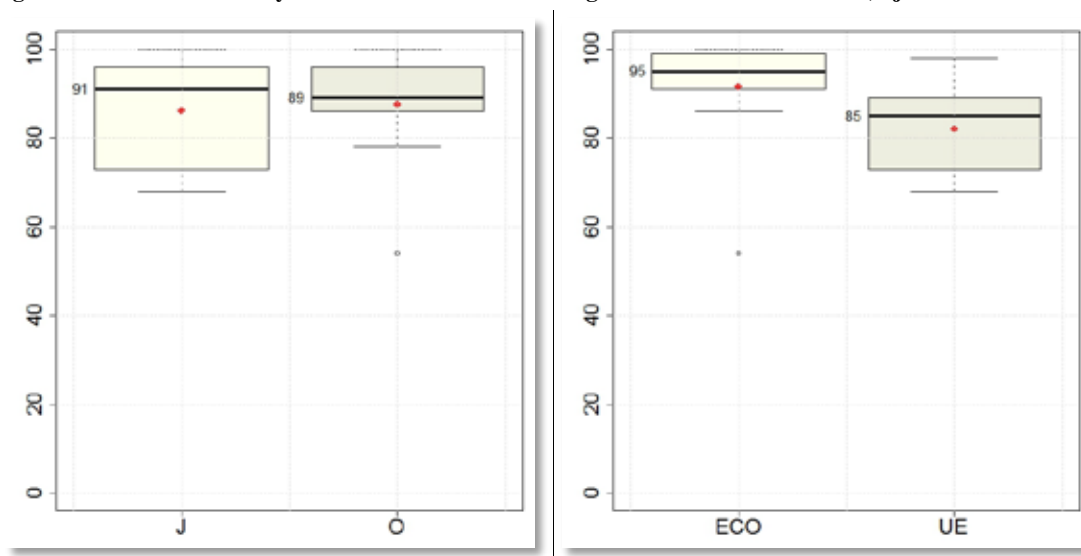
deux tâches ( $ME_{ECO} = 1$ ). Un test de *Wilcoxon* sur données appariées a en effet montré que ces deux distributions sont différentes ( $Z = 2.65, p < 0.01$ ). On en conclut que le document a eu une influence sur la réussite des tâches de navigation, et qu'en l'occurrence le document UE a causé moins d'échecs que le document ECO.

### 5.3.h. Effet de la stratégie de restitution sur l'utilisabilité subjective

Caractéristiques de l'indicateur : voir 4.4.h.

Effets de l'ordre et du document : Ni l'ordre (*Wilcoxon paired*  $Z = 1.41, p > 0.05$ ) ni la condition (*Wilcoxon paired*  $Z = -0.08, p > 0.05$ ) n'ont d'effet significatif sur cet indicateur. Cependant le document présente un effet significatif (*Wilcoxon paired*  $Z = 2.43, p < 0.05$ ). On présente donc les deux sous-ensembles de l'échantillon correspondant aux deux documents lus.

Figure 81 - Utilisabilité du système en fonction de la stratégie et du document abordé (sujets déficients visuels)



On constate graphiquement que quelle que soit la stratégie utilisée, les scores d'utilisabilité sont élevés ( $ME_J = 91$  et  $ME_O = 89$ ), 68 étant considéré pour cet indicateur comme « au-dessus de la moyenne » [Brooke 1996]. Cependant en ce qui concerne les documents abordés, la distribution de l'échantillon du document UE est clairement en dessous de celle de l'autre document ( $Q3_{UE} = 88 < Q1_{ECO} = 91$ ). Cela est confirmé par un test de *Wilcoxon* sur données appariées ( $Z = 2.43, p < 0.05$ ). On en conclut que le document présenté a bien eu une influence sur la perception de l'utilisabilité du système par les sujets, et que le document UE a suscité des scores moins bons que le document ECO.

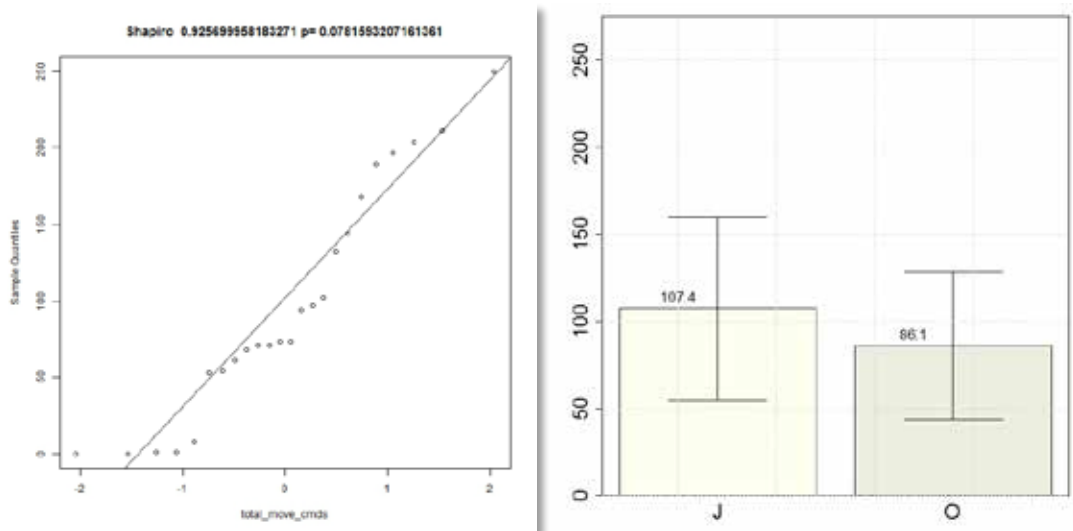
### 5.3.i. Effet de la stratégie de restitution sur l'utilisation des commandes

#### 5.3.i.1. Commandes de navigation

Quelle que soit la condition, les sujets pouvaient se déplacer dans le document à l'aide des touches « flèche » du clavier. On reporte ici le nombre de fois que ces touches ont été utilisées.

Normalité des données : Cet échantillon suit bien une loi normale (Shapiro-Wilk :  $W = 0.93, p > 0.05$ ) ; l'inspection graphique confirme la proximité des quantiles avec la droite de Henry.

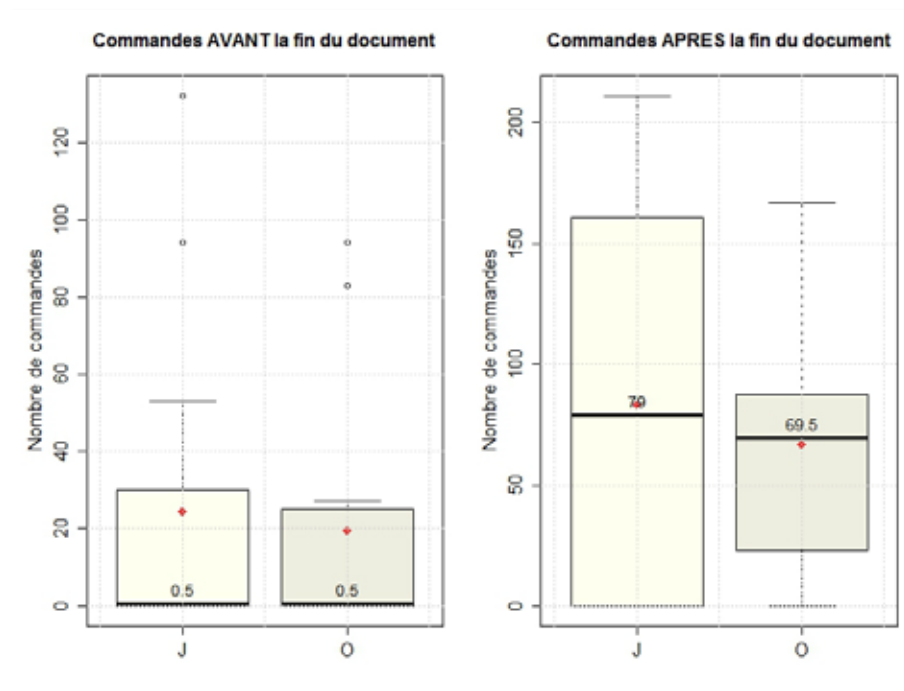
Effets des variables catégorielles : Ni l'ordre (*Student paired t*(11) = 0.63,  $p > 0.05$ ), ni le document (*Student paired t*(11) = 0.17,  $p > 0.05$ ) n'ont d'effet significatif sur cet indicateur. On présente donc les moyennes et intervalles de confiance des 2 sous-ensembles de l'échantillon correspondant aux 2 conditions expérimentales.



**Figure 82 – Commandes de navigation utilisées (sujets déficients visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance**

On observe que les sujets ont utilisés en moyenne ( $M_J = 107$ ) commandes de navigation en condition J, et ( $M_O = 86$ ) commandes en condition O, ce qui correspond respectivement pour une lecture de 15 minutes à 7 et 6 commandes par minute. Cependant, un test de *Student* sur données appariés montre une absence de différence entre ces moyennes ( $t(11) = 1.2$ ,  $p > 0.05$ ).

Nous avons observé lors des tests avec des sujets voyants que ces derniers effectuaient une première écoute exhaustive « passive » pour ensuite naviguer une fois la fin du document atteinte. Ci-dessous nous avons donc étudié l'utilisation des commandes par les sujets déficients visuels, avant d'avoir atteint une première fois la fin du document, et après. Ces données sont non normales (respectivement Shapiro-Wilk ( $W$ ) = 0.63 et ( $W$ ) = 0.89,  $p < 0.05$ ).



**Figure 83 - Commandes de navigation utilisées avant (gauche) et après (droite) la fin du document (sujets déficients visuels)**

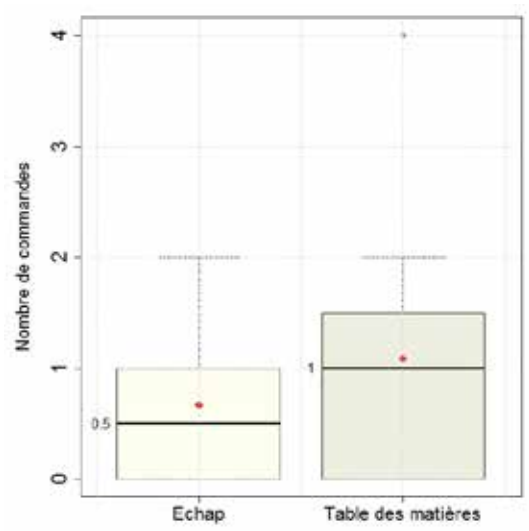
On observe graphiquement très peu de commandes utilisées avant la fin du document ( $ME_{J\_AVANT} = ME_{O\_AVANT} = 0.5$ ) ; 50% des sujets ont donc utilisé moins d'une commande avant la fin du document. On observe une utilisation beaucoup plus importante des commandes après la fin du document ( $ME_{J\_APRES} = 79$  et  $ME_{O\_APRES} = 70$ ), avec cependant une grande variabilité pour les commandes utilisées en condition J ( $Q3_{J\_APRES} - Q1_{J\_APRES} = 158$ ).

Un test de *Wilcoxon* pour données non appariées montre qu'effectivement en condition O il y a eu plus de commandes utilisées après la fin du document qu'avant ( $V = 35$ ,  $p < 0.05$ ). En condition J, la grande variabilité de l'échantillon ne donne aucune différence significative ( $V = 49$ ,  $p > 0.05$ ).

### 5.3.i.2. Commandes spécifiques au mode optimal

Dans le mode de lecture optimal, les sujets avaient accès à deux fonctions supplémentaires : ECHAP leur permettait de lire un message les situant dans le document, et ENTREE leur permettait d'accéder à la table des matières et de revenir au document (ici seules les appuis sur ENTREE pour passer dans la table des matières ont été comptabilisés).

**Figure 84 - Commandes du mode optimal utilisées (sujets déficients visuels)**



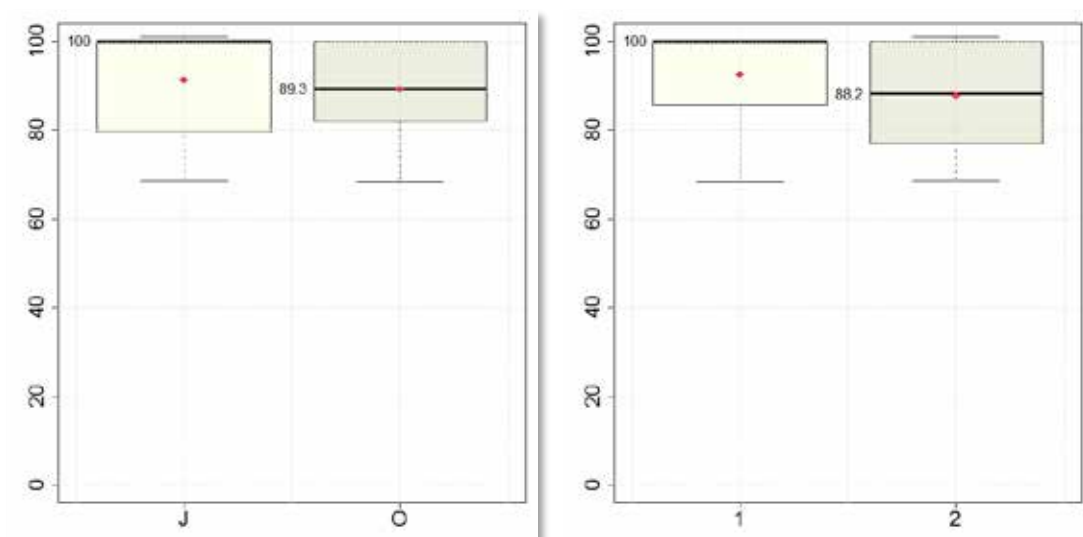
On observe une utilisation très faible des commandes spécifiques au mode optimal par les sujets. La touche ECHAP est utilisée une seule fois ou moins par 75% de sujets, et seuls 50% des sujets ont consulté au moins une fois la table des matières.

Concernant la table des matières, ces résultats n'indiquent cependant pas combien de temps les sujets ont passé à parcourir la table des matières. Nous avons récupéré ces temps à l'aide des fichiers traces. Cet échantillon de durées suit une loi normale (Shapiro-Wilk :  $W = 0.89$ ,  $p > 0.05$ ) ; les sujets ont en moyenne parcouru la table des matières pendant ( $M = 100$ ) secondes ( $SD = 99$ ), soit 1min40secs, ce qui représente environ 10% du temps total imparti pour le parcours du document en condition optimale. À noter que l'écart-type est tout de même de  $\pm 1\text{min}39\text{secs}$  (soit  $\pm 10\%$  du temps total imparti), ce qui signifie qu'environ 70% des sujets ont passé moins de 20% du temps total imparti à parcourir la table des matières.

### ***5.3.j. Effet de la stratégie de restitution sur les temps de parcours des documents***

Caractéristiques de l'indicateur : voir 4.4.j.

Effets des variables catégorielles : Ni le document (*Wilcoxon paired*  $Z = -0.14$ ,  $p > 0.05$ ) ni la condition (*Wilcoxon paired*  $Z = 0.84$ ,  $p > 0.05$ ) n'ont d'effet significatif sur cet indicateur. Cependant l'ordre de présentation influe sur ce dernier de manière significative (*Wilcoxon paired*  $Z = 2.1$ ,  $p < 0.05$ ). On présente donc les deux sous-ensembles de l'échantillon correspondant aux deux possibilités d'ordre pour la lecture (en premier ou en second).

**Figure 85 - Temps total de parcours des documents en fonction de la stratégie de restitution et de l'ordre de présentation (sujets déficients visuels)**

On observe une certaine variabilité dans les temps de parcours des sujets quelle que soit la condition (on rappelle que la durée de lecture d'un document représente environ 60% du temps total). Malgré cela, pour les deux stratégies on n'observe pas de différence significative entre les temps de parcours, montrant qu'une majorité des sujets a utilisé environ 90% ou plus du temps imparti ( $ME_0 = 89$ ).

Concernant l'ordre dans lequel les documents ont été abordés, 50% des sujets ont utilisé l'intégralité du temps imparti ( $ME_1 = Q_3 = 100$ ) pour les documents abordés en premiers. Pour les documents abordés en second, 50% des sujets ont utilisé 88% ou moins du temps imparti ( $ME_2 = 88$ ). Étant donné que ces deux distributions présentent une différence significative ( $Wilcoxon Z = 2.1, p < 0.05$ ), on en conclut que les sujets ont arrêté la lecture du document plus tôt pour les documents abordés en second que pour les documents abordés en premier.

### 5.3.k. Retours qualitatifs

Même chose que pour 4.4.k.

Thème	Nombre d'occurrences	Conditions correspondantes	Exemples
Concentration stable pendant la lecture	N = 6	O et J	- pas de changements majeurs - quelques toutes petites baisses de concentration...
Utilité de la table des matières	N = 7	O	- j'ai préféré la lecture avec la table des matières - J'ai bien aimé avoir une vision avec la table des matières
Utilité du son spatialisé	N = 7	O	- ça enrichie les infos sur la longueur d'une ligne, ça donne une notion - ça m'a permis de voir quand c'était la fin des paragraphes, alors on est plus attentifs on se dit ah oui on passe à autre chose
Son spatialisé inutile voir perturbant	N = 5	O	- je trouve ça déconcentre plus que ça aide, - moi je l'entendais mais ça ne m'a pas servi plus que ça
Utilité des icônes audio	N = 6	O	- même sans vraiment les intellectualiser on les repère, j'aime bien - quand on passe d'une partie à l'autre on arrive bien à se situer avec les bips, à visualiser le texte
Inutilité des icônes audio	N = 6	O	- Les bips pour les titres je suis passé outre - moi je l'entendais mais ça ne m'a pas servi plus que

			<i>ça</i>
Utilité de la voix off	N = 4	O	- par rapport à la vision globale du truc la voix off aide beaucoup - les deux vois ça me va bien déjà
Réalisme de la condition « lecteur d'écran »	N = 5	J	- c'est du classique ça... - oui bon c'est comme JAWS, rien de bien nouveau
Avis général positif	N = 3	O	- j'ai tout aimé - je suis tout à fait emballé. Ça me paraît être très utile
Simplicité du système appréciée par rapport à la complexité des systèmes existants	N = 3	O et J	- si pour naviguer sur une page internet ça pouvait être aussi simple j'aimerais bien l'utiliser fréquemment - Moi ce que je reproche aux lecteurs d'écran c'est qu'il y a trop de raccourcis : pas simple d'utilisation

### 5.3.1. Synthèse des résultats

Ci-dessous un tableau reprenant l'ensemble des résultats présentés précédemment. Les résultats présentent les différences significatives et absences de différences entre les différentes conditions expérimentales pour les variables étudiées.

Deux conditions étaient étudiées : J (Jaws, lecteurs d'écran) et O (optimale). Pour présenter les résultats de manière synthétique, on présente les différences significatives avec des opérateurs mathématiques d'inégalité (« > » et « < »). Par exemple si les résultats de la condition optimale (O) ont été supérieurs à ceux de la condition sans MFM (N), on reportera dans le tableau « O > N » dans la colonne des résultats significatifs. De la même façon, on présente l'absence de différences significative à l'aide de l'opérateur égal.

Sont présentés également dans le tableau les éventuels effets de l'ordre de présentation du document (1 ou 2 pour les documents respectivement abordés en premier ou en second), et les effets du document sur l'indicateur (document UE traitant de l'Union Européenne dans la mondialisation et le document ECO trait d'économie).

À noter que pour chaque variable du tableau, un lien vers la partie correspondante des résultats est inséré.

<b>Variable étudiée</b>	<b>Test utilisé</b>	<b>Différences significatives</b>	<b>Résultats vérifiant les hypothèses ?</b>
<a href="#">Plan : thèmes</a>	<i>Student</i>	O > J	Oui
<a href="#">Plan : distance</a>	<i>Wilcoxon</i>	-	Non
<a href="#">Plan : ordre</a>	<i>Wilcoxon</i>	-	Non
<a href="#">Plan : différence de structure</a>	<i>Wilcoxon</i>	O < J	oui
<a href="#">Compréhension</a>	<i>Student</i>	O > J	oui
<a href="#">Charge cognitive globale</a>	<i>Student</i>	-	non
<a href="#">Charge intrinsèque (contenu)</a>	<i>Student</i>	-	Non
<a href="#">Charge intrinsèque (TTS)</a>	<i>Wilcoxon</i>	O < J (tendance)	Partiellement
<a href="#">Charge extrinsèque (présentation du contenu)</a>	<i>Wilcoxon</i>	-	non
<a href="#">Charge d'apprentissage</a>	<i>Wilcoxon</i>	2 < 1	non

<a href="#"><u>Effizienz</u></a>	<i>Student</i>	$O > J$	oui
<a href="#"><u>Désorientation</u></a>	<i>Wilcoxon</i>	$O < J$	oui
<a href="#"><u>Navigation : temps</u></a>	<i>Student</i>	-	-
<a href="#"><u>Navigation : échecs</u></a>	<i>Wilcoxon</i>	$UE < ECO$	-
<a href="#"><u>Utilisabilité</u></a>	<i>Wilcoxon</i>	$UE < ECO$	-
<a href="#"><u>Commandes de navigation utilisées</u></a>	<i>Student</i>	-	-
<a href="#"><u>Commandes de navigation utilisées : avant/après</u></a>	<i>Wilcoxon</i>	$O_{\text{après}} > O_{\text{avant}}$	-
<a href="#"><u>Durée de parcours du document</u></a>	<i>Wilcoxon</i>	$2 < 1$	-

Tableau 35 - Synthèse des résultats : expérimentation sur sujets déficients visuels

## 5.4. Discussion

Nous avons testé 2 stratégies différentes de restitution de la MFM : une première stratégie reproduisait la restitution orale des lecteurs d'écran (restitution minimale de la mise en forme) et une autre tentait de restituer la MFM de manière optimale. Chaque sujet abordait successivement deux documents (aux contenus différents mais avec des structures équivalentes) avec chacune des 2 stratégies définies. Leur objectif à chaque fois était d'essayer de comprendre au mieux le contenu du document et sa structure ; ils devaient ensuite répondre à un questionnaire de compréhension, évaluer la charge cognitive induite par la tâche ainsi que la désorientation ressentie pendant cette dernière. Les scores de compréhension moyens et les connaissances antérieures faibles des sujets confirment la difficulté de la tâche, bien que celle-ci puisse juste être qualifiée de moyenne au regard des faibles scores de charge cognitive.

Comme pour l'expérimentation précédente, nos résultats montrent globalement l'intérêt de restituer la MFM à l'oral dans le cadre d'une tâche de compréhension de difficulté moyenne : à charge cognitive égale (donc effort mental équivalent), les performances sont meilleures et les sujets étaient moins désorientés. On retrouve cette conclusion dans les scores d'efficacité des sujets, qui montrent que le ratio performance sur ressources cognitives mises en jeu est meilleur en condition optimale (voir 5.3.e). On notera cependant que ce cadre expérimental ne permet pas de préciser les rôles qu'ont joués respectivement la restitution des indices locaux de mise en forme et de la table des matières, étant donné que la stratégie optimale comprenait les deux et l'autre stratégie aucun des deux.

Ces résultats confirment la supériorité de la stratégie optimale, déjà observée auprès de sujets non-déficients, mais qui était ici confrontée à une restitution avec laquelle les sujets étaient accoutumés. Il est ainsi intéressant d'observer que les non-voyants bénéficient également de la restitution de la structure du document en termes de compréhension et de rétention du plan, mais dans notre cas pas en termes de charge cognitive. Concernant ce dernier point, les scores de charge cognitive globale étant juste moyens, il est possible que le débit lent (pour ce type de population) de la synthèse vocale, et l'habitude des non-voyants à traiter l'information orale, leur aient permis de mieux faire face aux exigences de la tâche. Cette difficulté a pu être insuffisante pour que les indices de MFM, locaux comme globaux, aient pu soulager la charge mentale en limitant les inférences, ces indices-là ayant pu être inférés sur la base du discours (du contenu du texte indépendamment des



informations de MFM). De nouveaux tests avec un débit vocal plus élevé devraient être menés pour évaluer cette hypothèse.

Les résultats des scores de charge cognitive indiquent également que malgré l'ajout de nombreuses fonctionnalités en condition optimale (dont la table des matières et ses liens hypertextes), et malgré le fait que les sujets étaient tous familiers de la stratégie « lecteur d'écran », la charge cognitive n'a pas augmenté. Les sujets auraient donc soit rapidement intégré l'utilisation des fonctionnalités à leur disposition, soit ignoré les indices non-verbaux restitués. Étant donné que la population était partagée en 2 d'après les retours qualitatifs (il y a autant de sujets qui ont apprécié et utilisés les indices non-verbaux que de sujets qui les ont ignorés ou qui ont été gênés par leur présence), on peut supposer que les deux cas de figures étaient présents dans la population. Quoi qu'il en soit, il est donc envisageable d'ajouter de nouvelles fonctionnalités de lecture aux stratégies définies en restant dans des limites optimales de charge mentale pour une tâche de compréhension (une charge trop faible ou trop importante pénalisent l'apprentissage [Teigen 1994]).

Les retours qualitatifs des participants montrent qu'il est important de prendre en compte les préférences individuelles pour la lecture. Ainsi les utilisateurs déficients devraient pouvoir facilement paramétrer la restitution, les modalités de lecture, pour être les plus efficaces et les moins frustrés possible. Nous avons déjà vu dans les résultats de la première expérimentation (voir 3.4) que les sujets déficients avaient des préférences hétérogènes, que nous pourrions expliquer par le simple constat général que de très nombreux paramètres rentrent en compte quant à la manière qu'ont les individus de gérer et se représenter l'information. En fonction de l'âge d'apparition de la déficience, du type de déficience<sup>124</sup>, de l'éducation ou d'autres facteurs environnementaux, ou encore en fonction des technologies utilisées dans un cadre personnel, les stratégies compensatoires mises en place peuvent être très hétérogènes ([Vigo and Harper 2013] illustrent bien la difficulté de les recenser dans le cadre pourtant plus restreint d'une tâche de recherche d'information), ce qui aboutit même avec des déficiences similaires à des préférences souvent différentes. Cependant, la conception des technologies d'assistance ne doit pas être basée uniquement sur les préférences des utilisateurs qui sont d'ailleurs parfois en contradiction avec les performances [Nielsen and Levy 1994] ; la possibilité de paramétrer la restitution semble être un bon compromis, si tant est que la configuration soit elle-même accessible et utilisable.

Globalement les sujets ont apprécié l'utilisation du système quelle que soit la stratégie si l'on s'en réfère aux scores d'utilisabilité élevés. À ce propos, l'utilisabilité semble avoir été significativement meilleure pour un des deux textes utilisés (bien que les scores fussent élevés dans les deux cas), ce qui montre la difficulté d'évaluer précisément un système de lecture indépendamment du contenu, problématique qui peut être étendue aux lecteurs d'écran. Les scores d'utilisabilité sont cependant en cohérence avec le fait que la désorientation ressentie par les sujets était faible quelle que soit la stratégie utilisée ; la stratégie optimale a cependant observé des scores inférieurs à l'autre, ce qui valide le fait que les indices locaux et la table des matières aient permis une meilleure orientation dans le document. Nous pensons que, comme pour la charge cognitive, les faibles scores de désorientation sont liés à la grande habitude des sujets à parcourir des documents en utilisant des interfaces non-visuelles. Non seulement les sujets étaient familiers de ce genre de systèmes de

---

<sup>124</sup> Par exemple, le seul sujet de notre population ayant une vision partielle (1/50<sup>ème</sup>) nous a rapporté qu'il utilisait sa vue pour repérer la structure des documents, et qu'il accédait ensuite si besoin au contenu par synthèse vocale

lecture, mais ils étaient habitués à des systèmes beaucoup plus complexes que celui utilisé dans le cadre expérimental. Plusieurs sujets ont d'ailleurs rapporté avoir apprécié la simplicité de la lecture dans ce cadre expérimental.

En effet, même si la stratégie « lecteur d'écran » était représentative des technologies actuelles en termes d'oralisation, elle ne proposait que des fonctions de déplacements basiques alors que les lecteurs d'écrans incluent des dizaines de fonctions de déplacement, d'accès aux diverses propriétés du document et d'interaction avec le système informatique utilisé [Freedom Scientific 2014]. C'est cette « simplicité » du système qui aurait été appréciée (cf. scores d'utilisabilité et retours qualitatifs), et qui a pu limiter la désorientation des sujets. Cela soulève un point d'importance pour les lecteurs d'écrans et les systèmes de lecture en général qui est la pertinence de ces fonctions au regard de l'utilisation qui en est faite dans certains cadres comme la compréhension de documents. Il est d'ailleurs étonnant de ne trouver aucune référence à l'évaluation de l'utilisabilité des lecteurs d'écran dans la littérature. On rejoint encore une fois l'importance d'adopter une démarche holistique dans la conception de tels systèmes [Giraud et al. 2011].

On retrouve la même stratégie que celle utilisée dans l'expérimentation précédente par les sujets non-déficients ; ici également les sujets non-voyants ont effectué une première lecture exhaustive plutôt passive avant de rentrer dans une phase de navigation. Encore une fois il semble important de mener de nouvelles études quant aux stratégies de navigation dans le cadre de la compréhension de documents, et ce afin de proposer une lecture adaptée du document dans le cas d'une tâche de compréhension.

Enfin, on notera que les ressources cognitives engagées pour la compréhension du contenu n'ont pas diminué avec la suppression des contenus périphériques en condition optimale ; comme pour l'expérimentation avec les sujets non-déficients, cela peut indiquer que les contenus périphériques étaient présents en trop faible quantité ou de manière trop éparse. On peut émettre l'hypothèse que les non-voyants mettent en place des stratégies efficaces de filtrage attentionnel de l'information orale pour préserver leurs ressources cognitives.

En conclusion, la restitution de la MFM a été de manière générale bénéfique aux sujets non-voyants. L'ensemble des résultats ont une implication directe sur la conception de systèmes de lecture non-visuelle de documents dans le cadre de l'apprentissage : la structure du document doit être accessible à travers plusieurs fonctions, et des indices non verbaux de MFM peuvent être incorporés afin de permettre aux utilisateurs de mieux s'orienter lors de la navigation (quel que soit leur objectif, apprentissage ou autre). Ces deux points permettent ainsi d'optimiser l'efficacité de l'apprentissage. Il est également important que la restitution soit paramétrable pour respecter les préférences inter-individus particulièrement hétérogènes chez les déficients visuels ; cependant l'interface de configuration doit être accessible et utilisable pour que le paramétrage ait un quelconque bénéfice, ce qui nous semble être un sujet de recherche à part. Plus généralement, l'ensemble des interactions et modalités de restitution devraient présenter une complexité limitée, et devraient pouvoir être adaptables aux capacités de l'utilisateur : plusieurs sujets dans notre cas ont apprécié la simplicité de l'interaction proposée dans ce cadre expérimental, comparativement aux systèmes de lecture qu'ils utilisent habituellement. On peut en conclure que la conception de systèmes de lecture pour être optimale doit prendre en compte non seulement les caractéristiques de la population (préférences et expertise), les caractéristiques de la tâche (ici une tâche de

compréhension), mais aussi les architectures textuelles véhiculées par la MFM des textes et les fonctions de lecture qu'elle permet de mettre en place.

## **6. Conclusion sur l'ensemble des résultats**

---

---

Nous reprenons ici les conclusions des trois expérimentations présentées dans ce chapitre, qui évaluaient respectivement la restitution des titres (3), la restitution globale de la MFM<sup>125</sup> auprès de sujets non-déficients (4) et la restitution globale de la MFM auprès de sujets non-voyants (5). Les résultats de la première expérimentation sur les titres n'ayant donné aucun résultat significatif, la discussion portera principalement sur les résultats des deux dernières expérimentations, sauf exception quand cela est précisé.

Le point de départ de l'ensemble de ces travaux est le postulat que la mise en forme matérielle est une composante essentielle des documents pour la lecture visuelle, notamment dans le cadre de la compréhension (voir Chapitre I1.2), qui au-delà de l'esthétique du texte possède une dimension informative et fonctionnelle, dimensions qui ne sont que très peu prises en compte par les technologies actuelles (voir Chapitre II6). Ainsi, les trois expérimentations que nous avons menées avaient pour but d'évaluer l'impact de la restitution de la MFM des textes lors de leur accès par synthèse vocale. L'ensemble des résultats montre que cette restitution améliore la compréhension du texte et la rétention du plan (le second contribuant au premier [Lemarié et al. 2012]), et permet aux sujets de se sentir moins désorientés pendant l'accès au document. Ces résultats ont été validés à la fois auprès de sujets non-déficients et de sujets non-voyants ; pour ces deux types de populations, la restitution de la MFM a d'ailleurs permis une meilleure efficacité de compréhension (performance/ressources cognitives).

L'ensemble des hypothèses formulées pour les deux dernières expérimentations portaient sur ce qu'il était pertinent de restituer quant à la MFM (le « quoi »), la manière et le moment les plus pertinents pour restituer ces informations (le « comment » et le « quand »). Nous suivrons cette logique pour conclure sur l'ensemble des résultats, en commençant par la nature des informations restituées (le « quoi »). Les informations de MFM restituées peuvent être scindées en deux catégories pour les besoins de l'analyse, à savoir la restitution de la structure globale du document et de phénomènes dits « locaux » de MFM.

La structure globale était restituée à travers la possibilité pour les sujets d'accéder à la table des matières du document, et donc à l'organisation globale de ses thèmes. La restitution de la structure globale du document a en effet montré son importance pour la compréhension, en permettant des performances 1.5 à 2 fois supérieures et une charge mentale réduite pour les non-déficients, et équivalente pour les déficients visuels. Ces résultats mettent l'accent sur l'importance d'étudier la tâche réalisée sur la base du texte pour restituer la MFM, ici une tâche de compréhension. Dans notre cas, cette étude a montré que la compréhension passe par la construction d'une représentation globale cohérente de l'organisation des thèmes traités dans le texte (d'après les modèles psycho-cognitifs de compréhension), et que cette construction s'appuie notamment en lecture visuelle sur les titres signalés dans le texte : d'où la restitution de la structure globale des

---

<sup>125</sup> Le terme *mise en forme matérielle* (MFM) renvoie à un concept fondamental dans nos travaux. Pour plus de détails sur sa signification voir Chapitre I ou le Glossaire.

thèmes avec la table des matières. Nous avons évalué une méthode particulière d'accès à la structure globale, mais d'autres fonctions de navigation dans ce type de structure hiérarchique peuvent s'avérer pertinentes [Smith et al. 2004], par exemple naviguer entre les titres de même niveau hiérarchique.

L'ensemble des phénomènes locaux de MFM restitués informaient sur le lecteur sur la nature des segments textuels lus ainsi que leur contenu, et explicitaient ainsi les informations données localement<sup>126</sup> par la MFM dans le texte (ex. le fait que le segment soit une énumération, combien d'items comportait l'énumération et de quoi traitaient-ils, le niveau d'un titre, l'explicitation d'une énumération incluse dans le discours signalée par la syntaxe, etc. ; voir 4.2.c.3). La restitution de la MFM à un niveau local a permis de diminuer les ressources mentales engagées pour la compréhension chez les non-déficients ; ce résultat illustre le rôle de facilitateur de traitement de la MFM qui permet de limiter les inférences faites pendant la lecture. Cependant les non-voyants n'ont pas profité de cet effet d'allègement de la charge cognitive : nous pensons que leurs capacités supérieures de traitement de l'information orale leur ont permis de réaliser plus facilement ces inférences sur la base du discours, indépendamment de la restitution des phénomènes locaux de MFM. Ces capacités pourraient avoir été induites par l'adaptation neurologique à la déficience à travers la plasticité cérébrale [Hugdahl et al. 2004; Burton 2003] ; leur expertise des systèmes de lecture non-visuelle a également pu leur permettre de développer des stratégies de traitement de l'information plus efficaces. Afin d'évaluer la validité de cette hypothèse, il serait intéressant d'observer l'impact de la restitution des phénomènes de MFM lorsque le texte est oralisé à une vitesse supérieure ou lorsque le contenu est plus complexe, c'est-à-dire lorsque l'on augmente la difficulté de la tâche.

De manière générale, restituer la MFM ne semble pas avoir eu d'effet sur la charge cognitive des non-voyants là où elle a contribué à la diminuer pour les non-déficients. On peut donc supposer qu'il existerait un point d'équilibre à trouver pour la restitution, afin que la MFM (restituée) puisse jouer le même rôle de facilitateur de traitement qu'à l'écrit. En restituer « trop » aurait un effet inverse à l'allègement escompté, en obligeant les lecteurs à traiter des informations orales redondantes car déjà inférées sur la base du discours ; en restituer « trop peu » demanderait à contrario plus d'inférences de la part des lecteurs et par conséquent augmenterait la quantité de ressources cognitives allouées pour la tâche. Tout le problème de la restitution de la MFM se résume donc à trouver le bon compromis entre restituer pour faciliter les traitements et améliorer l'accessibilité de manière générale, et ne pas restituer pour éviter une redondance d'informations et une surcharge cognitive. Ce compromis doit être guidé à la fois par le degré d'ambiguïté et de complexité du contenu à restituer lorsqu'il est dépourvu de MFM (le verbal est-il suffisant ? voir Chapitre IV pour plus de détails sur les problématiques générales de la restitution de la MFM), les capacités de traitement cognitif des sujets et la complexité de la tâche à réaliser sur la base du texte.

Notre réflexion portait également sur « comment » restituer la MFM : nous utilisons à la fois des indices verbaux et non-verbaux pour restituer l'information véhiculée par les différents phénomènes de MFM des textes. Nos résultats montrent que quelle que soit la population étudiée, l'utilisation d'indices non-verbaux tend à réduire la charge cognitive et permet aux sujets de mieux s'orienter

<sup>126</sup> Le terme local est utilisé ici par opposition à la structure globale restituée, et fait référence aux indices micro et macrostructuraux intervenant dans la compréhension du texte (d'après le modèle de [Van Dijk and Kintsch 1983]), indices fournis par la mise en forme.

dans le texte. L'intérêt des indices non-verbaux est multiple : comme démontré, l'utilisation d'indices non-verbaux permet de donner des repères de navigation dans le document, et réduit ainsi la désorientation ressentie ; de plus, le non-verbal permet d'économiser l'espace audio en fournissant des informations d'une manière plus synthétique (par exemple avec des icônes audio), ou en parallèle de l'information verbale (en modulant les voix de synthèse ou en spatialisant le discours oralisé). Cette dernière remarque fait écho à un problème majeur rencontré par les déficients visuels, qui est le facteur temps [Sperandio and Uzan 2002]. Enfin, les indices non-verbaux permettent de rajouter une certaine saillance ou profondeur dans la restitution (notamment pour démarquer les frontières entre les segments de texte), par contraste entre le verbal, qui représente la modalité principale utilisée, et non-verbal utilisé plus minoritairement (voir Chapitre 12.2.b pour plus de détails sur le principe de contraste utilisé pour les marques visuelles de MFM). Les indices non-verbaux présentent cependant un inconvénient majeur qui est l'apprentissage. En effet en définissant une « grammaire » non-verbale, son utilisation ne peut être optimale qu'après apprentissage des règles définies par cette grammaire, d'où l'intérêt pour de futurs travaux d'étudier les capacités d'apprentissage des sujets d'une grammaire non-verbale plus complexe. Concernant ce dernier point les résultats des mesures de charge cognitive semblent indiquer que les sujets non-voyants ont utilisé peu de ressources pour interpréter la grammaire utilisée, ce qui laisse penser qu'une grammaire plus complexe pourrait être mise en place, tout en restant dans les limites optimales des ressources cognitives engagées pour la compréhension et l'apprentissage.

Les préférences individuelles des sujets non-voyants représentent un autre point intéressant relatif à la manière de restituer la MFM (le « comment »), et plus généralement pour l'accès non-visuel aux contenus. Nous avons observé dans les deux expérimentations menées avec des sujets non-voyants que ces derniers montraient des préférences hétérogènes quant aux modalités de restitution, et que ces préférences étaient très marquées puisque souvent mentionnées dans les retours qualitatifs là où, avec le même protocole, des sujets non-déficients plus nombreux n'ont que rarement mentionné de préférences. La diversité des déficiences et des stratégies compensatoires mises en place peuvent avoir contribué à cette hétérogénéité ; de plus, l'expertise des systèmes de lecture développée par les non-voyants a pu contribuer à ces avis plus tranchés sur les modalités de restitution, tandis que les sujets non-déficients n'avaient pas de point de comparaison pour se positionner. La première expérimentation sur les titres laissait d'ailleurs entrevoir le fait que les préférences puissent impacter les performances (ici de compréhension). Quoi qu'il en soit, il semble donc important que les systèmes de lecture non-visuelle puissent s'adapter aux préférences des utilisateurs ; cela représente une problématique délicate, puisque d'après ce que nous ont rapporté plusieurs formateurs d'un Institut pour Jeunes Aveugles la configuration d'un tel système en elle-même semble être une activité difficile, plutôt réservée à des utilisateurs experts.

Enfin, un dernier point portant sur la manière de restituer la MFM concerne le caractère *actif* ou *passif* de cette restitution ; ces notions n'ont à notre connaissance pas été définies en tant que telles dans la littérature, mais permettent ici une distinction intéressante entre deux méthodes de restitution de l'information dans des modalités non-visuelles. Pour nos expérimentations, la structure globale était restituée de manière *active* par l'appui sur une touche clavier, tandis que la restitution des phénomènes locaux de MFM était *passive*, dans le sens où les informations verbales ou non-verbales correspondantes étaient oralisées au fil de la lecture sans intervention active de la part des sujets. Une restitution active a l'avantage de ne pas imposer au sujet de traiter directement des informations supplémentaires en lui laissant le choix de pouvoir accéder ou non à certaines

informations ; cela permet également une plus grande adaptabilité de la restitution car le sujet peut mettre en place sa propre stratégie d'accès à l'information. Cependant, une restitution active implique que le sujet ait mémorisé l'action utilisateur correspondant à la restitution (ici un raccourci clavier), et qu'il ait au préalable une compréhension de la nature et de l'utilité des informations restituées à travers cette action, afin que la restitution puisse justement servir une stratégie de lecture. Ceci suppose que l'utilisateur possède des habiletés métacognitives élevées ; c'est pourtant le parti-pris des lecteurs d'écrans dont la restitution est principalement basée sur des raccourcis clavier (voir Chapitre II5.2). À contrario, une restitution passive rajoute des informations à traiter mais permet de guider le lecteur dans l'élaboration de sa représentation du contenu, sans nécessiter la mise en place de stratégies particulières du moment que la restitution peut être interprétée par le lecteur (en référence au cas d'une grammaire non-verbale). Là encore, nous pensons que plusieurs facteurs doivent être pris en compte pour trouver le juste équilibre entre restitution active et passive. L'utilité des informations de MFM restituées relativement à l'explicitation du contenu en est un premier : si elles sont d'importance pour la compréhension du contenu, une restitution passive sera plus pertinente. Le degré de complexité des informations à restituer en est un autre ; des informations trop complexes pourraient amener une surcharge cognitive. Enfin les capacités de traitement cognitif des utilisateurs et leur expertise du système utilisé peuvent également aider à trouver l'équilibre optimal.

L'ensemble de nos conclusions soulignent la nécessité d'adopter une approche *holistique* (c'est-à-dire globale) dans le cadre de la restitution de la MFM et plus généralement pour la conception de systèmes d'accès non-visuel aux documents, en prenant en compte à la fois les caractéristiques des documents, des utilisateurs et de la tâche réalisée dans un contexte donné. L'approche *holistique* diffère de l'approche dite *exhaustive* adoptée par les technologies actuelles et les travaux de recherche sur l'accessibilité web [Harper and Chen 2012; Al-Khalifa 2012], qui sont tournés vers l'amélioration de la couverture de restitution (de la disponibilité de l'information), c'est-à-dire la quantité d'informations restituées par rapport aux informations présentes dans le contenu d'origine. Ce type d'approche exhaustive de conception semble moins performant et plus demandeur cognitivement parlant qu'une approche holistique dans le cas des pages web [Giraud 2014]. Dans le cas de nos expérimentations, mettre en œuvre une approche holistique est passé par l'étude des caractéristiques de la MFM et de son utilité du point de vue de la tâche de compréhension, ce qui a permis de proposer une fonction d'accès à la structure globale qui s'est avérée plus que pertinente. De plus, en ce qui concerne la prise en compte des utilisateurs, ces travaux pointent du doigt le fait que les utilisateurs non-voyants montraient des préférences hétérogènes et possédaient des capacités de traitement de l'information orale plus performantes que leurs homologues non-déficients.

Enfin, en ce qui concerne les futurs travaux sur la restitution de la mise en forme, nous pensons qu'il sera intéressant d'explorer la restitution de la dimension fonctionnelle de la mise en forme, en phase avec la tâche réalisée. Cela implique donc d'étudier quelles sont les fonctions utilisées en lecture visuelle pour une tâche donnée (ex. recherche d'information), et dans un second temps d'étudier les possibilités de restitution non-visuelles de cette fonction ; on pourrait citer l'exemple dans la littérature d'une restitution de lecture en diagonale [Ahmed et al. 2012]. C'est ce que nous avons fait avec la restitution de la structure globale pour la compréhension. La pertinence de cette approche est liée au fait que la MFM ait émergé de besoins fonctionnels des lecteurs, qui a fait que les phénomènes de MFM structurant le contenu existent en partie pour répondre à ces besoins ; on en

conclut que pour que les documents soient accessibles (en termes d'utilisation) les fonctions de lecture doivent être restituées. Cette restitution ne doit pas être une reproduction des logiques de structuration visuelles à l'oral, mais une restitution « transmodale » [Maurel 2004] de l'information, c'est-à-dire transposant l'information et les fonctions d'une modalité vers une autre. La théorie de la MFM et du MAT sont un point d'entrée permettant de formaliser ces informations et fonctions, dont la restitution doit ensuite être conçue à travers une approche globale (ou *holistique*).

Cette dernière réflexion implique que les technologies actuelles doivent se concentrer sur ce type de logique de restitution fonctionnelle, en prise avec le sens de la MFM et son utilisation, plutôt que de proposer une liste extensive de fonctions génériques permettant d'accéder aux propriétés visuelles des textes (ex. accéder aux liens hypertextes, accéder aux énumérations). De tels systèmes comme les lecteurs d'écran semblent d'ailleurs être perçus comme difficiles d'utilisation, d'après les retours qualitatifs de l'expérimentation auprès de sujets non-voyants, et rajouter de nouvelles fonctions de lecture nécessitant l'apprentissage de nouveaux raccourcis clavier pourrait se faire au détriment de l'utilisabilité, voir n'être simplement pas utilisées au regard des nombreuses commandes déjà existantes [Freedom Scientific 2014]. Nous pensons que ces systèmes pourraient être plus utilisables et rendre ainsi les contenus plus accessibles en adoptant l'approche holistique préconisée ci-avant. Cependant leur cadre d'utilisation dépasse celui de l'interaction avec un ensemble de documents, et l'interaction avec le système d'exploitation et les applications installées ajoute de la complexité à l'utilisation dont il paraît difficile de se départir. Malgré ces difficultés, restituer la MFM et les fonctions associées lors de l'accès aux documents à travers ces systèmes semble essentiel dans une optique de rendre les contenus plus accessibles, en adoptant une approche holistique basée sur le constat que l'utilisation des documents sert différents objectifs pour lesquels la MFM joue un rôle prépondérant.

Nos travaux et les résultats obtenus pourraient également contribuer à la recherche sur le plan des méthodes et métriques employées. Tout d'abord, concernant les mesures, nous pensons qu'il est particulièrement intéressant de mesurer la charge cognitive pour ce type de tâche de lecture de document, afin de donner un éclairage complémentaire aux performances. En effet, bien souvent les utilisateurs non-voyants parviennent à réaliser les mêmes tâches (ex. comprendre, rechercher une information) mais avec un niveau d'efficacité et de satisfaction moins bon en raison du coût cognitif important associé. Ainsi, la charge cognitive se trouve être en lien direct avec une problématique plus générale de fatigue mentale des déficients visuels (voir Chapitre II7). Ensuite, la mesure de la désorientation semble être un indicateur prometteur, qui permet d'évaluer l'aisance de navigation des sujets, et se trouve également être particulièrement pertinente dans le cadre de l'accès non-visuel aux contenus, étant donné l'aspect intrinsèquement exploratoire de ce cadre (voir 2 pour plus de détails sur ce dernier point). Les mesures objectives de la désorientation retrouvées dans la littérature sont basées sur la navigation entre les pages web, au regard de l'ensemble des pages d'un site web ; la définition de métriques spécifiques à la désorientation intra-document pourrait constituer un champ de recherche pertinent dans le cas de la déficience visuelle, au regard de nos résultats. Enfin, les résultats complémentaires obtenus dans les expérimentations avec les non-déficients et les non-voyants montrent qu'il est primordial de pouvoir effectuer des évaluations auprès de la population cible, malgré les difficultés qu'il peut y avoir à trouver des volontaires. Cette difficulté nous a contraint dans notre cas à adopter un protocole expérimental différent avec des passations en intra-sujets pour les non-voyants, ce qui a engendré certaines complications et des

effets liés au matériel (notamment pour la première expérimentation qui utilisait 3 documents différents), n'ayant cependant pas eu d'influence sur les résultats principaux.

En conclusion, ces résultats montrent la pertinence de notre approche globale de la restitution de la MFM dans le cadre d'une tâche de compréhension de document ; de futures études pourront étendre la recherche autour de la restitution de la MFM à d'autres tâches, comme la recherche d'information, ou encore l'interaction avec champs d'un formulaire sur une page web. Les résultats obtenus contribuent donc en premier lieu à la conception des systèmes d'accès non-visuel aux documents, en général, et en particulier lorsqu'ils sont utilisés à des fins d'apprentissage. Cependant, de futures études devront être menées pour évaluer le bénéfice de la restitution de la MFM dans d'autres cadres d'utilisation des documents, comme la recherche d'information, ou la prise de notes pour résumer un contenu, par exemple. Étant donné notre approche globale de la restitution de la MFM, de futurs travaux devront évaluer plus précisément le rôle de chaque architecture textuelle et chaque fonction restituée ; par exemple l'impact de la spatialisation des blocs de texte. Enfin, de futures études devront également préciser comment utiliser de manière optimale les modalités non-verbales pendant la lecture, et comment intégrer l'ensemble de ces résultats au sein de systèmes avec une portée plus large, comme les lecteurs d'écrans.



# Conclusion générale et perspectives

Ces travaux se sont nourris d'outils, de méthodes et de résultats provenant de nombreux domaines scientifiques différents allant de la linguistique à l'interaction homme-machine, en passant par la psychologie cognitive ou encore la psychologie des apprentissages. Notre objectif principal était de mettre ces différents domaines à contribution pour concevoir et évaluer plusieurs restitutions des architectures textuelles véhiculées par la mise en forme, et ce, dans une approche holistique de l'accessibilité. Nos résultats montrent la pertinence d'une telle approche pour améliorer l'accessibilité des contenus textuels aux déficients visuels. Cependant, ils ne représentent qu'un premier pas dans l'étude de la restitution de la MFM dans des modalités non-visuelles.

Ci-après nous synthétiserons les différentes contributions théoriques et pratiques de nos travaux, ainsi que les perspectives pour ces derniers à moyen et long termes.

## 7. Synthèse des contributions

---

---

### 7.1. Contributions théoriques

---

Le Modèle d'Architectures Textuelles (MAT) permet de définir ce qu'est la mise en forme signifiante des textes, appelée Mise en Forme Matérielle (MFM) : le produit d'intentions auteuriales d'informer le lecteur quant à l'organisation du texte. Les marques typographiques et dispositionnelles de MFM dans les textes ont donc une sémantique ; elles représentent l'expression d'architectures textuelles, c'est-à-dire d'ensembles d'objets textuels (ex. paragraphe, titre) en relation (ex. relation de composition). La propriété de la MFM à pouvoir être développée discursivement sous forme de métaphrases nous a permis de définir une méthodologie de formalisation, qui constitue à la fois une opérationnalisation des principes énoncés par le MAT mais aussi leur extension. On y définit, entre autres, les différents axes de description des objets textuels : leurs statuts et leurs propriétés locales, ainsi que les relations et phénomènes transverses dans lesquels ils sont impliqués. L'autre contribution de cette partie de nos travaux est la production d'un modèle des conventions de MFM, c'est-à-dire une manière de représenter les utilisations de la MFM à travers différents domaines de connaissances ou même de manière universelle. Ce modèle a pour but de permettre de capitaliser les conventions partagées de MFM qui auront été formalisées, afin de proposer à terme une formalisation de plus en plus automatique. Nous avons d'ailleurs montré qu'à l'aide de techniques d'apprentissage automatique, une fois définis les indices typo-dispositionnels et lexico-syntaxiques des conventions à formaliser, leur annotation automatique était réalisable avec un taux d'erreur de l'ordre de 10 à 20%. Une formalisation exhaustive de toutes les utilisations des marques de MFM semble toutefois complexe, étant donné que chaque auteur est libre de définir ses propres conventions mais aussi libre de faire des erreurs de cohérence dans leur utilisation. L'intervention d'un annotateur humain dans le processus de formalisation reste donc nécessaire. D'un point de vue plus général, cette méthodologie met en avant une dimension quasi-universelle des textes écrits assez mal cernée, et souvent ignorée des technologies d'assistance, qu'est la MFM.

Nous avons vu dans le Chapitre II que les technologies d'assistance utilisées par les déficients visuels pour accéder aux documents numériques, principalement les lecteurs d'écran, ne restituent que très

partiellement les architectures textuelles. La sémantique correspondante est justement peu présente dans les métadonnées des documents, malgré l'existence et l'évolution de normes d'accessibilité. De plus, les technologies d'assistance adoptent une approche de restitution dite *exhaustive*, qui consiste à rendre accessible (ou plutôt *disponible*) un maximum des informations présentes dans le document (excluant les architectures textuelles qui sont très peu retranscrites dans les métadonnées) afin que les utilisateurs déficients visuels puissent accéder, supposément, aux mêmes informations que leurs homologues voyants. Notre position est qu'une telle approche est pénalisante en termes cognitifs et en termes d'utilisabilité, et qu'il serait pertinent de passer d'un tel paradigme d'accessibilité au sens de la disponibilité de l'information, à un paradigme d'accessibilité au sens de l'utilisabilité et de l'efficacité cognitive des interactions pour accéder à l'information. Dans ce cadre, nous prôtons une approche holistique de conception des interactions, dans une perspective d'ergonomie cognitive, qui, à la manière de l'*user-centered design*, prenne en compte les utilisateurs et les tâches qu'ils réalisent.

L'étude en situation réalisée auprès d'élèves déficients visuels inclus en classe ordinaire (Chapitre II7) a montré que ces derniers souffrent régulièrement de fatigue mentale dans leurs activités scolaires et ont des difficultés à construire une représentation globale des contenus, quelle que soit la modalité de présentation des informations. Ces résultats corroborent le fait que la séquentialité des modes non-visuels d'accès à l'information ne permet pas de filtrer les informations non-pertinentes, entraînant ainsi un surcoût cognitif et attentionnel, et que cette séquentialité pénalise également la construction de représentations mentales globales du document qui, on le sait empiriquement, sont importantes dans une tâche de compréhension et d'apprentissage. Les modes non-visuels d'accès à l'information ont ainsi des caractéristiques contraignantes ; leur étude semble cependant nécessaire à la conception d'interactions et d'alternatives permettant de compenser ces contraintes.

L'étude que nous avons réalisée au Chapitre IV pointe justement l'existence de différentes modalités et méthodes de restitution de l'information dans le cadre de l'accès à des documents textuels, afin d'optimiser le traitement cognitif des informations restituées. Par exemple, supprimer les informations non-pertinentes avec les objectifs de la tâche permet d'alléger la charge cognitive ; l'utilisation de modalités non-verbales peut être plus pertinente que le verbal pour véhiculer une information en situation de charge cognitive, et permet dans tous les cas d'alléger les contraintes temporelles de l'accès. Ainsi, nous préconisons de concevoir l'accès non-visuel aux contenus de manière holistique, en prenant en compte les objectifs de la tâche et les caractéristiques des modalités de restitution, afin de répondre au mieux aux questions *que restituer ?* (i.e. que filtrer ?) *comment (et quand) restituer ?* (i.e. quelles modalités ? quelle reformulation ? à quel moment pendant la lecture ?). Dans notre cas, l'aspect holistique s'est traduit par l'étude des caractéristiques de la MFM (de l'information à restituer), celles de la tâche de compréhension des documents écrits (objectif de la restitution), et enfin de celles des modalités utilisables pour restituer la MFM. Les réflexions et études produites sur ces différents points contribuent de manière générale à la conception d'interactions, dans le cadre de l'accès aux contenus pour les déficients visuels.

Un autre aspect théorique méritant d'être évoqué est la restitution de la dimension fonctionnelle de la MFM. Les capacités du système visuel sont mises à contribution pour permettre au lecteur d'exploiter les propriétés visuelles des textes, lui offrant ainsi un ensemble de fonctions de lecture adaptées à différentes tâches et contextes. On peut par exemple citer l'accès sélectif à un type de phénomène particulier (ex. rechercher un titre sur la page) ou encore l'anticipation de la fin des

segments de MFM à l'aide de la vision périphérique. Les rôles de telles fonctions en termes cognitifs ou pour la réalisation de différentes tâches utilisant des textes sont autant d'arguments en faveur de l'étude de leur restitution dans des modes non-visuels. Nous avons ici proposé une restitution de la fonction d'anticipation de la fin des blocs à l'aide de son spatialisé ; de futures études devraient permettre de dresser un panorama plus exhaustif des fonctions existantes et de leurs rôles, afin d'évaluer l'intérêt et la faisabilité de leur restitution.

Les résultats obtenus suite aux diverses évaluations de la restitution de la MFM soulignent l'importance de la mesure de la charge cognitive, qui permet de mettre en perspective les performances des sujets avec les ressources mentales qu'ils ont engagées pour les atteindre. Cela nous semble d'autant plus pertinent dans un cadre non-visuel étant donné la propension de ces modes à générer une surcharge cognitive. Toujours concernant les mesures, nous avons pu mesurer la compréhension à travers des questionnaires (inspirés d'études existantes en psychologie des apprentissages) évaluant notamment la représentation globale du contenu abordé ; dans le cadre de l'accès à des contenus pédagogiques numériques (ex. les interfaces web des MOOC<sup>127</sup> lorsque les cours sont consultés en ligne), cette métrique est rarement utilisée et pourrait contribuer à une évaluation plus complète de ce type d'outils. Enfin, nous avons pu mesurer le ressenti des sujets en termes d'orientation dans les documents (où suis-je ? d'où viens-je ?) ; le caractère exploratoire du « non-visuel » qui ne propose pas, contrairement au visuel, de forme d'ensemble unifiée, nous semble susceptible de perdre les utilisateurs dans l'interaction. Ainsi, cette métrique, initialement utilisée dans le cadre de la navigation hypertexte, nous semble particulièrement pertinente pour l'évaluation de futures interactions non-visuelles.

## **7.2. Implications et contributions pratiques**

---

L'étude en situation des difficultés et besoins des élèves déficients visuels inclus en classe ordinaire en matière d'accessibilité à l'information pédagogique a permis de mettre en évidence les outils et matériels à dispositions des élèves déficients visuels scolarisés en inclusion, et plus généralement a permis de dépeindre les contraintes de l'accès à l'information (pédagogique) pour les élèves. Les deux résultats principaux de l'étude ont porté sur la forte charge cognitive liée à la nécessité de maintenir une attention continue, et leur difficulté à pouvoir se représenter globalement un contenu pour le comprendre notamment à cause de l'aspect distribué des sources d'informations (discours oral du professeur, bloc-notes braille, documents en braille, auxiliaire de vie scolaire). L'ensemble des résultats, ayant permis de mettre en lumière les difficultés et besoins des apprenants déficients visuels, a servi à piloter de façon rationnelle les choix de conception concernant les stratégies de restitution évaluées. De plus, le portrait très général de la situation des apprenants déficients visuels en classe ordinaire, que l'étude a permis de dresser, renseigne sur leurs besoins et contraintes réelles et à ce titre pourrait servir à la formation des éducateurs spécialisés ou des professeurs de filière générale, pourrait contribuer à conception de dispositifs d'assistance logiciels et/ou matériels à destination des élèves, ou encore permettre de faire évoluer la transcription de documents braille papier.

La méthodologie de formalisation de la MFM développée permet de décrire potentiellement n'importe quelle architectures textuelles et donc n'importe quelles marques de MFM présentes

---

<sup>127</sup> Massive Open Online Course ; plateformes numériques de cours.

dans le texte, sous réserve qu'elles soient correctement interprétées. Le processus de formalisation défini est pour le moment réalisé manuellement ; cependant ce dernier a été défini dans une optique d'automatisation. De plus, une première implémentation *partielle* du processus de formalisation, à l'aide de techniques d'apprentissage automatique, a montré son efficacité sur un corpus d'articles scientifiques aux mises en forme hétérogènes. L'annotation réalisée reste cependant proche de la sémantique des langages actuels de balisage comme le HTML, et l'extension de cette approche algorithmique à une annotation plus fine et plus riche des phénomènes de MFM (ce pour quoi a été développé notre méthodologie de formalisation), reste une perspective.

READ, la plateforme logicielle que nous avons développé, permettra une fois publiée en open source la conception et l'évaluation de nouvelles interactions non-visuelles avec les documents numériques, et ce, à moindre coût. Les résultats de nos évaluations montrent qu'il existe une marge de progression quant à l'efficacité de ce type d'interactions ; par exemple, une stratégie de restitution de la MFM permet d'augmenter les scores de compréhension de plus de 50% par rapport à une stratégie classique. Cela montre que les méthodes classiques pour l'accès non-visuel aux documents peuvent être améliorées. De plus, ces dernières exploitent assez peu les possibilités des modalités non-visuelles, qui pourraient permettre aux déficients visuel d'accéder à de nouvelles fonctions de lecture, comme l'anticipation de la fin des objets textuels, qui était ici restituée par du son spatialisé. La combinaison de différentes modalités non-visuelles pour l'accès aux contenus semble d'autant plus pertinente que chaque modalité présente des avantages et inconvénients différents.

Les évaluations réalisées de différentes stratégies de restitution de la MFM ont plusieurs implications pratiques pour les technologies d'assistance. L'accès à une table des matières, soit une vue transverse de la structure organisationnelle du document, est essentiel pour la compréhension du contenu. Cette vue permet également de diminuer la charge cognitive ; ce résultat n'a cependant été validé que pour les utilisateurs voyants. On suppose que les capacités de traitement et l'habitude des stimuli audio ont permis aux sujets non-voyants de traiter de manière optimale le contenu, quelle que soit la stratégie de restitution employée. Cependant, la littérature montre que des informations redondantes et/ou non pertinentes avec les objectifs d'une tâche de compréhension pénalisent l'apprenant, ce qui laisse penser que la restitution de méta-informations dans les documents doit impérativement être pertinente quant aux objectifs de la tâche et non-redondante avec le discours. La restitution des indices locaux de MFM par des sons non-verbaux a permis de diminuer la désorientation des sujets pendant leur parcours du document ; ce résultat est particulièrement intéressant à l'heure où le verbal est modalité principale d'accès aux contenus. L'expérimentation auprès de sujets voyants a montré que la stratégie employée par les lecteurs d'écrans pour l'accès non-visuel n'était pas plus performante que l'accès à un fichier audio, ce qui souligne la marge de progression des technologies. Enfin, un résultat secondaire par rapport aux objectifs principaux des expérimentations, mais intéressant de par sa récurrence, est la grande diversité en termes d'habitudes et de préférences des utilisateurs déficients visuels. Ce constat nous amène à penser que les technologies d'assistance doivent être les plus paramétrables possibles afin de pouvoir s'adapter à cette hétérogénéité ; la question de l'accessibilité des interfaces de configuration revêt alors d'une importance particulière.

D'une manière plus générale, étudier la restitution de la MFM dans le cas de documents pédagogiques semble avoir des implications pratiques plutôt directes, dans la mesure où les manuels

scolaires tendent à être de plus en plus utilisés en version numérique, et où l'utilisation de tablettes tactiles et/ou d'ordinateurs en cours semble se démocratiser.

### **7.3. Limites des travaux**

---

Une première limite de ces travaux réside dans la faisabilité d'une formalisation de la MFM à grande échelle. En effet, nos résultats montrent que restituer la MFM est important pour comprendre un contenu textuel efficacement, dans le cadre d'un accès non-visuel. Cependant, nous avons également vu que la formalisation de la MFM ne pouvait pas être, pour le moment, totalement automatique. Au vu des difficultés de mise en place des normes actuelles d'accessibilité, dont l'application ne peut également être que partiellement automatique, on imagine qu'étendre les normes existantes, ou en créer de nouvelles, risque fort d'observer le même phénomène. Pour des raisons de coût, de complexité ou encore de manque de formation des développeurs et créateurs de contenus, la formalisation de la MFM risque donc de ne pas pouvoir être généralisée à des volumes importants de contenus numériques, si cette dernière n'est pas totalement automatique ; auquel cas elle pourrait être gérée par les technologies d'assistance indépendamment des contenus numériques en eux-mêmes.

Une deuxième limite concerne nos choix méthodologiques au regard de l'évaluation de la restitution de la MFM. Nous avons en effet choisi une approche globale, qui consistait à définir et comparer plusieurs stratégies, restituant plus ou moins la MFM. Nous voulions en effet montrer dans un premier temps la validité de notre hypothèse générale, qui était que restituer la MFM des textes dans des modes non-visuels pouvait engendrer certains bénéfices pour les utilisateurs. Le but était donc, dans un premier temps, d'observer l'effet global d'une restitution (conçue comme étant) optimale de la MFM et de la comparer à d'autres stratégies. Cependant, ces choix méthodologiques font qu'on ne peut cerner les effets potentiels des différentes informations de MFM restituées, indépendamment les unes des autres. De la même façon, on ne peut conclure sur les effets de chacune des modalités utilisées pour la restitution, relativement à une restitution purement verbale (lecture de texte par synthèse vocale). Ainsi, nos principaux résultats montrent que restituer la MFM améliore la compréhension du contenu, et dans certains cas permet de diminuer la charge cognitive ; cependant de futures études seront nécessaires pour comprendre plus finement les mécanismes ayant engendré ces résultats : informations restituées, et modalités de restitution de ces informations. Nous avons validé notre hypothèse principale, évoquée en introduction de ce manuscrit, qui était que la restitution de la MFM peut être bénéfique. Toutefois, l'influence dans les résultats obtenus de chacune des architectures textuelles restituées, de chacune des fonctions de lecture restituées et interactions proposées, et de chacune des modalités de restitution utilisées, reste à préciser.

Enfin, les résultats obtenus ne permettent pas de conclure directement sur l'efficacité des lecteurs d'écrans, principale technologie logicielle d'accès aux contenus utilisée par les déficients visuels, étant donné qu'ils incluent des fonctions de navigation qui n'étaient pas reprises dans les stratégies de restitution évaluées. Toutefois, les résultats permettent de dire qu'indépendamment de ces fonctions de navigation, leur stratégie de restitution des contenus textuels (méta-informations restituées, et modalités utilisées) est moins efficace que d'autres stratégies restituant la MFM.

---

## 8. Perspectives : futurs travaux

---

Une première piste d'étude, qui pourrait être partiellement abordée avec les données récoltées, serait une analyse fine des stratégies de navigation que les sujets ont employées pour le parcours des documents pendant les expérimentations. Une caractérisation fine des comportements de navigation, à la fois par une étude empirique et une observation des données, permettrait de définir des fonctions de navigation plus pertinentes avec la tâche de navigation non-visuelle dans un document. On peut citer comme exemple de ce type d'approche [Rohani Ghahari et al. 2012] qui ont développé une méthode de rétro-navigation des liens hypertextes visités sur la base d'observations des comportements et difficultés des utilisateurs de lecteurs d'écran.

Les perspectives les plus directes de nos travaux portent sur l'extension de notre approche de restitution de la MFM à d'autres cadres expérimentaux. Une première piste de travaux consisterait à étudier les effets de chacune des composantes de la restitution pour les stratégies définies, indépendamment les unes des autres. Par exemple, un protocole similaire aux dernières expérimentations pourrait être mis en place mais comparant une stratégie contrôle avec une autre, qui inclurait uniquement des icônes audio pour signaler la présence d'objets importants. Toujours dans cette optique il serait d'ailleurs intéressant d'évaluer les effets de la restitution de la fonction d'anticipation réalisée avec du son spatialisé, sachant que le pendant visuel de cette fonction de lecture permet d'optimiser la construction d'une représentation du contenu. Ensuite, l'évaluation des effets de la restitution de la MFM pour d'autres tâches comme la recherche d'informations (ou des tâches plus complexes incluant la navigation entre plusieurs documents couplée à de la prise de notes), sur d'autres types de documents comme les pages Web (dont la MFM aurait été formalisée), voir sur d'autres dispositifs comme les *smartphones* dont l'utilisation va croissante chez les déficients visuels<sup>128</sup>, constituent autant de pistes intéressantes pour de futurs travaux. Enfin, concevoir une stratégie de restitution de la MFM incluant le braille, une page en relief ou un dispositif vibrant comme modalités de restitution ouvrirait d'autres perspectives pour une restitution multimodale ; il nous semble d'ailleurs pertinent d'explorer l'utilisation d'autres modalités dont les caractéristiques pourraient s'avérer pertinentes pour mieux compenser les contraintes des modes non-visuels d'un point de vue cognitif et fonctionnel.

Plusieurs perspectives se dessinent également pour la méthodologie de formalisation de la MFM développée et le modèle de représentation des conventions de MFM qu'elle inclut. Une première serait leur application pour la modélisation de conventions de MFM comme l'ordre de lecture, des conventions liées à certains formats ou genre de documents ou encore à certains domaines de connaissances. Ce travail permettrait d'évaluer le pouvoir descriptif du modèle et de l'améliorer, mais nécessiterait tout d'abord la transposition du modèle en ontologie afin de le standardiser et de permettre sa diffusion. Dans la même optique, le développement d'une DTD (grammaire de balisage) standard serait nécessaire pour l'annotation des documents et donc l'incorporation des différentes informations de MFM aux fichiers. Ce dernier point soulève un ensemble de questions et de difficultés relatives aux standards d'accessibilité déjà existants. Ayant montré dans nos travaux l'impact de la restitution de la MFM pour les utilisateurs déficients visuels, il semblerait pertinent dans l'avenir que les documents numériques puissent incorporer les métadonnées correspondantes.

---

<sup>128</sup> Cf. Étude réalisée par WebAIM <http://webaim.org/projects/screenreadersurvey5/#mobiledesktopusage>

Cependant, il existe déjà plusieurs standards d'accessibilité web (WCAG et WAI-ARIA) qui sont actuellement assez peu appliqués et semble-t-il coûteux à mettre en place, sans compter les autres types de métadonnées existantes qu'il est possible d'incorporer aux documents numériques (liés à d'autres modèles ontologiques comme le Dublin-Core). Ainsi, l'annotation de la MFM devrait à terme pouvoir être réalisée de la manière la plus automatique possible afin d'en minimiser les contraintes pour favoriser sa mise en place par les créateurs et gestionnaires de contenus numériques. Techniquement parlant, les langages de balisage (utilisés pour la majorité des documents numériques) et l'intégration notable de balises plus orientées vers une annotation sémantique dans la version 5 de HTML permettent aisément d'incorporer les informations de MFM aux documents numériques. Cependant, il serait bon d'étudier dans quelle mesure l'annotation de la MFM pourrait réutiliser ou étendre les standards d'accessibilité existants, et notamment le standard ARIA qui propose certains marquages sémantiques des segments des pages web, toujours dans l'optique de favoriser son implémentation dans la nature. Plus généralement, une étude des recouvrements (redondances) sémantiques entre la modélisation de la MFM et les différents formats de fichiers et standards de métadonnées (dont ceux pour l'accessibilité), semble être une étape importante pour rendre la MFM accessible de la manière la plus universelle possible.

D'un point de vue théorique, la modélisation proposée de la MFM offre un nouvel angle d'analyse du contenu des documents, qui pourrait compléter la structure rhétorique ainsi que d'autres types de métadonnées dans le cadre d'extraction de connaissances, d'analyse du discours ou de gestion de bases de documents (bibliothèques numériques). Pour revenir à la situation de lecture visuelle pour laquelle la MFM est initialement prévue, notre modèle pourrait contribuer à l'étude des effets des signaux visuels sur les lecteurs, en offrant un point de vue complémentaire sur la sémantique la MFM et surtout en proposant une méthodologie pour sa formalisation. En effet, un modèle semblable au notre existe déjà, et présente les signaux dans une perspective plus en lien avec le discours du contenu, avec pour but de prédire leurs effets sur les lecteurs [Lemarié et al. 2008] ; ces deux modèles pourraient s'avérer complémentaires dans l'étude de la MFM, et une fusion pourrait être envisagée.

Méthodologiquement parlant, que ce soit pour la formalisation de la MFM ou sa restitution, notre approche pourrait être appliquée à des objets textuels particuliers comme par exemple les formules mathématiques, dont l'inscription spatiale et les marques syntaxiques pourraient être caractérisées dans le formalisme de la MFM (statuts, propriétés, relations) et restituées de manière similaire aux segments textuels ici étudiés. À noter que l'on retrouve des approches similaires à la nôtre pour la restitution des formules mathématiques, utilisant des modalités non-verbales [Bates and Fitzpatrick 2010]. De la même façon, la modélisation et la restitution des tableaux pourrait être réalisée à l'aide des méthodes et outils présentés dans ce manuscrit, et ainsi compléter une littérature existante portant notamment sur la navigation entre cellules. Notre approche pourrait également être étendue à la restitution d'autres fonctions de lecture, en prise avec la MFM, ou la restitution d'autres composantes de la MFM des documents comme par exemple sa portée esthétique.

On peut aussi évoquer des perspectives plus indirectes de l'étude de la mise en forme des textes. La première serait de permettre aux transpositeurs braille de mieux analyser les documents qu'ils doivent traiter et peut être de mieux prendre en compte la sémantique visuelle d'origine pour la retranscrire avec un minimum de pertes. La seconde concerne directement la MFM en tant qu'objet visuel ; l'étude de ses mécanismes de formation et de sa sémantique pourrait donner lieu à la

création de phénomènes de mise en forme particuliers ou encore à l'optimisation voire l'émergence de fonctions de lecture permises par la MFM. Comme souligné par [Meiselwitz et al. 2009] (p.216), les études et outils initialement destinés à compenser une déficience bénéficient souvent à une population plus large et dans d'autres cas d'utilisation que ceux pour lesquels ils ont été conçus initialement.

Nos travaux ont permis non seulement de rendre disponibles les architectures textuelles des documents numériques jusque-là ignorées, en les formalisant, mais ont également montré qu'une restitution *fonctionnelle* de ces informations (dans des modalités non-visuelles) est possible, et bénéfique aux utilisateurs déficients visuels. Pour concevoir une restitution fonctionnelle, qui puisse être efficiente cognitivement parlant, nous avons mis en place une démarche *holistique* pour la conception des stratégies de restitution, prenant en compte non seulement l'information à restituer, mais aussi le contexte de restitution (objectifs de la tâche, caractéristiques des utilisateurs, etc.). Notre positionnement est qu'une telle démarche est nécessaire pour rendre les technologies d'assistance plus efficaces et utilisables, au-delà du simple critère binaire d'information accessible ou non, de tâche faisable ou non, qui définit bien souvent la portée du terme *accessibilité* dans un contexte numérique. Nous espérons ainsi avoir modestement contribué à une définition de l'accessibilité plus large, et plus en prise avec l'utilisabilité des technologies pour les déficients visuels, qui pourrait dans le futur contribuer également à faire évoluer les travaux sur le sujet.



# Références

- ABRAS, C., MALONEY-KRICHMAR, D. AND PREECE, J., 2004. User-Centered Design. In William Sims Bainbridge, ed. *Berkshire Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications, pp. 445–456.
- AHMED, F., BORODIN, Y., PUZIS, Y. AND RAMAKRISHNAN, I.V., 2012. Why Read if You Can Skim : Towards Enabling Faster Screen Reading. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility - W4A2012*. ACM.
- AHMED, T., 2015. Privacy Concerns and Behaviors of People with Visual Impairments. In *CHI'15*. pp. 3523–3532.
- AHUJA, J.S. AND WEBSTER, J., 2001. Perceived disorientation: An examination of a new measure to assess web design effectiveness. *Interacting with Computers*, 14(1), pp.15–29.
- AL-KHALIFA, H.S., 2012. WCAG 2.0 Semi-automatic Accessibility Evaluation System: Design and Implementation. *Computer and Information Science*, 5(6), pp.73–87.
- ALLPORT, D. A, ANTONIS, B. AND REYNOLDS, P., 1972. On the division of attention: a disproof of the single channel hypothesis. *The Quarterly journal of experimental psychology*, 24(2), pp.225–235.
- AMMAR ALI, A., 2006. *Analyse des explorations haptiques de formes pour la conception d'un dispositif de suppléance perceptive dédié aux personnes aveugles*.
- ANTONACOPOULOS, A., CLAUSNER, C., PAPADOPOULOS, C. AND PLETSCHACHER, S., 2011. Historical document layout analysis competition. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, (August 2015), pp.1516–1520.
- ARCHAMBAULT, D., 2010. *Interaction et usages des modalités non visuelles, accessibilité des contenus complexes*.
- ARCHAMBAULT, D., 2009. Non Visual Access to Mathematical Contents : State of the Art and Prospective. In *Proceedings of the WEIMS Conference 2009 (The Workshop on E-Inclusion in Mathematics and Science)*.
- ARNE, J., AXELSSON, M., BERGENHOLM, E. AND RYBING, J., 2008. Skim reading of audio information. In *Swedish Language Technology Conference - SLTC'08*. pp. 23–24.
- ASAKAWA, C. AND TAKAGI, H., 2003. Maximum Listening Speed For The Blind. In *Proceedings of the 2003 International Conference on Auditory Display*. Boston,MA, pp. 276–279.
- AUSTIN, J., 1975. *How to do things with words*,
- AWDE, A., 2009. *Techniques d'interaction multimodales pour l'accès aux mathématiques par des personnes non-voyantes*.
- BACCINO, T. AND PYNTE, J., 1991. Le codage spatial dans la lecture. *L'année psychologique*, 91(2), pp.231–245.
- BACH-Y-RITA, P., TYLER, M.E. AND KACZMAREK, K. A., 2003. Seeing with the Brain. *International Journal of Human-Computer Interaction*, 15(2), pp.285–295.
- BADDELEY, A., 2003. Working memory: looking back and looking forward. *Nature reviews neuroscience*, 4(10), pp.829–839.
- BADDELEY, A.D. AND HITCH, G., 1974. Working-memory. *Science*, 255(5044), pp.556–559.

- BATEMAN, J., 2008. *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*, Palgrave Macmillan.
- BATES, E. AND FITZPATRICK, D., 2010. Spoken mathematics using prosody, earcons and spearcons. In *ICCHP 2010*. pp. 407–414.
- BAUMGARTNER, R. ET AL., 2010. Web 2.0 Vision for the Blind. In *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*.
- BECHHOFFER, S., BECHHOFFER, S., HARPER, S., HARPER, S., LUNN, D. AND LUNN, D., 2006. SADle: Semantic Annotation for Accessibility. In *The Semantic Web - ISWC 2006, 5th International Semantic Web Conference*.
- BELLIK, Y., 1995. *Interfaces multimodales : concepts, modèles et architectures*.
- BERNERS-LEE, T., HENDLER, J., LASSILA, O. AND OTHERS, 2001. The semantic web. *Scientific american*, 284(5), pp.28–37.
- BERZ, W.L., 1995. Working Memory in Music: A Theroretical Model. *Music Perception*, 12(3), pp.353–364.
- BICKFORD, J.O. AND FALCO, R.A., 2012. Technology for early Braille literacy: Comparison of traditional Braille instruction and instruction with an electronic notetaker. *Journal of Visual Impairment & Blindness*, 106(10), pp.679–693.
- BIGHAM, J.P., PRINCE, C.M. AND LADNER, R.E., 2008. WebAnywhere: a screen reader on-the-go. *Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, pp.73–82.
- BLATTNER, M., SUMIKAWA, D. AND GREENBERG, R., 1989. Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, 4(1), pp.11–44.
- BLENKHORN, P., 2000. Architecture and requirements for a windows screen reader. In *IEE Seminar on Speech and Language Processing for Disabled and Elderly People*. p. 1.
- BLOOM, B.S. AND KRATHWOHL, D.R., 1956. *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*, Longmans.
- BORODIN, Y. AND BIGHAM, J., 2008. Towards one world web with HearSay3. In *W4A'08*. ACM, pp. 5–6.
- BORODIN, Y., BIGHAM, J.P., DAUSCH, G., RAMAKRISHNAN, I. V AND BROOK, S., 2010. More than Meets the Eye : A Survey of Screen-Reader Browsing Strategies. In *W4A '10: Proceedings of the 2010 International Cross-Disciplinary Workshop on Web Accessibility*. Raleigh, USA: ACM.
- BOU-ISSA, Y., 2010. *Accessibilité aux informations visuelles des pages web pour les non-voyants*.
- BOWMAN, V., 2002. Reading between the lines: an evaluation of WindowEyes screen reader as a reference tool for teaching and learning. *Library Hi Tech*, 20(2), pp.162–168.
- BRAJNIK, G., 2008. A comparative test of web accessibility evaluation methods. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility - Assets '08*. pp. 113–120.
- BREWSTER, S. AND BROWN, L.M., 2004. Tactons : Structured Tactile Messages for Non-Visual Information Display. In *AUIC 2004*. pp. 15–23.
- BREWSTER, S., WRIGHT, P.C. AND EDWARDS, A D.N., 1995. Parallel Earcons - Reducing the Length of Audio Messages. *International Journal of Human-Computer Studies*, 43(22), pp.153–175.

- BREWSTER, S.A., WRIGHT, P.C. AND EDWARDS, A.D.N., 1993. An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. In *CHI'93*. ACM, pp. 222–227.
- BRITTON, B.K., GLYNN, S.M., MEYER, B.J. AND PENLAND, M.J., 1982. Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 74(1), p.51.
- BROOKE, J., 1996. SUS - A quick and dirty usability scale - Standard Usability questionnaire. *Usability evaluation in industry*, 189(194), pp.4–7.
- BROUWERS, A. AND OTHERS, 2015. *La recherche d'information dans un environnement hypermédia: le rôle des habiletés visuo-spatiales*. UCL.
- BROWN, A. AND HARPER, S., 2013. Dynamic injection of WAI-ARIA into web content. *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility - W4A '13*, (August), p.1.
- BROWN, L.M., BREWSTER, S. A AND PURCHASE, H.C., 2005. A first investigation into the effectiveness of Tactons. In *Eurohaptics Conference*. IEEE, pp. 167–176.
- BURTON, H., 2003. Visual cortex activity in early and late blind people. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(10), pp.4005–4011.
- BURTON, H., SNYDER, A.Z., CONTURO, T.E., AKBUDAK, E., OLLINGER, J.M. AND RAICHLER, M.E., 2002. Adaptive changes in early and late blind: a fMRI study of Braille reading. *Journal of Neurophysiology*, 87(1), pp.589–607.
- CAI, D., YU, S. AND WEN, J., 2003. *VIPS : a Vision-based Page Segmentation Algorithm*,
- CAUCHARD, F., 2008. *Empan perceptif en lecture et en recherche d'information dans un texte : influence des signaux visuels*.
- CHALI, Y., 1997. *L'expansion de texte Une approche basée sur l'explication par questions/réponses pour la génération de versions de textes*.
- CHAROLLES, M., 1997. L'encadrement du discours : UNIVERS , CHAMPS , DOMAINES ET ESPACES. *Cahiers de recherche Linguistique*, (6), pp.1–73.
- CHAUVEY, V., HATWELL, Y., VERINE, B., KAMINSKI, G. AND GENTAZ, E., 2012. Lexical references to sensory modalities in verbal descriptions of people and objects by congenitally blind, late blind and sighted adults. *PloS one*, 7(8), p.e44020.
- CHAYTOR, H.J., 1945. *From script to print: an introduction to medieval literature*, Cambridge University Press.
- CHI, M.T.H., GLASER, R. AND FARR, M.J., 1988. The nature of expertise. *Hillsdale (NJ): Lawrence Erlbaum*, 1, pp.448–451.
- CHIRAC, J., 2005. Law n°2005-102 for rights and opportunities equality. "Loi n°2005-102 pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées."
- CHRISTENSEN, L.B. AND STEVNS, T., 2012. Biblus – A Digital Library to Support Integration. In *Proceedings of the 16th international conference on Computers Helping People with Special Needs*. pp. 36–42.
- CLARK, J., 2006. To Hell With WCAG 2. *May 23, 2006*.
- COFINO, J., BARRETO, A., ABYARJOO, F. AND ORTEGA, F.R., 2014. Sonically-Enhanced Tabular Screen-Reading. In *International Technology and Persons with Disabilities Conference*. Northridge, CA.
- COIRIER, P., GAONACH, D. AND PASSERAULT, J.-M., 1996. *Psycholinguistique textuelle: une approche cognitive de la compréhension et de la production des textes*, Paris: Armand Colin.

- COMMISSION, D.R., 2004. *The Web: Access and Inclusion for Disabled People; a Formal Investigation*, The Stationery Office.
- DECLERCK, G., LENAY, C. AND KHATCHATOUROV, A., 2009. Rendre tangible le visible. *Irbm*, 30(5-6), pp.252–257.
- DEMEGLIO, M., HAKKINEN, M.T. AND KAWAMURA, H., 2002. Accessible Interface Design : Adaptive Multimedia Information System ( AMIS ). In *ICCHP 2002*. pp. 406–412.
- VAN DIJK, T.A. AND KINTSCH, W., 1983. *Strategies of discourse comprehension*, Academic Press New York.
- DOURISH, P., 2001. *Where the Action Is: The Foundations of Embodied Interaction*,
- DYSON, M.C., 2004. How physical text layout affects reading from screen. *Behaviour & Information Technology*, 23(6), pp.377–393.
- EL-GLALY, Y.N. AND QUEK, F., 2014. Digital Reading Support for The Blind by Multimodal Interaction. In *ICMI'14*. pp. 439–446.
- ERICSSON, A., NEIL, C., PAUL, F. AND ROBERT, H., 2006. *The Cambridge handbook of expertise and expert performance*, Cambridge University Press.
- ESPOSITO, F., FERILLI, S., BASILE, T. AND MAURO, N. DI, 2008. Machine Learning for digital document processing: from layout analysis to metadata extraction. *Studies in Computational Intelligence*, 90, pp.105–138.
- ETCHEVERRY, I., BACCINO, T., TERRIER, P., MARQUIÉ, J.-C. AND MOJAHID, M., 2012. Age differences in information finding tasks: Performance and visual exploration strategy with different web page layouts. *Computers in Human Behavior*, 28(5), pp.1670–1680.
- FAUCONNIER, J., SORIN, L., KAMEL, M., AUSSENAC-GILLES, N. AND MOJAHID, M., 2014. Détection automatique de la structure organisationnelle de documents à partir de marqueurs visuels et lexicaux. In *TALN -RECITAL 2014, Traitement Automatique du Langage Naturel*.
- FAYZRAKHMANOV, R., MAX, G., HOLZINGER, W., KR, B. AND BAUMGARTNER, R., 2010. A Unified Ontology-Based Web Page Model For Improving Accessibility. In *WWW 2010*.
- FEDERICI, S., BORSCHI, S. AND MELE, M.L., 2010. Usability evaluation with screen reader users: A video presentation of the PCTA's experimental setting and rules. *Cognitive Processing*, 11(3), pp.285–288.
- FERILLI, S., GRIECO, D., REDAVID, D. AND ESPOSITO, F., 2014. Abstract argumentation for reading order detection. In *Proceedings of the 2014 ACM symposium on Document engineering - DocEng '14*. New York, New York, USA: ACM Press, pp. 45–48.
- FIDEL, R. AND PEJTERSEN, A.M., 2004. From information behaviour research to the design of information systems: The Cognitive Work Analysis framework. *Information Research*, 10, pp.1–20.
- FRAISSE, P., 1984. Perception and estimation of time. *Annual review of psychology*, 35(1), pp.1–37.
- FRANCISCO-REVILLA, L. AND CROW, J., 2010. Interpretation of web page layouts by blind users. In *Proceedings of the 10th annual joint conference on Digital libraries - JCDL '10*. New York, New York, USA: ACM Press, p. 173.
- FREEDOM SCIENTIFIC, 2013. JAWS ARIA Support. , pp.1–11.
- FREEDOM SCIENTIFIC, 2014. JAWS keystrokes.

- FRICKE, J. AND BAEHRING, H., 1993. Design of a tactile graphic I/O tablet and its integration into a personal computer system for blind users. *Journal of Microcomputer Applications*, 16(3), pp.259–269.
- GADDE, P. AND BOLCHINI, D., 2014. From Screen Reading to Aural Glancing : Towards Instant Access to Key Page Sections. In *ASSETS'14*. pp. 67–74.
- GALIANO, A.R. AND BALTENNECK, N., 2007. Interactions verbales et déficience visuelle: le rôle de la vision dans la communication. *Revue électronique de psychologie sociale*, 1, pp.47–54.
- GAPENNE, O., ROVIRA, K., ALI AMMAR, A. AND LENAY, C., 2003. Tactos: Special computer interface for the reading and writing of 2D forms in blind people. *Universal Access in HCI, Inclusive Design in the Information Society*, 10, pp.1270–1274.
- GELLENBECK, E. AND STEFIK, A., 2009. Evaluating Prosodic Cues as a Means to Disambiguate Algebraic Expressions : An Empirical Study. , pp.139–146.
- GENDRON, B., 2010. Le coût des aides techniques. *Réadaptation*, 568.
- GIRAUD, S., 2014. *L'accessibilité des interfaces informatiques riches pour les déficients visuels*.
- GIRAUD, S., THEROUANNE, P., PSYCHOLOGIE, L. DE, NICE, U. DE AND ANTIPOLIS, S., 2010. Role of lexico-syntactic and prosodic cues in spoken comprehension of enumerations in sighted and blind adults. In *Multidisciplinary Perspectives on Signalling Text Organisation*. Moissac, France, pp. 17–20.
- GIRAUD, S., UZAN, G. AND THEROUANNE, P., 2011. L'accessibilité des interfaces informatiques pour les déficients visuels. In J. Dinet & C. Bastien, eds. *L'ergonomie des objets et environnements physiques et numériques*. Paris : Hermes - Sciences Lavoisier.
- GOLDMANN, S.R. AND WOLFE, M.B.W., 2001. Text comprehension: Models in psychology. *International encyclopedia of the social & behavioral sciences*, pp.15625–15631.
- GOOSE, S. AND MÖLLER, C., 1999. A 3D Audio Only Interactive Web Browser : Using Spatialization to Convey Hypermedia Document Structure. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*. ACM, pp. 363–371.
- GRANTHAM, D.W. AND WILKERSON, T.B., 1986. Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *The Journal of the acoustical society of America*, 79(6), pp.1939–1949.
- GRUBER, T., 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6), pp.907–928.
- GUERREIRO, J. AND GONÇALVES, D., 2014. Text-to-Speeches : Evaluating the Perception of Concurrent Speech by Blind People. In *ASSETS '14*. pp. 169–176.
- HA, J.H.J., HARALICK, R.M. AND PHILLIPS, I.T., 1995. Document page decomposition by the bounding-box project. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*.
- HANN, H., JEONG, Y.-S. AND JANG, W.-H., 2013. Best Practice of the DAISY Standard in Korea. In *IFLA World Library and Information Congress*.
- HANNETON, S., LENAY, C., GAPENNE, O., VERMANDEL, S. AND MARQUE, C., 1998. Dynamique de la reconnaissance de caracteres via une interface haptique. In *ARC'98*.
- HARPER, S. AND CHEN, A.Q., 2012. Web accessibility guidelines. *World Wide Web*, 15(1), pp.61–88.
- HARPER, S. AND YESILADA, Y., 2007. Web Authoring for Accessibility (WAfA ). *Journal of Web Semantics*, 5(3), pp.175–179.

- HARRIS, R., 1993. *La sémiologie de l'écriture*, CNRS éditions.
- HARRIS, Z.S., 1971. *Structures mathématiques du langage*. Dunod.
- HARTLEY, J., BARTLETT, S. AND BRANTHWAITE, A., 1980. Underlining can make a difference—sometimes. *The Journal of Educational Research*, 73(4), pp.218–224.
- HEADLEY, P.C., STREET, W.M., HRIBAR, V.E. AND PAWLUK, D.T. V, 2011. Displaying Braille and Graphics on a Mouse-like Tactile Display. In *ASSETS'11*. pp. 235–236.
- HENRY, S.L., 2005. Introduction to Web accessibility. In *Web Accessibility Initiative (WAI), World Wide Web Consortium (W3C)*.
- HERNANDEZ, N. AND GRAU, B., 2005. Détection Automatique de Structures Fines de Texte. In *Actes de la conférence Traitement Automatique de la Langue Naturelle (TALN)*. pp. 6–10.
- HEURLEY, L., 1994. *Traitement de textes procéduraux: étude de psycholinguistique cognitive des processus de production et de compréhension chez des adultes non experts*. Dijon.
- HILL, E., 1999. Braille: oral and silent reading speeds. *New Beacon*.
- HONG, S., 2012. An alternative option to dedicated braille notetakers for people with visual impairments: Universal technology for better access. *Journal of Visual Impairment & Blindness*, 106(10), p.650.
- HORNBAEK, K. AND FRØKJÆR, E., 2001. Reading of Electronic Documents : The Usability of Linear , Fisheye , and Overview + Detail Interfaces. In *CHI'01*. pp. 293–300.
- HUGDAHL, K. ET AL., 2004. Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds. *Cognitive brain research*, 19(1), pp.28–32.
- HUSSAIN, I., CHEN, L., MIRZA, H.T., MAJID, A. AND CHEN, G., 2012. Hybrid Auditory Feedback : A New Method for Mobility Assistance of the Visually Impaired. In *ASSETS'12*. pp. 255–256.
- HUSSAIN, I., CHEN, L., TURAB, H. AND HASSAN, G.C.S., 2014. Right mix of speech and non-speech : hybrid auditory feedback in mobility assistance of the visually impaired. *Universal Access in the Information Society*, March 2014, pp.1–10.
- HYÖNÄ, J., 1995. An eye movement analysis of topic-shift effect during repeated reading. *Journal of experimental psychology. Learning, memory, and cognition*, 21(5), pp.1365–73.
- INHOFF, A.W. AND WEGER, U.W., 2005. Memory for word location during reading: Eye movements to previously read words are spatially selective but not precise. *Memory & Cognition*, 33(3), pp.447–467.
- ISLAM, M.A., AHMED, F., BORODIN, Y. AND RAMAKRISHNAN, I. V, 2012. Thematic Organization of Web Content for Distraction-Free Text-to-Speech Narration Categories and Subject Descriptors. In *ASSETS*. pp. 17–24.
- JAMET, É., BETRANCOURT, M. AND BETRANCOURT, J.-F., 2008. *La compréhension des documents complexes*, Presses Universitaires de France.
- JONASSEN, D.H., 1985. *The technology of text: Principles for structuring, designing, and displaying text*, Educational Technology.
- JUST, M.A. AND CARPENTER, P.A., 1980. A theory of reading : From eye fixations to comprehension. *Psychological review*, 87(January), pp.329–354.

- JUVINA, I. AND VAN OOSTENDORP, H., 2006. Enhancing internet experience of visually impaired persons by means of dynamic highlighting and selective reading. *Information Design Journal*, 14(1), pp.71–81.
- JZOO.COM, Benefits of Bone Conduction Headphones.
- KAMEI-HANNAN, C. AND LAWSON, H., 2012. Impact of a Braille-Note on Writing: Evaluating the Process, Quality, and Attitudes of Three Students Who Are Visually Impaired. *Journal of Special Education Technology*, 27(3), pp.1–14.
- KAMEL, M. AND AUSSENAC-GILLES, N., 2008. How can document structure improve ontology learning? In *Workshop on Semantic Annotation and Knowledge Markup collocated with K-CAP*. pp. 1–8.
- KELLY, B. ET AL., 2007. Accessibility 2.0: People , Policies and Processes. In *W4A '07: Proceedings of the 2007 International Cross-Disciplinary Workshop on Web Accessibility*. pp. 138–147.
- KETTERLIN-GELLER, L.R. AND TINDAL, G., 2008. Embedded Technology : current and Future Pratices for Increasing Accessibility for All Students. *Journal of Special Education Technology*, 22(4).
- KIM, H. AND HIRTLE, S.C., 1995. Spatial metaphors and disorientation in hypertext browsing. *Behaviour & Information Technology*, 14(4), pp.239–250.
- KINTSCH, W., 1994. *The psychology of discourse processing.*, Academic Press.
- KLATZKY, R.L., MARSTON, J.R., GIUDICE, N. A, GOLLEDGE, R.G. AND LOOMIS, J.M., 2006. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of experimental psychology. Applied*, 12(4), pp.223–232.
- KONTARINIS, D. A, HOWE, R.D. AND HALL, P., 1995. Tactile Display of Vibratory Information in Teleoperation and Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 4(4), pp.387–402.
- KRATHWOHL, D.R., 2002. A Revision of Bloom’s Taxonomy. *Theory into Practice*, 41(4), pp.212–218.
- KREIMAN, J. AND PAPCUN, G., 1991. Comparing discrimination and recognition of unfamiliar voices. *Speech Communication*, 10(3), pp.265–275.
- LAFFERTY, J. AND MCCALLUM, A., 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the 18th International Conference on Machine Learning (ICML)*. pp. 282–289.
- LANE, D.M., NAPIER, H.A., PERES, S.C. AND SANDOR, A., 2005. Hidden Costs of Graphical User Interfaces: Failure to Make the Transition from Menus and Icon Toolbars to Keyboard Shortcuts. *International Journal of Human-Computer Interaction*, 18(2), pp.133–144.
- LASCARIDES, A. AND ASHER, N., 1991. Discourse relations and defeasible knowledge. In *Proceedings of the 29th. Annual Meeting of the Association for Computational Linguistics*. pp. 55–63.
- LAZAR, J., ALLEN, A., KLEINMAN, J. AND MALARKEY, C., 2007. What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users. *International Journal of Human-Computer Interaction*, 22(3), pp.247–269.
- LAZAR, J., DUDLEY-SPONAUGLE, A. AND GREENIDGE, K.-D., 2004. Improving web accessibility: a study of webmaster perceptions. *Computers in Human Behavior*, 20(2), pp.269–288.
- LEAHY, W. AND SWELLER, J., 2011. Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, 951(February), pp.943–951.

- LECOLINET, E. AND MOURET, G., 2005. TACTIBALL, TACTIPEN, TACTITAB Ou comment "toucher du doigt" les données de son ordinateur. In *Proceedings of the 17th international conference on Francophone sur l'Interaction Homme-Machine*. pp. 227–230.
- LEE, S., HONG, S.H. AND JEON, J.W., 2002. Designing a universal keyboard using chording gloves. *ACM SIGCAPH Computers and the Physically Handicapped*, (73-74), p.142.
- LEGGE, G.E., MADISON, C.M. AND MANSFIELD, J.S., 1999. Measuring Braille reading speed with the MNREAD test. *Visual Impairment Research*, 1(3), pp.131–145.
- LEMARIE, J., 2006a. *La compréhension des textes visuellement structurés - Le cas des énumérations*.
- LEMARIE, J., 2006b. La signalisation de la structure des textes comme élément facilitateur des traitements de 'information. In *JETCSIC 2008 - Journée d'Etude sur le Traitement Cognitif des Systèmes d'Information Complexes [séminaire sans actes publiés]*.
- LEMARIÉ, J., EYROLLE, H. AND CELLIER, J., 2006. Visual signals in text comprehension : How to restore them when oralizing a text via a speech synthesis ? *Computers in Human Behavior*, 22(6), pp.1096–1115.
- LEMARIÉ, J., LORCH, R.F., EYROLLE, H. AND VIRBEL, J., 2008. SARA: A Text-Based and Reader-Based Theory of Signaling. *Educational Psychologist*, 43(1), pp.27–48.
- LEMARIÉ, J., LORCH, R.F. AND PÉRY-WOODLEY, M.-P., 2012. Understanding How Headings Influence Text Processing. *Discours. Revue de linguistique, psycholinguistique et informatique.*, 10.
- LENAY, C., GAPENNE, O., HANNETON, S., MARQUE, C. AND GENOUELLE, C., 2003. *Sensory Substitution : Limits and Perspectives*,
- LEPORINI, B. AND PATERNO, F., 2008. Applying Web Usability Criteria for Vision-Impaired Users: Does It Really Improve Task Performance? *International Journal of Human-Computer Interaction*, 24(1), pp.17–47.
- LEPORINI, B. AND PATERNO, F., 2004. Increasing usability when interacting through screen readers. *Universal Access in the Information Society*, 3(1), pp.57–70.
- LEVESQUE, V. AND HAYWARD, V., 2003. Experimental evidence of lateral skin strain during tactile exploration. In *Proceedings of EUROHAPTICS*.
- LIANG, J., HA, J., HARALICK, R.M. AND PHILLIPS, I.T., 1996. Document layout structure extraction using bounding boxes of different entitles. In *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96*.
- LIN, D.Y.M., 2003. Hypertext for the aged: Effects of text topologies. *Computers in Human Behavior*, 19(2), pp.201–209.
- LORCH, R.F., 1989. Text-signaling devices and their effects on reading and memory processes. *Educational Psychology Review*, 1(3), pp.209–234.
- LORCH, R.F., CHEN, H.-T. AND LEMARIÉ, J., 2012. Communicating headings and preview sentences in text and speech. *Journal of Experimental Psychology: Applied*, 18(3), pp.265–276.
- LUC, C., 2000. *Représentation et composition des structures visuelles et rhétoriques du texte. Approche pour la génération de textes formatés*.
- LUC, C., MOJAHID, M. AND VIRBEL, J., 2002. Le modèle d'architecture de texte. In *Actes de Inscription Spatiale du Langage : Structures et Processus*. Prescott, pp. 41–50.



- LUC, C., MOJAHID, M., VIRBEL, J., GARCIA-DEBANC, C. AND PÉRY-WOODLEY, M., 1999. A Linguistic Approach to Some Parameters of Layout: A Study of Enumerations. In *Using Layout for the Generation, Understanding or Retrieval of Documents, AAAI Fall Symposium*. pp. 20–29.
- LUC, C. AND VIRBEL, J., 2001. Le modèle d'architecture textuelle: fondements et expérimentation. *Verbum*, (1), pp.103–123.
- LUITEN, J., AMES, W. AND ACKERSON, G., 1980. A meta-analysis of the effects of advance organizers on learning and retention. *American Educational Research Journal*, 17(2), pp.211–218.
- LUNN, D., BECHHOFFER, S. AND HARPER, S., 2008. A user evaluation of the SADle transcoder. In *ASSETS'08*. New York, New York, USA: ACM Press, p. 137.
- LUSK, K.E. AND CORN, A.L., 2006. Learning and Using Print and Braille: A Study of Dual-Media Learners, Part 2. *Journal of Visual Impairment & Blindness*, 100(11), pp.1–19.
- MAHMUD, J., 2007. CSurf : A Context-Driven Non-Visual Web-Browser. In *Proceedings of the 16th international conference on World Wide Web*. ACM, pp. 31–40.
- MAILLET, F., 2015. *Influence de la signalisation d'un texte et de l'imagerie visuelle sur la mémoire de localisation des mots*. Université Jean-Jaurès.
- MANN, W.C., 2006. Applications of Rhetorical Structure Theory. *Discourse studies*, 8(4), pp.567–588.
- MANN, W.C. AND THOMPSON, S.A., 1987. Rhetorical Structure theory description and constructions of text structures J. Kempen, ed. *Text*, 8(3), pp.85–96.
- MARIA, L. V. AND ANDERSON, S., 2003. Combining Speech And Earcons To Assist Menu Navigation. In *Proceedings of the 2003 International Conference on Auditory Display*. Boston, pp. 38–41.
- MATHET, Y., WIDLÖCHER, A., CNRS, U.M.R., CAEN, U. DE AND CEDEX, C., 2011. Une approche holiste et unifiée de l'alignement et de la mesure d'accord inter-annotateurs. In *TALN -RECITAL 2011*.
- MAUMET, L., 2007. L'accès à l'écrit des personnes déficientes visuelles. *Bulletin des Bibliothèques de France*, 52(3), pp.46–50.
- MAUREL, F. ET AL., 2012. Haptic Perception of Document Structure for Visually Impaired People on Handled Devices. In *Proceedings of the 4th International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2012)*. pp. 319–329.
- MAUREL, F., 2004. *Transmodalité et multimodalité écrit/oral : modélisation, traitement automatique et évaluation de stratégies de présentation des structures « visuo-architecturales » des textes*.
- MAUREL, F., 2006. Transposition automatique à l'oral des structures visuelles des textes. *Document numérique*, 9(1).
- MAUREL, F., LUC, C., MOJAHID, M., VIGOUROUX, N., VIRBEL, J. AND NESPOULOUS, J.-L., 2002. Transposition à l'oral des structures énumératives à partir de leurs paramètres formels. In *Inscription Spatiale du Langage structure et processus (ISLsp) 2002*, Toulouse, 29/01/2002-30/01/2002. IRIT, pp. 179–189.
- MAYER, R.E., 2003. *Learning and instruction*, Prentice Hall.
- MAYER, R.E., 2009. *Multimedia Learning*, Cambridge University Press.
- MCCARTHY, T., PAL, J. AND CUTRELL, E., 2013. The “voice” has it: screen reader adoption and switching behavior among vision impaired persons in India. *Assistive Technologies*, 25(4), pp.222–229.
- MCDONALD, S. AND STEVENSON, R.J., 1996. Disorientation in hypertext: The effects of three text structures on navigation performance. *Applied Ergonomics*, 27(1), pp.61–68.

- MCNAMARA, D.S., KINTSCH, E., SONGER, N.B. AND KINTSCH, W., 1996. Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and instruction*, 14(1), pp.1–43.
- MEISELWITZ, G., WENTZ, B. AND LAZAR, J., 2009. Universal Usability: Past, Present, and Future. *Foundations and Trends in Human–Computer Interaction*, 3(4), pp.213–333.
- MEYER, B.J.F., BRANDT, D.M. AND BLUTH, G.J., 1980. Use of Top-Level Structure in Text: Key for Reading Comprehension of Ninth-Grade Students. *Reading Research Quarterly*, 16(1), pp.72–103.
- MILLAR, S., 2003. *Reading by touch*, Routledge.
- MILLER, G. A., 1994. The magical number seven, plus or minus two: some limits on our capacity for processing information. 1956. *Psychological review*, 101(2), pp.343–352.
- MINATANI, K., 2012. Development of a DAISY Player That Utilizes a Braille Display for Document Structure Presentation. In *ICCHP 2012 - LCNS 7382*. pp. 515–522.
- MINISTERE DES AFFAIRES SOCIALES AND COMMISSION EVOLUTION DU BRAILLE FRANÇAIS, 1999. *Code de transcription en braille des textes imprimés*,
- MOJAHID, M., 1990. *Une architecture multi-agents en PAO pour le formatage de documents structurés*. Toulouse 3.
- MOTTI, L.G. AND BURGER, D., 2009. Adapting diagrams for daisy books. *DAISY '09*, pp.67–71.
- MURPHY, E., BATES, E. AND FITZPATRICK, D., 2010. Designing auditory cues to enhance spoken mathematics for visually impaired users. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '10*. New York, New York, USA: ACM Press, p. 75.
- MURPHY, E., KUBER, R., MCALLISTER, G., STRAIN, P. AND YU, W., 2007. An empirical investigation into the difficulties experienced by visually impaired Internet users. *Universal Access in the Information Society*, 7(1-2), pp.79–91.
- MURPHY, E., KUBER, R., STRAIN, P., MCALLISTER, G. AND YU, W., 2007. DEVELOPING SOUNDS FOR A MULTIMODAL INTERFACE: CONVEYING SPATIAL INFORMATION TO VISUALLY IMPAIRED WEB USERS. In *Proceedings of the 13th International Conference on Auditory Display*. pp. 348–355.
- NATIONAL FEDERATION OF THE BLIND, 2009. *The Braille Literacy Crisis in America : Facing the Truth, Reversing the Trend, Empowering the Blind*,
- NESPOULOUS, J.-L., DE MATTOS PIMENTA, M.-A. AND VIRBEL, J., 2002. Des objets textuels à leur traitement par le cerveau/esprit humain. In *Inscription Spatiale du Langage : structure et processus (ISLsp)*. pp. 77–89.
- NEVILE, L., 2005. User-centred accessibility supported by distributed, cumulative authoring.
- NEWELL, A. AND GREGOR, P., 2000. “User sensitive inclusive design”—in search of a new paradigm. In *CUU'00 - Conference on Univesal Accessibility*. pp. 39–44.
- NIELSEN, J., 1993. Usability inspection methods. In *ACM SIGCHI Bulletin*. pp. 28–33.
- NIELSEN, J. AND LEVY, J., 1994. Measuring usability: preference vs. performance. *Communications of the ACM*, 37(4), pp.66–75.
- NORMAN, D.A. AND DRAPER, S.W., 1986. User Centered System Design. *New Perspectives on Human-Computer Interaction*, Lawrence Erlbaum Associates Inc.
- NUNNALLY, J., 1978. Psychometric methods.

- O'HARA, K., 1996. *Towards a Typology of Reading Goals*.
- LOUDEN, H. DEN, NOORDMAN, L. AND TERKEN, J., 2002. The Prosodic Realization of Organizational Features of Texts. In *Proceedings of the speech prosody conference*. pp. 543–546.
- PAAS, F., TUOVINEN, J.E., TABBERS, H. AND VAN GERVEN, P.W.M., 2003. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38(1), pp.63–71.
- PAAS, F.G. AND VAN MERRIËNBOER, J.J.G., 1994. Paas Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), pp.351–371.
- PARENTE, P., 2008. *Clique: perceptually based, task oriented auditory display for gui applications*.
- PASCUAL, E., 1996. Integrating Text Formatting and Text Generation. In *Trends in Natural Language Generation An Artificial Intelligence Perspective*. pp. 205–221.
- PASCUAL, E., 1991. *Représentation de l'architecture textuelle et génération de texte*.
- PASCUAL, E. AND PÉRY-WOODLEY, M., 1991. Modèles de texte pour la définition. In *Premières Journées Scientifiques et Techniques du Réseau francophone de l'Ingénierie de la Langue de l'AUFELF-UREF*. pp. 137–146.
- PASCUAL, E. AND VIRBEL, J., 1996. Semantic and Layout Properties of Text Punctuation. In *Proceedings of the Association for Computational Linguistics Workshop on Punctuation*. Santa Cruz, USA, pp. 41–48.
- PAYNE, S.J. AND READER, W.R., 2006. Constructing structure maps of multiple on-line texts. *International Journal of Human-Computer Studies*, 64(5), pp.461–474.
- PERAYA, D., 1995. Vers une théorie des paratextes : images mentales et images matérielles. *Recherches en communication*, 4, pp.1–38.
- PERAYA, D. AND NYSSÉN, C., 1995. Les paratextes dans les manuels scolaires de biologie et d'économie. *Pratiques et Théorie*, 78.
- PERSSON, H., ÅHMAN, H., YNGLING, A.A. AND GULLIKSEN, J., 2014. Universal design, inclusive design, accessible design, design for all: different concepts—one goal? *Universal Access in the Information Society*.
- PETRIE, H., KING, N., HAMILTON, F. AND WEISEN, M., 2005. The accessibility of museum websites: results from an English investigation and international comparison. In *HCI International 2005*. pp. 22–27.
- PHIPPS, L. AND KELLY, B., 2006. Holistic approaches to e-learning accessibility. *Alt-J : Association for Learning Technology Journal*, 14(1), pp.69–78.
- PIEPER, M., MORASCH, H. AND PIÉLA, G., 2003. Bridging the educational divide. *Universal Access in the Information Society*, 2(3), pp.243–254.
- PIETRZAK, T., 2008. *Contributions à la dissémination d'informations haptiques dans un environnement multimodal*.
- PONTELLI, E., GILLAN, D., GUPTA, G. AND KARSHMER, A., 2002. Intelligent non-visual navigation of complex HTML structures. *Universal Access in the Information Society*, 2(1), pp.56–69.
- POWER, C., FREIRE, A., PETRIE, H. AND SWALLOW, D., 2012. Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In *SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 433–442.

- POWER, R., SCOTT, D. AND BOUAYAD-AGHA, N., 2003. Document Structure. *Computational Linguistics*, 29(2), pp.211–260.
- PRESCHER, D., WEBER, G. AND SPINDLER, M., 2010. A tactile windowing system for blind users. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '10*. New York, New York, USA: ACM Press, p. 91.
- PRING, L., 2008. Psychological characteristics of children with visual impairments: Learning, memory and imagery. *British Journal of Visual Impairment*, 26(2), pp.159–169.
- PRING, L. AND RUSTED, J., 1985. Pictures for the blind: An investigation of the influence of pictures on recall of text by blind children. *British Journal of Developmental Psychology*, 3(1), pp.41–45.
- PUZIS, Y., BORODIN, Y., SOVIK, A., MELNYK, V. AND RAMAKRISHNAN, I. V., 2015. Affordable Web Accessibility : A Case for Cheaper ARIA. In *W4A'15*. ACM.
- QUEK, F. AND OLIVEIRA, F., 2013. Enabling the blind to see gestures. *ACM Transactions on Computer-Human Interaction*, 20(1), p.32.
- RAMAKRISHNAN, C., PATNIA, A., HOVY, E. AND BURNS, G.A., 2012. Layout-aware text extraction from full-text PDF of scientific articles. *Source code for biology and medicine*, 7(1), p.7.
- RAO, K., 2006. *Usability and Impact of Braille and Audio Output Technologies for Students with Visual Impairments or Blindness*. State University of New York at Buffalo.
- RASMUSSEN, J., 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3), pp.257–266.
- RAYNAL, M. AND SERRURIER, M., 2002. CYNTHIA: An HTML Browser for Visually Handicapped People. In *ICCHP 2002 - LNCS 2398*. pp. 353–359.
- RICHARD, J., 1999. Lecture et compréhension. In R. Ghiglione & J. Richard, eds. *Cours de psychologie - Champs et théories*. Paris: Dunod, pp. 10–34.
- ROHANI GHAHARI, R., FERATI, M., YANG, T. AND BOLCHINI, D., 2012. Back navigation shortcuts for screen reader users. *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*, p.1.
- ROTHKOPF, E.Z., 1971. Incidental memory for location of information in text. *Journal of Verbal Learning and Verbal Behavior*, 10(6), pp.608–613.
- RUMSEY, F., 2001. *Spatial audio*, Taylor & Francis.
- RYLES, R., 1996. The impact of braille reading skills on employment, income, education, and reading habits. *Journal of Visual Impairment and Blindness*, 90, pp.219–226.
- SADATO, N., PASCUAL-LEONE, A. AND GRAFMAN, J., 1996. Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, 380, pp.526–528.
- SANDER, M.-S., BOURNOT, M.-C., LELIÈVRE, F. AND TALLEC, A., 2005. Les personnes ayant un handicap visuel. Les apports de l'enquête Handicaps - Incapacités - Dépendance. *DREES, Etudes et Résultats*, 416, pp.1–12.
- SCHMID, S., 2001. *Représentation organisationnelle et compréhension (rôle de la mise en forme matérielle dans la lecture)*. Université de Nice.
- SCHMID, S. AND BACCINO, T., 2001. Stratégies de lecture pour les textes à consigne. *Langages*, 35(141), pp.105–124.

- SCHRÖDER, M., 2009. Expressive Speech Synthesis: Past, Present, and Possible Futures. In J. Tao & T. Tan, eds. *Affective Information Processing SE - 7*. Springer London, pp. 111–126.
- SCOTT, W. A., 1955. Reliability of content analysis: The case of nominal scaling. *Public Opinion Quarterly*, 19, pp.321–325.
- SEARLE, C.L., BRAIDA, L.D., DAVIS, M.F. AND COLBURN, H.S., 1976. Model for auditory localization. *The Journal of the Acoustical Society of America*, 60(5), pp.1164–1175.
- SHINOHARA, M., SHIMIZU, Y. AND MOCHIZUKI, A., 1998. Three-dimensional tactile display for the blind. *Rehabilitation Engineering, IEEE Transactions on*, 6(3), pp.249–256.
- SHNEIDERMAN, B., 2000. Universal Usability. *Communications of the ACM*, 43(5).
- SMITH, A.C., COOK, J.S., FRANCONI, J.M., HOSSAIN, A., ANWAR, M. AND RAHMAN, M.F., 2004. Nonvisual tool for navigating hierarchical structures. In *Proceedings of the ACM SIGACCESS conference on Computers and accessibility - ASSETS '04*. New York, New York, USA: ACM, pp. 133–139.
- SORIN, L., LEMARIÉ, J., AUSSENAC-GILLES, N., MOJAHID, M. AND ORIOLA, B., 2014. Communicating text structure to blind people with Text-To-Speech. In K. Miesenberger, D. Fels, D. Archambault, P. Peñáz, & W. Zagler, eds. *ICCHP 2014 - Computers Helping People with Special Needs (LNCS vol. 8547)*. Springer Berlin Heidelberg, pp. 61–67.
- SORIN, L., LEMARIE, J., DUPUIS, A., AUSSENAC, N. AND MOJAHID, M., 2015. How educational information access can impact learning in class for blind students – A first step in the process of user-centered design of assistive technologies. *[SUBMITTED] Universal Access in the Information Society*.
- SORIN, L., LEMARIÉ, J. AND MOJAHID, M., 2015. READ : a (research) platform for evaluating non- visual access methods to digital documents. In *ASSETS'15*. pp. 3–4.
- SORIN, L., MOJAHID, M., AUSSENAC-GILLES, N. AND LEMARIÉ, J., 2013. Improving the accessibility of digital documents for blind users : contributions of the Textual Architecture Model. In C. Stephanidis & M. Antona, eds. *HCI International 2013*. Las Vegas, Nevada: Springer-Verlag Berlin Heidelberg, pp. 423–431.
- SPERANDIO, J.-C., 2007. Concevoir des objets techniques pour une population normale, c'est-à-dire comprenant aussi des personnes handicapées ou très âgées. *Perspectives interdisciplinaires sur le travail et la santé*, 9(2).
- SPERANDIO, J.-C. AND UZAN, G., 2002. Ergonomie des aides techniques informatiques pour personnes handicapées. *Handicap*, 93, pp.57–83.
- SPERANDIO, J.C., UZAN, G. AND JOBARD, N., 2004. Difficult{é}s rencontr{é}es par les aveugles et d{é}ficients visuels pour la consultation des sites Web sur les transports et le tourisme.
- SPOONER, S., 2014. " What page, Miss?" Enhancing Text Accessibility with DAISY (Digital Accessible Information SYstem). *Journal of Visual Impairment & Blindness (Online)*, 108(3), p.201.
- SPYRIDAKIS, J., 1989. Signaling Effects: A review of the research. *Journal of technical writing and communication*, 19(3), pp.227–240.
- STEINBERG, M.-A. AND MURPHY, K.M., 2012. An Annotated Review of Current Research Supporting the Use of Kurzweil 3000 in Higher Education.
- STONE, R.J., 2001. Haptic feedback: A brief history from telepresence to virtual reality. In *Haptic Human-Computer Interaction*. Springer, pp. 1–16.

- SULLIVAN, H., SAHASRABUDHE, S., LIIMATAINEN, J. AND HAKKINEN, M., 2014. Opportunities and Limitations of Haptic Technologies for Non-visual Access to 2D and 3D Graphics. In *ICCHP 2014 - Computers Helping People with Special Needs (LNCS vol. 8547)*. pp. 8–11.
- SWELLER, J., 1994. Cognitive Load Theory , Learning Difficulty, and Instructional Design. *Learning and Instruction*, 4, pp.295–312.
- SWELLER, J., 2010. Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), pp.123–138.
- SWELLER, J., KALYUGA, S. AND AYRES, P., 2011. *Cognitive load theory*, Springer.
- TAN, H.Z., GRAY, R., YOUNG, J.J. AND TRAYLOR, R., 2003. A Haptic Back Display for Attentional and Directional Cueing. *Haptics-e*, 3(1), pp.1–20.
- TEIGEN, K.H., 1994. Yerkes-Dodson: A law for all seasons. *Theory & Psychology*, 4(4), pp.525–547.
- THEOFANOS, M.F. AND REDISH, J., 2003. *Bridging the gap - Between accessibility and usability*,
- THIESSEN, P., 2011. WAI-ARIA live regions and HTML5. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility - W4A '11*. New York, New York, USA: ACM Press.
- TROUVAIN, J., 2005. On the comprehension of extremely fast synthetic speech. *Saarland Working Papers in Linguistics*, 1, pp.5–13.
- TSIAKALI, T.K. AND ARGYROPOULOS, V.S., 2015. An Investigation of the Perspectives of Teachers on the Literacy Media Selections for Students with Visual Disability : The Role of Assistive Technology. In *ICEAPVI*. pp. 12–14.
- UNESCO, 2008. International Conference on Education. In *48th session - Inclusive Education: The Way of the Future*. Geneva.
- UNESCO, 1994. The Salamanca Statement and Framework for action on special needs education. In *Conference on Special Needs Education; Access and Quality*. Salamanca, Spain: Unesco.
- UZAN, G., 2005. *Ergonomie cognitive du handicap visuel : une contribution à la conception d'aides informatiques*.
- VANLIERDE, A. AND WANET-DEFALQUE, M.-C., 2005. The role of visual experience in mental imagery. *Journal of Visual Impairment & Blindness (JVIB)*, 99(03).
- VICENTE, K.J., 1999. *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*, CRC Press.
- VIGO, M. AND BRAJNIK, G., 2011. Automatic web accessibility metrics: Where we are and where we can go. *Interacting with Computers*, 23(2), pp.137–155.
- VIGO, M. AND HARPER, S., 2013. Coping tactics employed by visually disabled users on the web. *International Journal of Human-Computer Studies*, 71(11), pp.1013–1025.
- VIGO, M., LEPORINI, B. AND PATERNO, F., 2009. Enriching web information scent for blind users. In *Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility - ASSETS '09*. New York, New York, USA: ACM Press, p. 123.
- VIRBEL, J. ET AL., 2005. Approches cognitives de la spatialisation du langage. De la modélisation de structures spatiolinguistiques des textes à l'expérimentation psycholinguistique : le cas d'un objet textuel, l'énumération. In C. Thinus-Blanc & J. Bullier, eds. *Agir dans l'espace*. Paris, pp. 233–254.

- VIRBEL, J., 2002. Eléments d'analyse du titre. In *Inscription Spatiale du Langage: structures et processus*. pp. 123–132.
- VIRBEL, J., 1985. Langage et métalangage dans le texte du point de vue de l'édition en informatique textuelle. *Cahiers de grammaire*, 10, pp.5–72.
- VIRBEL, J., 1989. The Contribution of Linguistic Knowledge to the Interpretation of Text Structures. In J. André, V. Quint, & R. Furuta, eds. *Structured documents*. New York, NY, USA: Cambridge University Press, pp. 161–180.
- VITU, F., 2005. Visual extraction processes and regressive saccades in reading. *Cognitive processes in eye guidance*, pp.1–32.
- W3C - WORLD WIDE WEB CONSORTIUM, 2010. Accessible Rich Internet Applications (WAI- ARIA) 1.0, Working Draft. , pp.1–100.
- WALKER, B.N., NANCE, A. AND LINDSAY, J., 2006. Spearcons : Speech-based earcons improve navigation performance in auditory menus. In *Proceedings of the 12th International Conference on Auditory Display*. pp. 63–68.
- WALLER, R., 1987. *The typographic contribution to language: Towards a model of typographic genres and their underlying structures*.
- WAN, C.Y., WOOD, A.G., REUTENS, D.C. AND WILSON, S.J., 2010. Early but not late-blindness leads to enhanced auditory perception. *Neuropsychologia*, 48(1), pp.344–8.
- WERNER, S., KRIEG-BRÜCKNER, B., MALLOT, H.A., SCHWEIZER, K. AND FREKSA, C., 1997. Spatial Cognition : The Role of Landmark , Route , and Survey Knowledge in Human and Robot Navigation 1 The German Priority Program on Spatial Cognition. *Informatik*, pp.41–50.
- WERTHEIMER, M. AND RIEZLER, K., 1944. Gestalt theory. *Social Research*, pp.78–99.
- WHO - WORLD HEALTH ORGANIZATION, 2015. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10).
- WHO - WORLD HEALTH ORGANIZATION, 2012. *Rapport Mondial sur le Handicap*,
- WIAZOWSKI, J., 2014. Can Braille Be Revived ? A Possible Impact of High-End Braille and Mainstream Technology on the Revival of Tactile Literacy Medium. *Assistive Technologies*, 26, pp.227–230.
- WILEY, J., GOLDMAN, S.R., GRAESSER, A.C., SANCHEZ, C.A., ASH, I.K. AND HEMMERICH, J.A., 2009. Source Evaluation, Comprehension, and Learning in Internet Science Inquiry Tasks. *American Educational Research Journal*, 46(4), pp.1060–1106.
- WILHELM, E., SCHWARZ, T., JAWOREK, G., VOIGT, A. AND RAPP, B.E., 2014. Towards Displaying Graphics on a Cheap, Large-Scale Braille Display. In *ICCHP 2014 - Computers Helping People with Special Needs (LNCS vol. 8547)*. pp. 662–669.
- WILHITE, S.C., 1989. Headings as memory facilitators: The importance of prior knowledge. *Journal of Educational Psychology*, 81(1), p.115.
- WILLIAMS, J.R., 1998. Guidelines for the Use of Multimedia in Instruction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(20), pp.1447–1451.
- WITTENSTEIN, S.H. AND PARDEE, M.L., 1996. Teachers' voices: Comments on braille and literacy from the field. *Journal of Visual Impairment and Blindness*, 90, pp.201–209.
- WRIGHT, P., 2000. *The Psychology of Layout : Consequences of the Visual Structure of Documents*,

- YALLA, P. AND WALKER, B.N., 2008. Advanced Auditory Menus: Design And Evaluation of Auditory Scroll Bars. In *ASSETS'08*. pp. 105–112.
- YANG, T., GADDE, P., MORSE, R. AND BOLCHINI, D., 2013. Bypassing Lists : Accelerating Screen-Reader Fact- Finding with Guided Tours. In *ASSETS'13*. pp. 1–8.
- YESILADA, Y., STEVENS, R., HARPER, S. AND GOBLE, C., 2007. Evaluating DANTE: Semantic Transcoding for Visually Disabled Users. *ACM Transactions on Computer-Human Interaction*, 14(3), pp.13–30.
- YU, W., RAMLOLL, R. AND BREWSTER, S. A., 2001. Haptic graphs for blind computer users. In *Haptic human-computer interaction*. pp. 41–51.
- ZECHMEISTER, E.B., MCKILLIP, J., PASKO, S. AND BESPALC, D., 1975. Visual memory for place on the page. *The Journal of general psychology*, 92(1), pp.43–52.
- ZHAO, H., PLAISANT, C., SHNEIDERMAN, B. AND LAZAR, J., 2008. Data Sonification for Users with Visual Impairment. *ACM Transactions on Computer-Human Interaction*, 15(1), pp.1–28.

## Annexes

### 1. Exemple de document dont la MFM a été formalisée

**Je fais le point sur mes connaissances**

**1. Utiliser le vocabulaire**

**a)** Dans les calculs (1) et (2) suivants, citer : une somme, un produit, un terme, un facteur, un nombre positif, un nombre négatif.  
 (1)  $(-3) + (+8) = (+5)$       (2)  $6 \times 7 = 42$

**b)** Quelle est la distance à zéro de  $-4$  ? de  $+9$  ?

**Zéro**

Pour réactiver mes connaissances

Exercice 13 p. 17

La formalisation se traduit par un ensemble de métaphrases, que l'on présentera ici sous forme discursive mais aussi traduites dans un vocabulaire XML de balises et de propriétés ; on remarque que cette version XML bien que proche en apparence d'un fichier HTML présente de nombreuses différences dans le découpage du contenu et des attributs des balises, sans compter les relations entre segments également décrites dans le fichier.

Afin de faciliter la lecture du métatexte et des fichiers, on retrouve ci-dessous les identifiants de chacun des blocs.



titre1

**Je fais le point sur mes connaissances**

1. Utiliser le vocabulaire titre2

a) Dans les calculs (1) et (2) suivants, citer : une somme, un produit, un terme, un facteur, un nombre positif, un nombre négatif.

(1)  $(-3) + (+8) = (+5)$  (2)  $6 \times 7 = 42$

b) Quelle est la distance à zéro de  $-4$  ? de  $+9$  ?

Zéro

im0

regroup1

Pour réactiver mes connaissances

titre3

Exercice 13 p. 17

bdt7

bdt1, bdt2, bdt3

bdt4, bdt5 et bdt6

mq1, mq2 et mq3

Ci-dessous le métatexte formalisant la MFM (mise en forme matérielle) de cet extrait.

==== Unités textuelles (ne font pas partie du métatexte) =====

UT1 = "Je fais le point sur mes connaissances"

UT2 = "1. Utiliser le vocabulaire"

UT3 = "1."

UT4 = "a) Dans les calculs (1) et (2) suivants, citer : une somme, un produit, un terme, un facteur, un nombre positif, un nombre négatif."

UT5 = "a)"

UT6 = "(1)"

UT7 = "(2)"

UT8 = "(1)(-3)+(8)=(+5)"

UT9 = "(2) 6x7=42"

UT10 = "b) Quelle est la distance à zéro de -4 ? de +9 ?"

UT11 = "b)"

UT12 = "Pour réactiver mes connaissances"

UT13 = "Exercice 13 p.17"

==== Métatexte correspondant à l'extrait =====

L'auteur attache UT1 à titre1

L'auteur attache UT2 à titre2

L'auteur attache UT3 à mq1

L'auteur attache UT4 à bdt1

L'auteur attache UT5 à mq2

L'auteur attache UT6 à bdt2

L'auteur attache UT7 à bdt3

L'auteur attache UT8 à bdt4

L'auteur attache UT9 à bdt5

L'auteur attache UT10 à bdt6

L'auteur attache UT11 à mq5

L'auteur attache UT12 à titre3

L'auteur attache UT13 à bdt7

L'auteur crée un texte doc0

L'auteur identifie un titre titre1, une section sec1, un titre titre2, un titre titre3, un bloc de texte bdt7, et une énumération enum0

L'auteur compose sec1 de titre1, titre2, titre3, bdt7 et enum0  
 L'auteur intitule sec1 par titre1  
 L'auteur labellise sec1 par titre1  
 L'auteur donne le niveau 1 à titre1  
 L'auteur identifie une section sec2  
 L'auteur compose sec2 de titre2, titre3, bdt7 et enum0  
 L'auteur numérote titre2 à l'aide de mq1  
 L'auteur intitule sec2 par titre2  
 L'auteur labellise sec2 par titre2  
 L'auteur donne le niveau 2 à titre2  
 L'auteur organise enum0 en item1 et item2  
 L'auteur donne le statut d'exercice à enum0  
 L'auteur numérote item1 à l'aide mq2  
 L'auteur labellise item1 par mq2  
 L'auteur numérote item2 à l'aide mq3  
 L'auteur labellise item1 par mq2  
 L'auteur donne le statut de question à item1 et item2  
 L'auteur distingue des blocs de texte bdt1, bdt2, bdt3, bdt4, bdt5 et bdt6  
 L'auteur compose item1 de bdt1, bdt4 et bdt5  
 L'auteur donne le statut de label à bdt2 et bdt3  
 L'auteur donne le statut de calcul à bdt4 et bdt5  
 L'auteur labellise bdt4 par bdt2  
 L'auteur labellise bdt5 par bdt3  
 L'auteur compose item2 de bdt6  
 L'auteur distingue un regroupement regroup1  
 L'auteur compose regroup1 de titre3 et bdt7  
 L'auteur intitule regroup1 par titre3  
 L'auteur donne le statut de renvoi à bdt7  
 L'auteur associe bdt7 à item1  
 L'auteur distingue une image im0  
 L'auteur illustre item2 par im0

Ci-dessous la version XML de la formalisation.

```

<document id="doc0">
<titre id="titre1" niveau="1">Je fais le point sur mes connaissances</titre>
<section id="sec1">
<titre id="titre2" niveau="2"> <marqueur id="mq1">1.</marqueur> Utiliser le
vocabulaire</titre>
<section id="sec2">
<enum statut="exercice" id="enum1">
<item statut="question" id="item1">
<bdt id="bdt1">
<marqueur id="mq2"> a) </marqueur>
Dans les calculs <bdt statut="label" id="bdt2">(1)</bdt>
et <bdt statut="label" id="bdt3">(2)</bdt> suivants, citer : une somme, un produit,
un terme, un facteur, un nombre positif, un nombre négatif.
</bdt>
<bdt statut="calcul" id="bdt4"> (1) (- 3) + (+ 8) = (+ 5) </bdt>
<bdt statut="calcul" id="bdt5"> (2) 6 x 7 = 42 </bdt>
</item>
<item statut="question" id="item2">
<bdt id="bdt6">
<marqueur id="mq5"> b)</marqueur> Quelle est la distance à zéro de - 4
? de + 9 ?</bdt>
</item>
</enum>
<image id="img1"/>
</section>

```

```

<regroup id="rgrp1">
  <titre id="titre3">Pour réactiver mes connaissances</titre>
  <bdt statut="renvoi" id="bdt7"> Exercice 13 p.17</bdt>
</regroup>

</section>
</document>

<relations>
  <!-- RELATIONS dérivées des statuts -->
  <rel origine="titre1" cible="sec1" verbe="intituler"/>
  <rel origine="titre1" cible="sec1" verbe="labelliser"/>

  <rel origine="titre2" cible="sec2" verbe="intituler"/>
  <rel origine="titre2" cible="sec2" verbe="labelliser"/>

  <rel origine="titre3" cible="regroup1" verbe="intituler" />

  <rel origine="bdt2" cible="bdt4" verbe="labelliser/> <!-- ce sont des
références aux calculs-->
  <rel origine="bdt3" cible="bdt5" verbe="labelliser"/>

  <rel origine="enum1" cible="item1;item2" verbe="énumérer"/>

  <rel origine="mq1" cible="titre2" verbe="numéroter"/>

  <rel origine="mq2" cible="item1" verbe="numéroter"/>
  <rel origine="mq2" cible="item1" verbe="labelliser"/>

  <rel origine="mq3" cible="item2" verbe="numéroter"/>
  <rel origine="mq3" cible="item2" verbe="labelliser"/>

  <rel origine="bdt2" cible="bdt4" verbe="labelliser"/>
  <rel origine="bdt3" cible="bdt5" verbe="labelliser"/>

  <!-- RELATIONS DE COMPOSITION -->
  <rel origine="titre2;section2" cible="sec1" verbe="composer"/>
  <rel origine="enum1; image1" cible="sec2" verbe="composer"/>
  <rel origine="item1;item2" cible="enum1" verbe="composer"/>
  <rel origine="mq2;bdt1;bdt2;bdt3;bdt4;bdt5" cible="item1" verbe="composer"/>
  <rel origine="mq3; bdt6" cible="item2" verbe="composer"/>
  <rel origine="titre3;bdt7" cible="regroup1" verbe="composer"/>
  <rel origine="bdt" cible="sec2" verbe="composer"/>

  <!-- AUTRES RELATIONS -->
  <rel origine="img1" cible="item2" verbe="illustrer"/>
  <rel origine="bdt7" cible="item2" verbe="associer"/>
</relations>

```

## 2. Étude en situation

### 2.1. Questionnaires ayant guidé les entretiens de l'étude en situation

Le questionnaire ci-dessous a été utilisé pour conduire les entretiens auprès d'éducateurs spécialisés de l'Institut des Jeunes Aveugles (IJA) de Toulouse.

### 2.1.a. Questionnaire destiné aux éducateurs spécialisés

Introduction : « Bonjour à vous, merci d'avoir accepté de bien vouloir participer à cet entretien. Je m'appelle *Laurent Sorin*, je suis un jeune chercheur et je travaille à l'université Paul Sabatier, à Toulouse. Je suis accompagné d'Antoine Dupuis, étudiant en master d'ergonomie à l'UTM, qui va conduire avec moi cet entretien. Dans notre travail, nous nous intéressons au handicap visuel et à l'apprentissage, et c'est pour cela que nous avons souhaité avoir cet entretien avec vous. Je précise de nouveau, comme inscrit dans le formulaire que vous avez signé, que votre nom n'apparaîtra nulle part et que l'entretien a pour but le recueil d'information à des fins de recherche scientifique.

Quel est votre rôle en tant qu'enseignant spécialisé à l'IJA ? Quelles sont vos activités ? Comment s'articulent-elles avec les professeurs de vos élèves en établissement classique ? »

[...]

« Dans le cadre de nos recherches, nous nous intéressons en particulier au cas des non-voyants ; ce sont donc sur vos élèves non-voyants que nos questions vont porter. »

Questions / Points à aborder
En fonction du sous-domaine des mathématiques abordé, avez-vous l'impression que les élèves NV aient des facilités ou au contraire des difficultés sur le plan de l'apprentissage ?
Diriez-vous que les élèves NV ont plus de difficultés que les autres ? Comment ces difficultés se manifestent-elles ? A quoi sont-elles dues selon vous ? Est-ce que les mêmes difficultés sont aussi présentes chez les élèves malvoyants que vous suivez ?
Concernant leur niveau en mathématiques, comment se situent les élèves NV que vous avez dans vos classes par rapport à la moyenne de la classe, en fonction du cours de maths ?
Au niveau de la <b>sélection des contenus à travailler en cours de soutien</b> : sur quels critères cette sélection est-elle faite ? Quel pourcentage du cours est travaillé ? Vous arrive-t-il de demander au service compétent des retranscriptions en braille supplémentaires par rapport à celles demandées par leur professeur ? De réaliser des adaptations du contenu vous-même ?
Comment utilisez-vous le manuel de l'élève pour préparer vos cours ? Vous arrive-t-il de reformuler son contenu dans des documents
Matériel : quel type de matériel est utilisé en classe ? (bloc-note braille type IRIS, Perkins...) Est-ce que les bloc-notes sont généralement bien acceptés par les élèves ? est-ce qu'il ya des problèmes récurrents avec ?

Comment procédez-vous pour préparer un cours ? Pour préparer un examen ?

Est-ce qu'il y a des « façons de faire » systématiques en amont du cours ? (donner une vue d'ensemble ?)

En quoi selon-vous la préparation d'un cours avec un DV peut être différente de celle qui pourrait être faite dans un cours de soutien avec un voyant ?

Qu'est-ce que l'enseignement des Mathématiques aux NV a de spécifique par rapport aux autres matières ?

Quels outils / méthodes sont utilisés spécifiquement pour l'enseignement des Maths ?

Nombre d'années d'enseignement : .....

### **2.1.b. Questionnaire destiné aux élèves déficients visuels**

Introduction : Bonjour, merci d'avoir accepté de bien vouloir participer à cet entretien. Je m'appelle Laurent, je suis un jeune chercheur et je travaille à l'université Paul Sabatier à Toulouse. Dans mon travail je m'intéresse aux personnes en situation de handicap visuel, et c'est pour cela que j'ai souhaité avoir cet entretien avec toi. Tu suis donc tes cours au sein d'une classe ordinaire avec d'autres élèves voyants, j'imagine que tu dois parfois t'adapter pour suivre. Comment t'y prends-tu pour suivre les cours en classe ? Et pour travailler à la maison ?

#### Questions / Points à aborder

Quelles sont les **difficultés**, les problèmes que tu rencontres de **manière générale** ? par exemple, des problèmes de **temps**, de **matériel**, des **parties manquantes dans le cours** ?

**A quoi sont dues** ces difficultés selon toi ?

As-tu des façons de faire ou des astuces qui te permettent de **compenser** ou de limiter les difficultés que tu rencontres ?

On a discuté des difficultés que tu rencontrais parfois mais y a-t-il des éléments pour lesquels au contraire tu as l'impression d'avoir des **facilités** par rapport à tes camarades de classes (en cours de maths ou dans d'autres cours en général) ?

Le rôle des **enseignants spécialisés IJA** est de t'aider dans tes apprentissages. Peux-tu m'expliquer à quels niveaux ils t'aident ?

Pour prendre des notes ou lire un cours, quels types d'outils spécifiques utilises-tu ?  
Rendre un devoir au professeur ?

Est-ce que tu te sers souvent du **contenu du manuel** (retranscriptions)? quand c'est le cas, dans quel but est-ce ?

En cours (outils) : pour prendre des notes ? faire un exercice ? Accéder au contenu du manuel ? rendre quelque chose au professeur ? communiquer avec tes camarades ?

Est-ce que cela arrive fréquemment / peu fréquemment que tu rencontres des erreurs / bugs avec les outils informatiques que tu utilises ?

Est-ce que tes camarades t'aident à suivre les cours ? Est-ce que tu es toujours / souvent /jamais à côté de la même personne ?
Est-ce que rester attentif en cours te demande un effort important ?
Comment fais-tu pour savoir dans quelle partie/chapitre du cours on se trouve ? Est-ce que tu te poses la question ?
De quels outils disposes-tu pour lire les équations/figures/tableaux ? Même question pour leur production ?
S'il y a des choses que tu n'as pas comprises dans le cours, comment fais-tu ? te sers-tu d'internet pour rechercher des infos complémentaires ou des explications ? Demandes au professeur ? A tes camarades ?
D'après ton expérience d'élève, vois-tu des moyens pour faciliter l'apprentissage des mathématiques en classe ordinaire pour les déficients visuels ?
Est-ce que tu connais le braille informatique à 8 points ? Est-ce que tu t'en sers ? Même question pour la notation mathématique ?
Est-ce que tu préférerais avoir des cours sur ordinateur ? est-ce que tu serais friand de plus d'exercices sur ordinateurs ?

### ***2.1.c. Questionnaire destiné aux professeurs des établissements publics***

Introduction : « Bonjour à vous, merci d'avoir accepté de bien vouloir participer à cet entretien. Je m'appelle Laurent Sorin, je suis un jeune chercheur et je travaille à l'université Paul Sabatier, à Toulouse. Dans mon travail je m'intéresse au handicap visuel et à l'apprentissage, et c'est pour cela que j'ai souhaité avoir cet entretien avec vous. Je précise de nouveau, comme inscrit dans le formulaire que vous avez signé, que votre nom n'apparaîtra nulle part et que l'entretien a pour but le recueil d'information à des fins de recherche scientifique.

Vous enseignez donc les mathématiques dans l'établissement [xxxxx] et notamment à une ou plusieurs classes accueillant un élève déficient visuel. Comment faites-vous en sorte que ces élèves puissent suivre vos cours ? «

Questions / Points à aborder
En fonction du sous-domaine des mathématiques abordé, avez-vous l'impression que les élèves DV aient des facilités ou au contraire des difficultés sur le plan de l'apprentissage ?
Diriez-vous que les élèves DV ont plus de difficulté que les autres ?
Comment ces difficultés se manifestent-elles ?
À quoi sont-elles dues selon vous ?
Les élèves DV demandent-ils de l'aide ? Posent-ils beaucoup de questions ?

Sur le plan de l'évaluation des connaissances, diriez-vous que vous avez les mêmes exigences envers les élèves DV ou est-ce que vous les adaptez en fonction de leurs possibilités ?
Concernant leur niveau en mathématiques, comment se situent les élèves DV que vous avez dans vos classes par rapport à la moyenne de la classe, en fonction du cours de maths ?
En tant qu'enseignant, avez-vous des façons de faire particulières pour enseigner à une classe comportant un élève DV ? Quelles sont, selon vous, les points problématiques dans un tel contexte ?
Sur le point particulier de l'évaluation des connaissances, comment procédez-vous ?
Le rôle des formateurs IJA est d'aider l'élève dans ses apprentissages. Pourriez-vous nous expliquer comment vous travaillez avec eux ?
Au niveau de la <b>sélection des contenus à retranscrire</b> : sur quels critères cette sélection est faite ? quel pourcentage du cours est retranscrit ?
Avez-vous suivi une ou des formations spécialisées ou des initiations/sensibilisations autour de la déficience visuelle et de la gestion d'un élève déficient visuel ?

Nombre d'années d'enseignement : .....

### ***2.1.d. Questionnaire destiné aux transpositeurs braille professionnels de l'institut***

« Bonjour, merci d'avoir accepté de participer à cet entretien. Je m'appelle Antoine Dupuis, je suis étudiant en Master d'ergonomie et actuellement en stage à l'IRIT où je participe au projet MathArchiTact aux côtés de Laurent Sorin. Dans ce travail nous nous intéressons au handicap visuel et à l'apprentissage, c'est pour cela qu'après avoir discuté avec des enseignants spécialisés nous avons souhaité nous entretenir également avec vous.

Je précise de nouveau, comme inscrit dans le formulaire que vous avez signé, que votre nom n'apparaîtra nulle part et que l'entretien a pour but le recueil d'informations à des fins de recherche scientifique. »

- § Pouvez-vous m'expliquer en quoi consiste votre travail ? (question très générale et ouverte pour d'abord laisser l'interviewé s'exprimer spontanément et librement)
- § Qui vous transmet des demandes de retranscription dans le cadre scolaire ? (enseignant, enseignant spécialisé IJA, autre?)
- § À quoi ressemble une demande ? Est-ce plutôt une consigne générale « réduisez... », ou précisant bien ce qui doit être fait ?
- § Avez-vous besoin d'échanger avec le demandeur avant de retranscrire ? Qui sollicite les éventuels échanges ?
- § Comment vous retranscrivez l'architecture des documents ? (structure, titres, mise en valeur de la structure, décalages/disposition) est-ce que c'est utile ? réellement facilitateur pour les aveugles ?
- § Selon vous, la retranscription d'un document en mathématiques est-elle différente que pour une autre discipline ? (problèmes, difficultés, facilités...) Pourquoi ?

- § Est-ce que retranscrire des formules mathématiques, des graphiques, des dessins... est plus long que le reste ?
- § Est-ce que vous retranscrivez seulement des exercices ? Ou également du cours, des exemples, des explications ?
- § Est-ce que vous réduisez le contenu que l'on transmet ? Est-ce que vous y ajoutez des éléments ? (explications par exemple) Est-ce que vous reformulez certaines choses ? (pour les maths, sinon pour le reste)

Nombre d'années de pratique : .....

## 2.2. Grille d'observation utilisée pour l'étude en situation

La grille ci-dessous a été utilisée pour prendre notes des différents évènements observés durant des cours de mathématiques, auxquels assistaient les élèves déficients visuels suivis pendant l'étude.

	Activité enseignant										Interaction Ens è DV ; DV è Ens EI è DV ; DV è EI Acc è DV ; DV è Acc	Observations éventuelles
	Dire / Lire / Ecrire / Montrer											
	On note le support en rapport avec l'activité  Tableau / Polycopié / Manuel élève / Cours enseignant / Autre <u>dessiner</u> : Courbe, Equation, Tableau, MFM ? <u>Montrer</u> : Geste / posture											
	Dire	Lire	Ecrire	Montre	Dessine	Support						
						Tabl	Poly	EI- Man	Pr - Cours	Autre		
1	...	...	...	...	...	...	...	...	...	...	...	...
2	...	...	...	...	...	...	...	...	...	...	...	...

## 2.3. Extraits des entretiens réalisés pour l'étude en situation

Abréviations: CT = citation, STD = élève (student), SPT = éducateur spécialisé (special education teacher), TE = professeur (school teacher), TR = transcripteur braille. Des notes de contexte sont parfois insérées dans les citations entre crochets []].

Chaque groupe de citations identifie un problème présenté dans les résultats.

[CT 1] *blind students never have access to their textbooks in their entirety*

-STD3: *All the exercises aren't transcribed, only what the teacher asks us. There are teachers that ask a lot and others that ask little [...] the book does not exist. Books already fully transcribed, we buy them. Here it wasn't, so as we can't transcribe everything and as she takes notes well...*

-TE3: *The textbook isn't used often in mathematics except for the exercises.*



-TR2: *if there are 140 exercises within the chapter, we ask the teacher to make a selection [...] according to his educational objectives*

[CT 2] *transcribing documents into braille is a time consuming process*

-TR2: *I take seven hours for a chapter of about 30 pages, it is not that much but we have to describe everything, rewrite everything, all of that is complex*

-TR1: *we don't have any software capable of transcribing automatically everything in Maths, so we have to prepare the documents*

[CT 3] *transcribers have a heavy workload*

-SPT3: [...] *the transcription service is overwhelmed*

-TR2: *we have less and less time, as there is more and more youngsters arriving*

-STD1: *Sometimes they don't send the braille documents on time so it's complicated...*

[CT 4] *blind students are unable to use as much content as their sighted peers*

-SPT2: *We chose quality rather than quantity with the students. Braille is slow to read, so we propose less exercises, less examples...*

-TR2: *every time they have such enormous volumes that I think it is best for them... yeah, well it is better if we can keep only the most important information to understand*

-SPT1: *It isn't always obvious to know what to keep. Generally we do simple. Even more for poor braille reader, the transcriber avoid overloading the documents to make reading easier*

[CT 5] *typography is considered an aesthetics ornament*

-TR2: *for Braille I remove all the embellishments, that is to say bold, centered, underline, colors, etc.*

-SPT1: *When we adapt a document in braille or in a magnified version: we mess up the typography, we can't keep everything. In braille there is almost no typography [...] so we lose information*

[CT 6] *braille readers don't like to have too many interruptions in their reading stream*

-SPT4:STD3 *prefers paper, as on her braille note-taker there are only 40 characters on the braille terminal so she needs to go back and forth all the time whereas with paper it is more continuous, she can read easier*

-SPT1: *Good braille readers don't like empty spaces, they always need something under their fingers.*

[CT 7] *(Tactile exploration) form a global representation*

-SPT4:ST3 *has a very good mental representation, she understood how book works. There is still a layout in braille, an indentation, which helps finding without reading everything*

-TR2: *If there is a lot of information on a map we will try splitting it up in two or three maps. So the fingers don't lose themselves and the youth builds easily a reliable mental representation*

[CT 8] *students need to carry heavy documents*

-STD1: *In braille it makes a big volume.*

-TE1:STD1 *has two schoolbags extremely heavy and loaded, and she takes way too much stuff [...] she thinks she needs everything all the time*

[CT 9] *students have to be very organized*

-ST4: *When I see my students we can work on methodology for organizing documents*

-ST3: *We teach them to organize their braille transcriptions in folders and to empty it regularly*

-STD1: *I'd like to do more stuff on the computer. I would have less paper. Because it is complicated to organize...*

[CT 10] *expertise in mathematics braille code greatly influence the information access*

- STD1: *I don't know the mathematics braille notation very well...*
- STD2: *For the braille mathematics symbol, either I learnt them and forgot or I didn't learn them. But since I don't like mathematics braille I don't bother learning it*
- ST4: *If they don't use the mathematics braille notation they forget it*

[CT 11] *Most of the students are actually poor braille readers*

- ST4: *All my students don't read braille fluently. We have youngsters with difficulties for reading braille [...] They don't read their lessons, they don't like braille*
- STD2: *I don't read braille very fast*

[CT 12] *students have trouble retrieving the global structure of the document*

- SPT4: *They haven't the exercise structure in mind if they don't read everything.*
- SPT1: *They haven't an overview of the support. For instance, an example with an exercise beside it is not possible in braille. It takes more organization*
- SPT3: *They have to understand the architecture, check the paragraphs, and read the whole document to represent the page structure*
- STD2: *with the exercises there are several spaces between lines... I go down gradually and when I find an empty space I check the number*

[CT 13] *the ambient noise of the classroom bothers the blind students*

- STD3: *Sometimes it's hard to stay focused, sometimes not, it's even harder when the others are excited and chat loudly*
- STD4: *in order to better follow the classes I would need less ambient noise*
- STD5: *My classmates don't really help. It's a class with little attention for the lessons*

[CT 14] *students need to use a special code to print mathematics content properly*

- STD2: *[in his braille note-taker] To write maths I use dot 6 and dot 3 [braille character] and it writes everything in maths if it is [the file] in the right folder, if I put it in other folders it won't work. In other folders we have to use the "dot-7-dot-8-character" to print it properly for the teacher, otherwise it prints accents...*
- STD4: *I use a special notation to print stuff for the teacher...*

[CT 15] *braille note-takers happens to regularly break down*

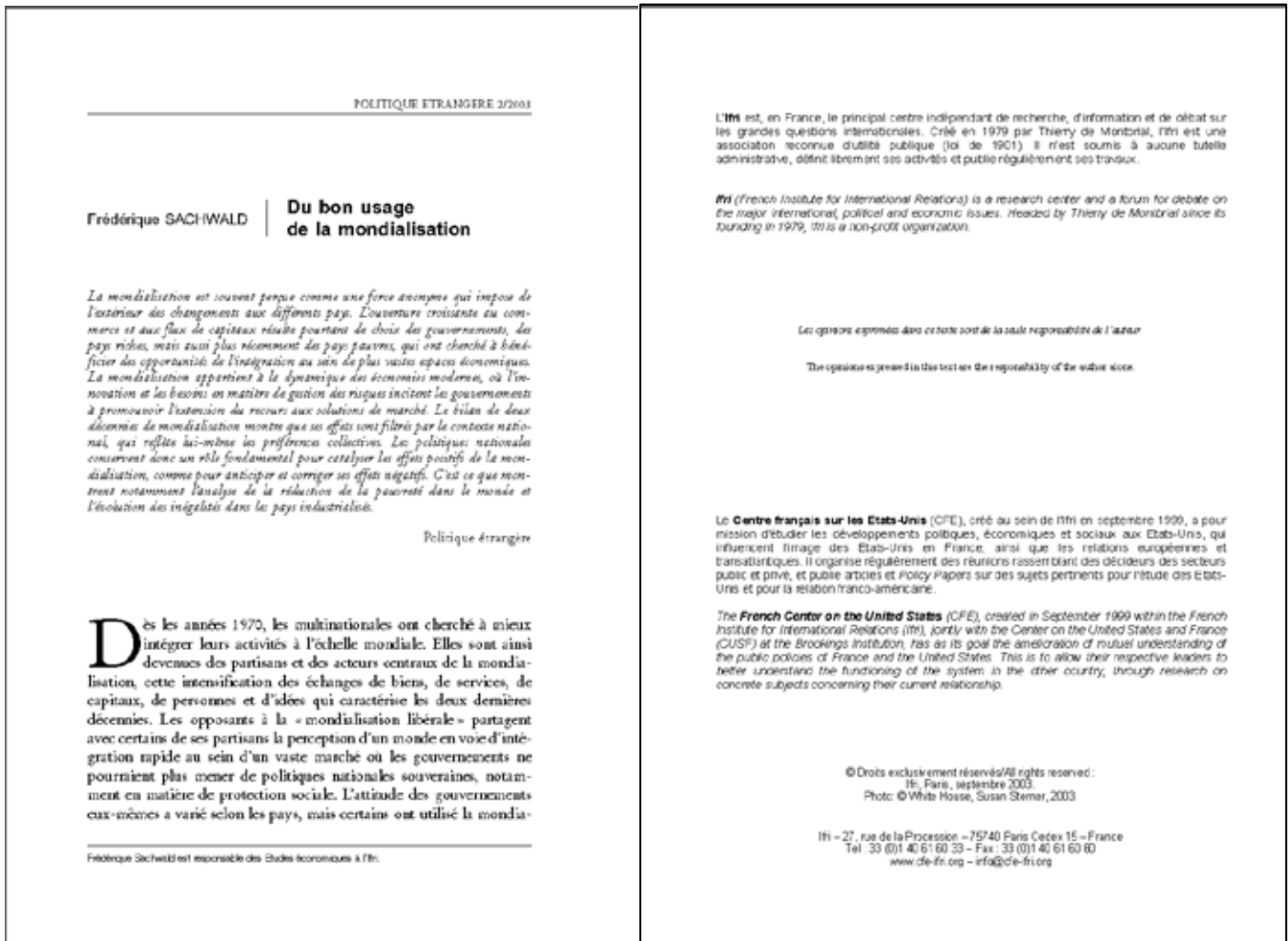
- SPT3: *After one or two years they all [every braille note-takers] happen to breakdown, systematically; 90% of them...*
- TE3: *STD2 has his braille note-taker at every class. It broke down though at the beginning of the year for a while...*
- STD2: *the machine works generally, except when it stops and I have to start all over...*

[CT 16] *good braille readers braille use white spaces on braille documents to find information*

- SPT1: *On braille documents you can find white spaces, indents, a disposition which allows reader to find information without reading everything. STD4 is capable of doing so, but I haven't many students who can do it*

### 3. Extraits de documents formalisés automatiquement par apprentissage automatique

Ci-après plusieurs images montrant des extraits de documents issus du corpus LING et d'autres du corpus GEOP, utilisés pour entraîner et évaluer notre système de formalisation automatique présenté en Chapitre III.7.



Extraits de documents du corpus GEOP (articles de géopolitique)

## G21 et « Quadrilatérale » : quelques indicateurs

G 21	1986*		2000	
	G21	Quadr	G 21	Quadr
% Population mondiale	51,1	15,1	51,7	13,5
% PIB mondial**	9,3	75,6	11,4	73,4
% Exportations mondiales de biens et services**	6,4	65,7	11,2	61,5
<b>Israël-Chine</b>				
% Population mondiale	29,3		30,9	
% PIB mondial**	8,0		8,4	
% Exportations mondiales de biens et services**	5,2		6,9	

Source : World Development Indicators / Banque Mondiale

\* Les données pour 1986 représentent des approximations à raison des données économiques non disponibles pour Cuba et l'Équateur

\*\* USD dollars constants de 1985

- *Petits pays pauvres et grandes ONG riches<sup>11</sup> : l'influence de la « diplomatie non gouvernementale ».*

Celles des ONG qui ne refusaient pas en bloc l'existence de l'OMC, particulièrement dans le monde anglo-saxon, revendiquaient depuis plusieurs années des moyens d'influence sur une institution qui ne leur reconnaissait pas de statut. Avant le cycle de Doha, cette revendication se concentrait sur le mécanisme de règlement des différends. Depuis le lancement du cycle, elle porte également sur les négociations. Cancun a manifesté l'évolution des méthodes d'intervention de certaines grandes ONG, évolution déjà très perceptible, dans la sphère de l'ONG, lors du sommet de Johannesburg : celles-ci assistent ou influencent des petits États, dépourvus de capacités suffisantes de négociation, pour orienter l'agenda. Onfam revendique ainsi d'avoir contribué à l'élaboration du dossier du coton, qui a eu une influence psychologique décisive sur le sort de la conférence de Cancun.

A cette assistance technique, les ONG de terrain ajoutent leur capacité de sensibilisation politique des populations. A Cancun, plusieurs ministres africains ont indiqué sans voile qu'ils se trouveraient politiquement le dos au mur si aucun mouvement n'était fait sur le dossier du coton, apportant avec eux des pétitions de dizaines de milliers de producteurs, ou, comme la Ministre du Sénégal, une résolution votée à l'unanimité de son parlement. Le Commissaire européen à l'agriculture, M. Fischler, ne s'est pas trompé sur le rôle de ces nouveaux acteurs : il imputait une partie de la responsabilité de l'échec de Cancun à certaines ONG, avant d'estimer que la redéfinition des relations entre membres de l'OMC et ONG devait être la première priorité de réforme de l'organisation.<sup>12</sup>

#### Le libre-échange ne suffit plus : géographie des doutes face à la globalisation

La question stratégique posée par Cancun est de savoir si les membres de l'OMC estiment encore avoir un intérêt à promouvoir une libéralisation multilatérale globale. L'échec de Cancun a réveillé les doutes fondamentaux des participants à la négociation.

<sup>11</sup> Le chiffre d'affaires du dernier exercice d'Onfam UK était en 2002 de 139,3 millions de livres (source « annual review and accounts 2001/2002 ») soit environ 316 MUSD. Le budget courant du Burkina Faso était de 342,4 MUSD, celui du Bénin de 342,4 MUSD celui du Tchad de 154,9 MUSD (source Banque Mondiale - World Development Indicators).

<sup>12</sup> Tobias Buck « No fresh WTO talks without reform » or *« Les pays riches »* - FT - 20 septembre 2000

8

## I. Une querelle Exécutif/Législatif

La question des pouvoirs de guerre n'est qu'une composante de la rivalité opposant l'Administration au Congrès, qui caractérise l'ensemble de la vie politique aux États-Unis. Dès lors que les conditions d'une intervention militaire sur un théâtre extérieur se trouvent remplies, les parlementaires mettent en avant les prérogatives qui leur sont offertes par la Constitution. Moins présents pendant la guerre froide, durant laquelle ils ont laissé au président une plus grande liberté d'action, les membres du Congrès sont revenus, depuis le début des années 1980, à une lecture plus fidèle des textes rédigés par les pères fondateurs, notamment à l'occasion de la guerre du Golfe (1990-1991). Il convient donc de s'attarder sur le principe constitutionnel de cette répartition des pouvoirs, puis sur la lecture qui en a été faite en fonction des circonstances et de la perception de la menace.

### 1. La Constitution et le War Powers Act

La Constitution des États-Unis définit les pouvoirs du Congrès en matière de relations internationales par « la défense des intérêts nationaux, la régulation du commerce, la déclaration de guerre, et le soutien aux forces armées<sup>13</sup> ». Ces prérogatives sont souvent rappelées par ceux qui défendent le rôle des parlementaires dans les affaires étrangères, en particulier depuis la fin de la guerre froide<sup>14</sup>. De son côté, le président ne peut déclencher une guerre ni mener d'opérations militaires qu'avec l'accord de deux tiers des sénateurs, tout comme il ne peut, seul et sans l'aval du Congrès, nommer des ambassadeurs<sup>15</sup>. En outre, le Congrès peut engager une procédure de destitution du

<sup>13</sup> L'article I, section 8 de la Constitution énonce les pouvoirs du Congrès comme suit : « To [...] provide for the common Defence and general Welfare of the United States; [...] to regulate Commerce with foreign Nations, and among the several States, and with the Indian Tribes; [...] to define and punish piracy and felonies committed on the high seas, and offences against the law of Nations; to declare war, grant letters of marque and reprisal, and make rules concerning captures on land and waters; to raise and support armies [...] to provide and maintain a Navy; [...] and to make all laws which shall be necessary and proper for carrying into execution the foregoing powers, and all other powers vested by this Constitution in the Government of the United States, or in any Department or Office thereof. »

<sup>14</sup> Louis Fisher, *Presidential War Powers*, Lawrence, Kansas University Press, 1995 : « The framers vested in Congress explicit control over the initiation and authorization of war, power over foreign commerce, approval of treaties, confirmation of ambassadors, power of the purse, and other authorities over external affairs. »

<sup>15</sup> Constitution des États-Unis, article II, sections 2 et 3 : « The President shall be Commander in Chief of the Army and Navy [...] He shall have power, by and with the advice and consent of the Senate, to make

9

**Principes d'une grammaire prédictive du discours**  
(français langue étrangère et maternelle)

Muriel Barbazan  
Laboratoire Jacques-Louis  
Lagarange (Toulouse-Le Mirail)  
muriel.barbazan@univ-tlse2.fr

**Abstract**

La grammaire actuelle a du mal à articuler le système traditionnel aux investigations plus récentes des « analyses de discours », qui développent la linguistique énonciative (interactionniste) et la linguistique textuelle, secondée par la psycholinguistique. Ces difficultés sont encore plus marquées en FLE que dans le domaine du FLM. Afin de comprendre les causes de ces difficultés interactives globales tant sur le plan théorique que sur le plan didactique, nous explorons les contraintes liées au statut de l'enseignant et les décalages paradigmatiques énonciatifs et textuels qui doivent inévitablement gêner ce cadre trop restrictif. Ce parcours problématique mettra en évidence l'origine épistémologique de l'hétérogénéité actuelle des domaines grammaticaux, mais aussi le paradoxe de leur complémentarité insaisissable du point de vue cognitif (développement ontologique et gestion mentale du langage). Il sera alors possible de justifier les principes d'une reformulation descriptive articulant l'héritage syntaxique aux fonctions énonciative et de structuration du texte, tant au niveau de la globalité textuelle qu'au niveau des moyens grammaticaux qui en permettent l'élaboration (discours, construction, système verbal, processus / reconstructions, système phonétique, unités de structure du texte).

**1 Introduction**

La grammaire, plus encore en français langue étrangère (FLE) qu'en français langue maternelle (FLM), a du mal à se présenter comme un domaine homogène, qu'articule de façon pertinente l'héritage traditionnal et les avancées récentes de la linguistique énonciative et des grammaires de texte. Si le FLM semble être en meilleure voie de réorganisation (Ducrot, Lieber et al. 2001), le FLE a des difficultés à intégrer en un système cohérent des notions énonciatives ou textuelles encore trop pérorantes, et qui ne sont pas reliées aux connaissances grammaticales traditionnelles : on est face à une juxtaposition de plusieurs systèmes, sans fil conducteur permettant de repérer dans cet ensemble un cheminement cohérent. Il faudrait pouvoir sortir de cette impression « d'embouteillage de différents dispositifs d'organisation de la langue, du discours à l'énoncé, du texte à la phrase, même si l'on ne sait pas encore très bien comment s'articulent dimension discursive ou énonciative et dimension formelle de l'organisation de la langue » (Viguer 2004, 36). Un des défis actuels posés au grammairien est donc de tenter d'homogénéiser ces divers concepts en une discipline plus cohérente. Tant que ce n'est pas fait, l'apprenant n'a pas d'autre choix que d'essayer d'acquiescer comme il peut ses connaissances entre elles :

« Il appartient toujours aux élèves d'établir les liaisons nécessaires entre ces différents niveaux d'organisation de la langue (du discours aux éléments de la phrase ou du mot), l'essentiel étant de proposer des activités qui facilitent ce lien en relation. » (Viguer 2004, 36)

Ce point de vue sous-entend tout de même un peu de fatalisme, ce il revient à situer la didactique en aval de modifications qui ne concernent pas la seule linguistique. Deux raisons semblent justifier un point de vue

1990, Encyclopédie philosophique universelle, II. Sur les notions philosophiques, ditomente, vol 1 : 1313)

Notre objectif est de rechercher en quoi le « complément » apparaît comme une connaissance nouvelle qui va se stabiliser dans le discours gramm. En quoi est-il un savoir nouveau sur la langue ? Incidemment, cette question en soulève d'autres, dont celle du statut particulier de l'invention en sciences du langage, et de la relativisation de la notion de « progrès » dans la science grammaticale : peut-on dire que le « complément » est mieux que le « régime » en tant qu'outil de l'analyse syntaxique ?

Nous proposons d'examiner ici un moment charnière dans l'histoire de la syntaxe : le passage du « régime » au « complément », en nous demandant dans quelle mesure l'émergence du « complément » apparaît comme une « invention ». Dans un premier temps, nous tentons d'identifier les traits de « nouveauté » du complément en comparant les deux notions. Dans un second temps, nous avançons comme élément de réponse principal que la construction du concept de complément ressortit en fait à l'évolution globale d'un faisceau de notions et de représentations grammaticales tenant à la structure de la proposition (la transitivité verbale, la notion de « genre verbal », le nombre et la forme des constituants dans le modèle propositionnel adopté).

**1 Les concepts de régime et de complément : quelles similitudes ? quelles différences ? quelle nouveauté ?**

On peut définir un concept comme un complexe constitué d'une définition, manipulant un contenu idéel, impliquant parfois d'autres concepts, relié à une terminologie. On peut ajouter une série de faits auxquels la définition s'applique, soit-ils représentés par des exemples. Chacun de ces éléments se trouve lié au concept à l'histoire, et le concept ne forme pas une entité de signification stable ou fixe. L'histoire du complément n'est pas l'histoire d'une invention ex nihilo mais c'est celle d'une construction longue et complexe, à partir de divers termes hérités ou forgés, et à partir d'autres notions provenant de la logique ou de la grammaire. Le « complément » se construit notamment en interférence avec le « régime ». C'est en relevant les points de convergence et les points de divergence que l'on observe le mieux la nouveauté existant entre les deux notions.

**1.1 Les points de convergence entre le régime et le complément**

Premièrement, il faut noter qu'aucun de ces deux termes ne provient du domaine épistémologique et technique de la logique (il l'inverse d'un attribut, « prédicat », ou « objet »).

La notion de régime provient du vocabulaire de Platon mais elle se développe au Moyen-Âge :

La notion médiévale de *rektion* semble avoir pour origine l'idée d'origine qui se trouve chez Platon ; certains nous, par leur nature, exigent un cas oblique, comme j'ai demandé le prêtre (Fosler, 1983 : 139).

Pierre Hélie explique ainsi que le terme de *regr* est employé pour désigner la relation du verbe au nominatif au moyen d'une métaphore militaire ; le verbe gouverne le nominatif comme le chef gouverne son armée (ibid., et Chevalier, 1968 : 55). Le régime est défini à l'aide de différentes images dont le point commun est l'idée d'un rapport de domination. Cependant, sa définition est variable et la notion n'est pas encore totalement stable au Moyen-Âge. Alexandre de Villeneuve<sup>10</sup> emploie le terme de *regimen* dans le chapitre 2 du *Doctrinale puerorum*, associé à ceux de *rektor*, *regno* et *guberna* (Colombat, 1999 : 428-429). Le terme est traduit et emprunté directement par les premiers grammairiens du français. La dépendance au verbe est pensée en termes de « régime » et les fonctions sont décrites à l'aide du système des cas. L'idée que le verbe « gouverne » le nom qui le suit n'est à dire qu'il impose un cas à sa suite, s'implante solidement dans la grammaire française. Quelques grammairiens, comme Palgrave (1530) ou Ramus (1562, 1572)<sup>11</sup>, tentent de différencier les types de régime, ce qui est souvent traité par le traitement du pronom ou de la préposition et de l'adverbe<sup>12</sup>. Le substantif « complément » n'apparaît lui qu'au milieu du 18<sup>ème</sup> siècle, de façon minoritaire au sein d'un réseau terminologique varié,

<p><b>3 Utilisation du concept dans le domaine du lexique construit</b></p> <p>Il ne s'agit pas ici de dresser un historique exhaustif de la notion telle qu'elle a été utilisée en grammaire et en linguistique de l'Antiquité à nos jours, mais seulement d'en broser les grandes lignes, en contrastant les observations sur le domaine de la formation des mots (pour un point historique plus documenté, je renvoie à Chevalier &amp; Delesalle, 1986, qui montrent la permanence du concept du XVIII<sup>e</sup> siècle à la fin du XIX<sup>e</sup> siècle, ainsi qu'à Biala, 1991 : 107-113).</p> <p><b>3.1 De l'Antiquité jusqu'au milieu du XIX<sup>e</sup> siècle</b></p> <p>Sans entrer dans le détail :</p> <ul style="list-style-type: none"> <li>– les grammairiens grecs de l'Antiquité opposaient la notion d'analogie, alors synonyme de régularité, à celle d'anomalie, synonyme d'irrégularité, de désordre, de dissemblance. Cette dichotomie a donné lieu à la célèbre controverse entre analogistes, pour qui la langue était fondamentalement régulière, et anomalistes, qui soutenaient la thèse inverse<sup>7</sup> ;</li> <li>– cette conception a perduré jusqu'à Port-Royal, où l'analogie est considérée comme une « sorte de projection des règles de la raison dans une langue particulière » (Chevalier &amp; Delesalle, 1986 : 332) : elle s'oppose alors à la notion d'usage (Arnauld &amp; Lancelot, 1660 : 83), en tant que susceptible de venir priver ce qu'on appelle alors « l'analogie de la langue » ;</li> <li>– Vaugelas, à la même époque, associe au contraire analogie et (bon) usage : pour lui, la première est le « processus d'établissement » du second (Chevalier &amp; Delesalle, 1986 : 333).</li> </ul> <p>En tout état de cause, pendant cette longue période, l'analogie n'est jamais conçue comme un principe d'explication, pas plus qu'elle n'est particulièrement évoquée à propos du lexique construit. Ce n'est qu'à partir de Desmazes qu'elle devient tout à tour :</p> <ul style="list-style-type: none"> <li>– un principe organisateur des langues, permettant de les classer en langues analogues (où analogie signifie « ce qui suit l'ordre naturel », i.e. l'ordre SV O), langues transpositives, langues mixtes,</li> <li>– un principe organisateur à l'intérieur d'une langue, permettant de séparer l'acceptable du non acceptable,</li> <li>– un principe explicatif, s'opposant chez certains grammairiens à l'arbitraire<sup>8</sup>.</li> </ul> <p>Pour la 1<sup>re</sup> édition du dictionnaire de l'Académie (1694), « [l]es mots nouveaux ne peuvent guère s'introduire qu'à l'aide de l'ANALOGIE »<sup>9</sup>.</p> <p>La première moitié du XIX<sup>e</sup> siècle s'intéresse peu à l'analogie : l'objectif de la grammaire comparée n'est pas de découvrir les principes dynamiques à l'œuvre dans les langues, mais d'en découvrir les origines. L'analogie est par conséquent très en marge de ce programme de recherche.</p> <p><b>3.2 Du milieu du XIX<sup>e</sup> siècle à Saussure</b></p> <p>On assiste en revanche à une réémergence du concept chez les philologues travaillant sur l'évolution du latin au français. Pour eux, l'analogie est surtout un facteur de trouble : ainsi l'exemple, cité par Chevalier &amp; Delesalle (1986 : 344), de l'extension du /s/ au cas sujet de tous les noms, considérée comme une régularisation abusive, arbitraire, allant à l'encontre des « lois naturelles » que sont les lois d'évolution phonétique.</p> <p>Les néo-grammairiens s'emparent à leur tour du concept et le redéfinissent. L'analogie devient alors une force dynamique, précisément parce qu'elle contrecarre les lois phonétiques : ainsi, pour Leskien (1876) cité d'après Touratier (1988 : 140-141), les lois phonétiques et l'analogie constituent les seuls facteurs susceptibles d'expliquer la forme que présente à une époque donnée la déclinaison d'une langue.</p>	<ul style="list-style-type: none"> <li>– Un assouplissement des contraintes spatio-temporelles et ainsi la possibilité de toucher de nouveau public n'ayant pas le temps de se déplacer une ou deux fois par semaine. Cet argument vaut notamment pour l'enseignement universitaire des langues aux spécialistes d'autres disciplines.</li> <li>– Une individualisation plus grande de la formation : les apprenants directs de l'étranger davantage présents sur ce simple et voir offrir des riches et ressources supplémentaires en ligne.</li> <li>– Une diversification des formes d'expression : en classe, on communique le plus souvent à l'oral, par petits groupes ou en grand groupe. À distance, il est possible d'avoir une communication écrite collective débouchant sur une socialisation des productions écrites.</li> <li>– Une place plus importante donnée à l'autoformation, à condition bien sûr que le public soit suffisamment autonome pour travailler en dehors de la présence de l'enseignant.</li> </ul> <p><b>2.2 Le lien social à distance</b></p> <p>Dans le cas de formations entièrement à distance, comme c'est le cas de plusieurs masters de FLÉ (Orléans, Rouen, Le Mans), Internet permet d'une part de créer un lien plus interactif entre les enseignants et les étudiants, d'une part et surtout d'établir un lien social entre les étudiants. On s'est ainsi également rendu compte, à la fin des années 1990, que la faible participation des étudiants aux premiers forums qui accompagnaient les formations prenait plus d'un tiers : souffrant et d'un manque d'objectifs fixés à ces forums que d'une réticence à échanger via Internet. Dès que des tâches collectives tirant parti de l'outil ont été proposées, la participation s'est fortement accrue et, à côté d'échanges de nature socio-cognitive, on a commencé à observer des phénomènes d'ordre socio-affectif et l'établissement d'un lien social (Devollet &amp; Margenet, 2004). Engagé depuis 1995 dans la migration vers Internet d'un master de FLÉ auparavant délivré par correspondance sous forme traditionnelle, Margenet (2002, 2003) a pu analyser les échanges écrits apprivoisés qui se faisaient jour dans ce cadre et notamment mettre en relation les tâches proposées avec la nature des échanges entre pairs et avec le tuteur.</p> <p><b>2.3 Communication exolingue pour l'appropriation des langues</b></p> <p>Si l'est un domaine où les échanges en ligne ont tenté à se multiplier, c'est bien celui de l'apprentissage des langues. La tradition du « pen friend » existait depuis longtemps, mais Internet a permis de servir ce genre de pratiques. Si l'est toujours possible de faire connaître des apprenants en mode un à un, avec les Tardieu par Internet, de nombreux projets associent des groupes d'échange en ligne de manière institutionnelle ou sont développés ces dernières années : créés parfois sur l'intercompréhension à l'intérieur d'une famille de langues (Dulmeijer), souvent sur la pratique à partir de deux langues, visent presque toujours un échange au plan culturel, impliquant dans certains cas de futures enseignants de langue, ces projets sont d'une grande diversité et on se contentera ici de renvoyer à deux recueils consacrés à cette question : <i>Ouvrir l'interculturalité</i> (D'Arbo, 2007) et <i>L'Écriture, Échanges et Apprentissage</i> (Margenet et al., 2007). Plusieurs thèses à ce sujet ont la qualification en sciences du langage, ont déjà été soutenues ou ce sujet.</p> <p><b>2.4 L'écrit numérique source de réflexivité</b></p> <p>Dans le prolongement de l'approche anthropologique de Goolbsy, Lévy (1990) a qualifié les TIC de « technologies intellectuelles ». Plus concrètement, des chercheurs anglais en sciences cognitives, Stephen &amp; Penberon (1990), ont relevé les divers modes d'interactions de la cognition liés aux différents artefacts, du post-it aux assistants logiciels à l'écriture. Dans cette même optique, Lamy &amp; Goodfellow (1998) parlent de « conversations réflexives [...] par conférences synchrones » et Margenet (2004) souligne deux caractéristiques particulièrement intéressantes [des forums] : la souplesse chronologique qu'autorise le temps différé et la permanence de l'écrit qui fait du forum l'équivalent d'un</p>
--	--

Extraits de documents du corpus LING (articles de linguistique)

## 4. Expérimentations sur la restitution de la MFM

### 4.1. Retours qualitatifs des sujets voyants

Retours des sujets voyants recueillis dans le cadre de l'expérimentation présentée en 4. Les retours sont classés par catégorie ; on précise pour chaque retour le code du sujet correspondant ainsi que la condition expérimentale qu'il a passée (O/X/J/N : optimale, exhaustive, lecteur d'écran et sans MFM).

#### -Apprentissage / mémorisation

20 – X, 57 – O, 56 – X, 53 – O, 45 – O, 42 – J, 38 – J, 30 – J : On a perdu cette habitude d'écouter, ça fait bizarre au début au niveau de la concentration on a vraiment l'impression de devoir faire un effort. [ou encore] je n'ai pas une mémoire auditive, je suis visuel. [ou encore] j'avais du mal à me concentrer [ou encore] Comme je ne visualise pas les mots je reste sur une phrase et puis ça part, et je décroche comme ça va assez vite. Je ne pouvais pas me poser sur des mots ou sur des phrases. [ou encore] Je trouve que

c'est difficile de retenir le texte je n'ai pas l'habitude d'apprendre comme ça. [ou encore] Je suis pas du tout habitée j'ai jamais écoutée des documents donc réfléchir à la structure c'est dur, avec l'habitude on s'améliore

45 – O : c'est bizarre car je retiens bien à l'oral d'habitude. Mais pour apprendre je me fais des fiches et j'utilise le visuel pour la structure et je me relis, donc j'ai l'oral pour apprendre. Là je n'arrive pas à structure ce que j'intégrais

57 – O : Moi pour apprendre j'apprends le plan d'abord ; en fait j'ai une mémoire schématique, donc d'abord le plan et après quelques idées que je surligne et je les raccroche au plan.

#### Compréhension du fond et de la forme

54 – J, 44 – X, 28 – X, 24 – X : Je me concentrais plus sur la structure que sur le fond... [ou encore] Ma concentration était plus portée sur la forme et pas le fond, comment se baladait le son au niveau spatial et des sons plus ou moins aigus, mais ma concentration n'était pas sur le document.

48 – X, 45 – O : Moi je le sentais c'était soit la forme soit le fond (pour la compréhension). [ou encore] j'étais trop focalisé sur la structure ; et même sur fond j'étais parasité par la structure. Je m'intéressais plus à la structure au début [...]

48 – X : si j'avais intégré, c'est un apprentissage, la manière de présenter la structure, à ce moment-là je pourrai me concentrer sur le fond.

## Autres

12 – X : Ca m'a paru court [temps vs. contenu].

24 – X : Pendant les retours je me rappelais de toutes les questions.

16 – X : Là je ne me suis pas senti guidé, j'ai eu l'impression de devoir beaucoup travailler.

## Contenu

58 – J, 2 – J : J'ai pas du tout compris que c'était des notes de bas de page, ce n'était pas les remarques alors, les remarques c'était encore autre chose ? [ou encore] Les 1/2/3 qui arrivent en bas de page je ne sais pas à quoi ça correspond.

60 – X, 57 – O, 38 – J, 34 – J : l'économie ce n'est pas mon truc. [ou encore] Sur un texte d'informatique ça aurait mieux marché ! [ou encore] Ça aurait été un autre sujet que l'économie s'aurait été mieux. [ou encore] C'est plus le contenu qui ne m'a pas passionné.

46 – J, 35 – N : les notes de bas de page au milieu [ les numéros ] c'est perturbant par ce que j'ai l'impression d'un coup qu'il y a un titre au milieu de la phrase.

57 – O : Là le plus dur c'était les rubriques parce que j'avais pas à les rattacher.

## Navigation & fonctions implémentées

27 – N : pas essayé de revenir en arrière car c'était trop rapide (maîtrisait pas la fonction).

9 – O : La touche ECHAP ne m'a servi à rien.

57 – O : Tout ce qui est navigation c'est super bien fait ; t'arrive bien à te représenter où tu en es. Y'a pas trop de touches non plus donc on s'éparpille pas trop non plus.

37 – O : Je trouve la table des matières vachement bien.

41 – O : je retrouve vraiment bien la structure c'était assez limpide à parcourir.

28 – X : facile de se représenter le document pendant.

## Restitution orale du logiciel : Métatexte

37 – O : les amorces des listes lues par une voix différente ça distrait, ça donne l'impression qu'on sort de la rubrique en cours de lecture qu'on change de paragraphe, là on pense que c'est plus important parce que ça change de voix. On n'a pas l'impression d'approfondir quelque chose mais de retourner en arrière...

9 – O, 33 – O : les amorces dans la VG perturbent quand on revient en arrière [ou encore] Quand t'es dans la table des matières il dit titre [UN META TEXTE] un avec toutes les sous-sections, puis juste après il redit les titre des sous sections alors que je viens de les entendre.

3 – N, 41 – O, 31 – J, 58 – J, 52 – X, 41 – O, 37 – O, 1-O : l'appellation titre de niveau 1 / niveau 2 suivie de la numérotation des titres est perturbante [ou encore] Lors de la lecture des titres ou des listes, même avec la voix masculine, lors du chiffre du niveau et le descriptif des éléments, y'a pas assez de temps : donc on entend le numéro et directement après la suite, donc ça peut être difficile de distinguer [ou encore] Peut être qu'il serait plus simple d'utiliser le terme « Grand 1 petit A » plutôt que « titre de niveau 1 ; 2 » [ou encore] Son interprétation des parties quand elle dit « 1a » ça aurait été bien qu'elle dise « grand 1 » par exemple.

48 – X, 45 – O : Les changements de voix tu t'y repère bien. [ou encore] La voix masculine c'était un peu plus la voix off, ça de le savoir ça m'a permis de me repérer, ça m'a aidé à structurer le texte.

## Sons spatialisé & icônes audio

48 – X, 28 – X, 9 – O : J'ai pas du tout remarqué le son spatialisé pendant le texte alors que dans l'explication je l'entendais. *[ou encore]* J'ai zappé l'histoire du son qui passe de gauche à droite.

36 – X : le son spatialisé de droite à gauche m'a gêné.

24 X : Avec le son spatialisé on n'entend pas bien ce qu'il dit à la fin d'un paragraphe parce que le volume diminue quand la voix va sur la droite.

29 – O : Le son spatialisé c'est super bien.

32 – X, 56 – X, 48 – X : Les sons des titres je les ai pas forcément utilisé ; l'histoire de grave / aigu j'ai pas du tout perçu. Je me repérais plus aux mots qu'aux sons *[ou encore]* je n'ai pas trouvé les différents sons intuitifs.

29 – O, 56 – X, 53 – O : Les sons des titres pour me repérer ça m'a aidé *[ou encore]* C'est plus les repères des paragraphes que j'ai trouvé utile, les sons par exemple. *[ou encore]* : J'ai senti le son plus aigu au titre de niveau 1.

## Voix de synthèse en elle même

59 – N, 58 – J, 50 – J, 44 – X, 3 – N : Pour moi le rythme de la diction était trop rapide pour bien comprendre le contenu. Ça manquait d'aération. *[ou encore]* Ça m'aurait aidé d'avoir un rythme plus lent pour la voix.

55 – N, 46 – J, 45 – O, 43 – N, 36 – X, 35 – N, 26 – J, 23 – N, 19 – N, 8-J, 4 – X, 8 - X : Parfois la synthèse avale des mots. Voix de synthèse difficilement compréhensible sur certains mots. *[ou encore]* Moi c'est la voix robotique que j'ai du mal à bien comprendre.

55 – N, 47 – N, 31 – N : y'a aucune structure ! *[ou encore]* on sent pas les parties c'est dur de repérer les fins de phrases, les enchainements des parties !

49 – O : il y a des coupures (les emphases) qui font que la voix coupe l'article avant la coupure.

## Suggestions / idées : Restitution Orale

2 – J : Ca m'aurait vachement aidé d'avoir un truc qui dit « fin de la remarque ».

13 – O : C'est un peu dur quand on est dans des modes (table des matières et je pensais que j'étais en mode lecture) donc ça serait bien d'avoir plus d'indicateurs pour savoir où on est exactement.

24 – X : Parfois quand je parcourais le document je me disais que s'aurait été bien d'avoir un son en plus, y'avait le son des titres et un autre son pour le reste. Par exemple un son pour une image.

59 – N : Ce serait bien d'avoir section / sous-section à chaque fois d'annoncées, pour bien comprendre.

49 – O : j'aurais aimé avoir les rubriques incluses dans le sommaire.

## Suggestions / idées : Fonctions liées aux titres et à la table des matières

31 – O : aurait aimé pouvoir naviguer par niveau hiérarchique dans la table des matières (les titres de niveau X uniquement).

33 – O : flèches du haut et bas pour aller de titre en titre.

40 – X, 15 – N, 18 – J : Pour naviguer plus facilement entre les chapitres il faudrait des fonctions *[ou encore]* Ce n'est pas si pratique d'utiliser que les flèches pour se déplacer, pour trouver un titre par exemple faut bien écouter les bruits, ça serait mieux de pouvoir directement se déplacer entre les titres ou les sous titres.

59 – N, 58 – J, 50 – J, 40 – X, 39 – N, 35 – N, 31 – N, 27 – J, 22 – J, 14- J , 6 – J : Le seul truc qui manque c'est l'arbre, ça serait plus simple de se balader, tu vois je prends tous les frères de (comme « vi hi ? »). Il faudrait une **table des matières**, surtout pour les gros documents. *[ou encore]* « il faudrait pouvoir naviguer dans la structure *[ou encore]* C'est bien mais si je veux répéter la section c'est pas si facile avec les touches comme ça ; ça m'a découragé d'utiliser les touches de navigation car c'était difficile de cibler la partie du texte où je voulais revenir.

9 – O : J'aurai bien aimé avoir la possibilité de choisir la granularité dans le mode entrée.

## Suggestions / idées : Autres



4 - X : manque une fonction pour revenir au début du document.

32 - X : fonction pour sauter toute une liste (et aller à la suite)

52 - X, 45 - O ce serait bien de pouvoir mettre en pause au milieu du paragraphe, ça m'a pas manqué pour la tâche mais pour les paragraphes longs ça me semble utile [ou encore] Ce qui m'a manqué c'était de revenir en arrière au sein d'un paragraphe.

33 - O : avec un peu d'apprentissage on pourrait rajouter plus de fonctions utiles.

41 - O, 40 - X, 38 - J, 34 - J, 32 - X, 31 - N: j'aurais bien aimé d'autres touches [fonctions] pour sauter les pages et des touches pour sauter les chapitres / sous chapitres. [ou encore] J'aurais aimé pouvoir chercher des mots clés. [ou encore] pouvoir aller directement aux notes de bas de page.

56 - X, 53 - O, 31 -J, 28 - X : j'aurais besoin de pouvoir prendre des notes en parallèle.

16 - X: il n'y a pas de composante esthétique à l'audio, contrairement au visuel qui peut imprimer une certaine émotion esthétique (et qui aide d'ailleurs probablement à se rappeler du document, on sait que les émotions sont des marqueurs mnésiques aident à ancrer les choses.

33 - O : s'aurait été bien de pouvoir avoir une seule image globale en tactile.

45 - O : Ce que je trouvais bizarre c'était d'appuyer sur les flèches ou espace pour sortir du mode ECHAP ; pourquoi ne pas appuyer de nouveau sur ECHAP comme pour la table des matières ?

60 - X : Mais ça pas applicable qu'aux non-voyants moi j'ai une fille qui est dyspraxique donc elle n'arrive pas à comprendre ce qui est écrit.

## 4.2. Retours qualitatifs des sujets déficients visuels

Retours des sujets déficients visuels recueillis dans le cadre de l'expérimentation présentée en Chapitre V.5. Pour chaque retour on précise le numéro du sujet ; quand la remarque est spécifique à une des 2 stratégies de restitution évaluées (optimale - O ou lecteur d'écran - J) on précisera le code de la stratégie.

### Ressenti général

7,11, 10 : j'ai tout aimé [...] J'espère que ce système sera réutilisé ! Je suis tout à fait emballé. Ça me paraît être très utile [...] mais tout le plus que vous apportez ici en termes de structuration, je trouve que c'est un réel plus et je trouverais ça dommage de pas l'utiliser dans une lecture pour les non-voyants.

9 : moi je suis plus plage braille de toutes façons. La TTS au bout d'un moment ça me gave, d'où le fait que je n'ai pas relu le texte. Quand je lis quelque chose en audio il faut que je fasse autre chose, que je sois active, je ne peux pas rester planter devant l'ordi.

### Contenu

4 : Dans l'autre texte les notes de bas de page m'ont un peu perturbé.

4 : y a des choses qui ressemblent à des titre et on a une deuxième structure du document qui est la structure par page.

### Navigation & fonctions implémentées

1, 2, 3, 8, 4, 7, 12 : j'ai préféré la lecture avec la table des matières [...] pour un document complexe j'ai préféré (O) [...] le premier document (O) était meilleur, cela simplifie les choses [...] table des matières OK [...] Après pour naviguer c'est bien d'avoir entrée pour se déplacer [...] J'ai beaucoup utilisé la table des matières. [...] Avec l'annonce du plan on a déjà un schéma dans notre tête c'est plus facile de retenir. [...] J'ai bien aimé avoir une vision avec la table des matières.

2 : la table des matières c'était un peu lourd je n'ai pas tout compris.

4 : (dans la table des matières) C'est bien d'annoncer dans le début tous les titres qu'il y a par parties.

5 : Le déplacement est pas réactif par rapport aux touches : quand j'appui j'ai l'impression qu'il faut que j'attende avant d'avoir la voix.

8, 10, 12 : si pour naviguer sur une page internet ça pouvait être aussi simple j'aimerais bien l'utiliser fréquemment, JAWS c'est plus compliqué ça utilise une série de raccourcis clavier, ça lit une série de liens, pour les sauter il faut descendre [...] Moi ce que je reproche à JAWS c'est qu'il y a 1500000 raccourcis et que c'est pas simple d'utilisation ; après on apprend les touches dont on a besoin et on ne se sert que de celles-là ; [...] sur JAWS par exemple pour lire un texte c'est la croix et la bannière. Y'a 100000 raccourcis claviers j'ai jamais tout retenu ; ne serait-ce que pour lire un texte faut appuyer sur 4 touches en même temps, je les connais toujours pas moi.

11 : J'aime bien que ça soit par paragraphe, parce que c'est pénible par ligne d'avoir tout le temps à cliquer sur la flèche du bas, c'est pratique.

### Restitution orale du logiciel

1, 2, 3, 4, 6, 7, 10: son 3D très bien, super [...] je n'ai pas fait trop attention mais ça peut être pas mal [...] Le son spatialisé est surprenant, ça enrichie les infos sur la longueur d'une ligne, ça donne une notion, ce n'est pas indispensable mais pour un cours ça me paraît être un avantage. [...] J'ai bien aimé le coté spatialisé, ça donne une info supplémentaire sur la longueur du paragraphe, alors après je ne sais pas ce qu'on en fait dans le phénomène de mémorisation. [...] Le 3DS je pensais que ça allait être perturbant [...] mais pour ce type de document où on a besoin de comprendre la structure et de suivre les paragraphes, là ouais. [...] Le 3DS ça m'a permis de voir quand c'était la fin des paragraphes, alors on est plus attentifs on se dit ah oui on passe à autre chose

5, 9, 10 : Lorsque la voix part à droite ça m'endort. Ce déplacement gauche/droite ce n'est pas logique [...] Moi le 3DS ça me va pas du tout ; certaines phrases étaient coupées [...] Après ce que je trouve inutile c'est le 3DS, je trouve ça déconcentre plus que ça aide, parce qu'on entend moins bien à des moments. Moi je ne vois pas trop l'utilité.

6, 10 : moi je l'entendais mais ça m'a pas servi plus que ça [...] Le 3DS m'a pas aidé, c'est tellement bien structuré qu'on s'en fiche, sauf si le paragraphe est très long et qu'on se dit mince ou est-ce que j'en suis ? C'est un plus mais c'est nécessaire.

1, 2,3 ,4,7 ,10 : les bips c'est bien [...] même sans vraiment les intellectualiser on les repère, j'aime bien [...] Les sons annonçant les titre dans le premier étaient très bien, le premier doc été le meilleur, cela simplifie les choses [...] Les titres c'est bien le son, peut être un petit peu plus fort [...] Les BIP je les ai utilisé pour me balader dans la table des matières et le document, quand on passe d'une partie à l'autre on arrive bien à se situer, à visualiser le texte [...] Les BIP c'est pas mal surtout quand on recherche quelque chose

4, 5, 8, 9, 12 : Les bips pour les titres je suis passé outre [...] Les BIP donnent une information mais y'a tellement de choses on ne peut pas faire attention à tout. Ça aide, à se localiser dans le document [...] moi je l'entendais mais ça m'a pas servi plus que ça. Déjà deux voix différentes ça suffit pour se repérer [...] les BIP j'y ai pas fait attention trop si tu révises ça peut être utile tu peux naviguer plus vite, tu retiens à quelle hauteur est une sous partie donc tu peux aller plus vite

5, 8, 9,12 : J'ai fait attention aux deux voix, est ce que ça m'a été utile : peut-être. [...] par rapport à la vision globale du truc la voix off aide beaucoup [...] les deux voix ça me va bien déjà [...] Après pourquoi pas avoir une voix off, pourquoi pas ça change.

10 : c'est bien que ça fasse une pause quand ça reprend au début du paragraphe

10 1J, la prosodie est parfois bizarre

3 : Une fois qu'on a compris a quoi le son qui annonçait les listes servait il est très utile.

4 : La mise en évidence de certains mots je ne sais pas quoi en penser, je ne sais pas dire qu'elle influence ça a sur ma compréhension

9 : certaines phrases étaient coupées [restitution des emphases au sein des paragraphes]

### Suggestions & idées

1 : éloquence a une prosodie particulière en fin de paragraphe ?? (NB : n'est pas en fin de phrase plutôt ?)

10 : si je n'avais pas eu de visuel pour avoir la structure, j'aurais écrit un plan pour avoir la structure

; moi pour apprendre je prends des notes pour me faire une synthèse du truc.

10 : Chaque déficient est différent après, c'est ça la difficulté y'en a pas 2 qui se ressemblent ; les aveugles c'est plus homogène ; déjà pas 2 qui ont la même acuité visuelle ; et s'ils ont la même acuité visuelle sur papier ils ont pas eu la même éducation, la même formation, et se servent pas des outils de la même façon ; certains vont se comporter comme des voyants d'autres comme des aveugles ; autre exemples y'a des gens qui ont la même acuité visuelle que moi et qui font 1000000 fois moins de choses que moi ; ça dépend de plein de paramètres.

10 : Si ce système doit devenir un lecteur d'écran, il faut aussi prendre en charge l'écriture.

6 : L'aigu/grave pour les BIP c'est dur à distinguer, faudrait plus d'écarts de tons.

5 : J'aurais aimé pouvoir distinguer la lecture du déplacement. Dans la table des matières se déplacer de haut en bas. Il faudrait dissocier la fonction de lecture du déplacement.

6 : pour moi un texte se navigue de haut en bas, j'aurais utilisé les flèches haut et bas plutôt que droite et gauche (ça marche comme ça sous Windows).

6 : Si on ne sait pas qu'on est dans la table des matières la voix ne le dit pas, on aurait pu mettre une touche pour savoir dans quel mode on est.

6, 4 : J'aurais aimé pouvoir naviguer au sein de chaque paragraphe, de phrase en phrase, pas forcément tout me retaper depuis le début.

### 4.3. Fonctions R utilisées pour les tests statistiques

---

Ci-dessous la liste de fonctions utilisées dans le langage de programmation *R* pour les tests statistiques appliqués aux données des expérimentations 4 et Chapitre V5. Les fonctions sont toutes présentes dans la bibliothèque « stats » incluse par défaut dans l'environnement de travail du logiciel *R*. Une description plus complète de chacune de ces fonctions peut être trouvée ici : <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>

- § **aov** : analyse de la variance à 1 facteur ; comparaison des moyennes des échantillons.
- § **TukeyHSD** : test post-hoc de comparaisons multiples réalisés sur les résultats de la fonction **aov**.
- § **Kruskal.test** : comparaison des rangs des observations
- § **kruskalmc** : test post-hoc de comparaisons multiples sur les rangs des observations (bibliothèques de fonctions *R* *pgirmess*)
- § **t.test** : test de Student de comparaison des moyennes sur données appariées.
- § **wilcox.test** : test de Wilcoxon de comparaison des rangs des échantillons sur données appariées.

## 4.4. Documents utilisés pour les expérimentations 2 et 3

Le document portant sur l'inflation a été utilisée pour les deux expérimentations ; le document portant sur l'union européenne n'a été utilisé que pour la 3<sup>ème</sup> expérimentation, auprès de sujets déficients visuels. Les versions originales (mises en forme) de ces deux documents sont reprises ci-après :

# L'inflation et la politique de stabilité du niveau général des prix

Distinguons les différentes notions liées au niveau des prix :

**L'inflation** : l'inflation se définit comme la **hausse continue du niveau général des prix**. Toute hausse des prix n'est pas forcément synonyme d'inflation ; pour être qualifié de phénomène inflationniste, le mouvement de hausse des prix doit être général, c'est-à-dire concerner la majorité des branches de l'économie ; il faut, en fait, que le mouvement de hausse l'emporte sur les baisses de prix et se maintienne dans le temps (une hausse en entraîne d'autres).

**La désinflation** : la désinflation se définit comme le **ralentissement durable du rythme de hausse** du niveau général des prix. Ainsi, lorsque le taux d'inflation passe de 10 % à 5 %, on dit qu'il y a désinflation.

**La déflation** : la déflation correspond à la **baisse continue** du niveau général des prix (taux négatifs).

**La stagflation** : la stagflation correspond à la conjonction d'un **ralentissement** de l'activité économique (baisse du PIB), d'une **forte inflation** et du **chômage**.

- **1. Mesure et évolution du phénomène inflationniste**
- **A. La mesure de l'inflation**

On mesure l'inflation en calculant l'**Indice des Prix à la Consommation** (l'IPC). Ce calcul s'effectue tous les mois. C'est l'INSEE qui est chargé de l'établir.

### • Comment calcule-t-on un indice des prix ?

Chaque mois, l'INSEE<sup>129</sup> mesure l'**évolution des prix** des produits jugés **représentatifs** de la consommation des ménages français. Pour cela, il établit un indice des prix ; cet indice est composé de 305 postes de dépenses (alimentation, chauffage, logement, habillement...), soit un suivi de plus de 200 000 prix.

*Remarque* : Pour répondre à un des critères de convergence de Maastricht (la stabilité des prix), des harmonisations entre les différents indices ont été menées dans chaque État membre. Aujourd'hui, les indices des prix à la consommation harmonisée (IPCH) permettent des comparaisons entre pays européens. Toutefois, l'indice national reste l'indice de référence pour les négociations salariales.

Chacun des postes de dépenses sélectionnés est pondéré (c'est-à-dire qu'à chaque poste est affecté un coefficient d'importance). Cet échantillon est fixé pour l'année et actualisé tous les ans.

<sup>129</sup> Institut National de la statistique et des études économiques

## • Comment calcule-t-on un taux d'inflation ?

Le taux d'inflation correspond au taux de **variation de l'indice des prix à la consommation** entre deux années ; on le calcul de la manière suivante :

$$\text{Taux d'inflation} = \frac{\text{Indice de l'année (N+1)} - \text{Indice de l'année (N)}}{\text{Indice de l'année (N)}} \times 100$$

## • B. L'évolution de l'inflation au cours de ces dernières décennies

### • Avant 1974, une inflation « de croissance »

La période de l'après-guerre est une période marquée par un taux de croissance et ce pour au moins 3 raisons :

- les investissements pour la reconstruction
- le développement du crédit
- une forte demande

Cette forte croissance s'accompagne d'inflation puisque **la demande est souvent supérieure à l'offre.**

*Point Historique :* La seconde guerre mondiale prit fin le 2 Septembre 1945 par une victoire des alliés contre l'Axe. La reconstruction fut des plus difficiles pour les nombreux pays affaiblis par la guerre.

fort

### • De 1974 au début des années 1980 : une période d'inflation élevée

L'inflation est l'une des manifestations majeures de la crise économique née en 1974. Durant cette période, les économies développées connaissent une situation de **stagflation** :

- les entreprises doivent **augmenter leurs prix** parce que **les coûts de production augmentent** (prix du pétrole<sup>130</sup>)

- les consommateurs **empruntent** car ils espèrent que la hausse des prix diminuera le poids des remboursements; ils consomment davantage, d'où, une nouvelle hausse des prix (demande supérieure à l'offre) ;

- les hausses de prix entraînent des **hausses de salaires** (les salariés vont revendiquer des augmentations afin de préserver leur pouvoir d'achat) : c'est ce que l'on appelle une **spirale inflationniste**

- la **croissance** quant à elle, stagne ; le chômage commence à se développer

### • À partir de 1983 : la désinflation

Une politique monétaire **restrictive** (hausse des taux d'intérêt qui génère une baisse des crédits demandés) est menée pour lutter contre l'inflation ; elle aboutit au milieu des années 1980 à une **désinflation** dans la plupart des pays développés. Enfin, **l'adoption de l'euro** impose une **maîtrise de l'inflation** qui devient l'une des préoccupations majeures des gouvernements européens. En France, le taux d'inflation est passé de près de 14 % en 1981 à moins de 2 % en 2005.

### • La crainte de la déflation

<sup>130</sup> Il peut être intéressant de voir des documents sur les chocs pétroliers de 1973 et 1979

Les taux d'inflation actuels sont faibles et la crainte de la déflation se fait sentir. En effet, la déflation diminue les **marges bénéficiaires des entreprises**, les obligeant à réduire leur niveau de production et donc parfois à licencier. Cela entraîne un accroissement du chômage, une tendance à la baisse des salaires et le risque de voir s'installer une **récession** économique.

## · 2. Les causes de l'inflation

*Remarque :* Il existe d'autres causes de l'inflation que celle présentées ci-après, notamment l'excès de masse monétaire, les prix d'importation ainsi que certains facteurs psychologiques.

### · A. L'inflation par les coûts

Les coûts sont le principal élément constitutif du prix de vente d'un bien<sup>131</sup>. Ainsi, lorsque les coûts de production augmentent, cette hausse se répercute automatiquement sur le prix de vente au consommateur. L'accroissement des coûts de production peut provenir de :

- des **revendications salariales** qui aboutissent à une hausse du coût du travail
  - demandes d'augmentations de salaire (primes de risques...)
  - demandes de diminution du temps de travail
- une **hausse des prix des matières premières** (on parlera dans ce cas d'inflation importée parce que la hausse des prix résulte ici de celle de produits étrangers importés en France) ;
- certaines **mesures gouvernementales** renchérissent le prix de revient :
  - une hausse des prélèvements obligatoires (hausse de la TVA, des charges sociales)
  - un alourdissement des contraintes en matière de sécurité
- des **stratégies commerciales** menées par les entreprises : dépensent en matière publicitaire, promotionnelle

Quelle qu'en soit l'origine, la **hausse des coûts de production entraîne une hausse du prix de vente**, donc un phénomène inflationniste. En outre, l'inflation se propage à toute l'économie parce que les prix de vente de certaines entreprises constituent les éléments du coût de revient d'autres entreprises.

### · B. L'inflation par la demande : l'analyse keynésienne

Pour l'analyse keynésienne<sup>132</sup>, il y a inflation, lorsque **la demande excède l'offre**. L'origine de l'inflation s'explique alors par **l'inélasticité de l'offre par rapport à la demande** (c'est-à-dire l'incapacité de l'offre à répondre à la demande en raison, entre autres, du plein emploi des facteurs, de l'insuffisance des stocks, d'une productivité insuffisante des facteurs travail et capital...).

**Plusieurs éléments** peuvent expliquer l'**excès de demande globale** : l'accroissement des dépenses publiques ; une politique monétaire expansive (la baisse des taux d'intérêt favorise les crédits) ; une augmentation des revenus (hausse du SMIC, des retraites, des salaires...) ; les changements de comportements d'achat : phénomènes de mode (téléphones portables), anticipations positives des consommateurs sur l'avenir...

<sup>131</sup> Le prix de vente d'un bien est constitué d'une part des coûts de production et d'autre part de la marge bénéficiaire fixée par le vendeur.

<sup>132</sup> Voir documentation sur John Maynard Keynes et le Keynésianisme



### · C. Les explications structurelles de l'inflation

L'inflation trouve ici son explication dans les dysfonctionnements de l'économie : les **formes de marchés non concurrentiels** : monopoles, cartels, oligopoles qui imposent des prix plus élevés que ceux qui résulteraient du libre jeu de la concurrence ; les **conflits entre groupes sociaux** (patronat et syndicats principalement) à l'occasion du partage de la valeur ajoutée (grèves pour préserver le pouvoir d'achat des salariés) ; les **systèmes d'indexation de certains types de revenus** sur l'évolution des prix (SMIC, prix garantis de certains produits agricoles...).

*Remarque* : D'après Raymond Barre, les prix résultant de situations de concurrence imparfaites dans l'industrie ou les prix fixés par les pouvoirs publics dans le secteur agricole contribuent également à l'inflation.

### · 3. Les différents effets de l'inflation

Ci-après une liste des effets positifs et négatifs de l'inflation :

- L'inflation entraîne une **diminution du pouvoir d'achat de la monnaie** (avec la même somme d'argent, on peut acheter moins de biens). Les agents économiques se débarrassent alors de leurs liquidités au profit de « **valeurs refuges** » (or, tableaux, biens immobiliers, devises fortes...). On parle de « fuite devant la monnaie ».

- Les titulaires de revenus fixes (retraités, épargnants ayant des placements à taux fixe, les salariés non indexés) subissent une **perte automatique de leur pouvoir d'achat**.

- L'inflation **diminue la compétitivité des produits nationaux** par rapport à leurs concurrents internationaux. La balance des paiements présente alors un solde négatif.

- L'inflation **allège le poids de la dette** : les entreprises, en augmentant leurs prix (donc leur chiffre d'affaires en valeur), diminuent le poids relatif de leurs emprunts ; en outre, les entreprises désireuses d'éviter les conflits sociaux, accorderont plus facilement des augmentations de salaires. Leurs salaires ayant augmenté, les ménages pourront donc plus facilement supporter le poids de leurs dettes.

- En période inflationniste, les **salaires augmentent**, les recettes fiscales de l'État (impôts et cotisations sociales principalement) sont donc plus importantes.

## L'Union Européenne et la mondialisation

### · 1. l'UE : Un pôle majeur de la mondialisation

La **mondialisation** désigne l'ouverture croissante des économies nationales sur le monde. Elle implique une augmentation des flux d'échanges de marchandises, de circulation des capitaux et des informations ainsi que des déplacements des populations. Ce processus conduit à une interdépendance des États et à une contraction de l'espace mondial.

#### · A. Le premier pôle commercial de la planète

##### • L'union fait la force

Première **puissance économique de la planète** devant les États-Unis, la Chine et le Japon, la production de richesses de l'UE représente 28% du PIB (Produit intérieur brut) mondial. Elle constitue également la **première puissance commerciale** et assure par exemple 38% du commerce mondial de marchandises. Il faut toutefois noter

que deux tiers de ces échanges correspondent à un commerce intra-communautaire (entre états membres). Ces économies européennes sont donc très complémentaires, elles assurent un rayonnement international dans les trois secteurs d'activités.

*Remarque* : cinq états membres de l'UE font partie des dix principales puissances économiques mondiales : l'Allemagne, la France, le Royaume-Uni, l'Italie et l'Espagne.

#### • **Les points forts de l'UE dans le domaine commercial**

Il faut insister sur les points forts qui permettent à l'UE de réaliser 17% des échanges mondiaux :

- Elle constitue la **deuxième puissance agroalimentaire** mondiale, avec une production excédentaire qui fournit une puissante industrie, restant malgré tout dépendante des subventions de la politique agricole commune (PAC européenne)
- l'UE est **un géant industriel** (industrie sidérurgique, industrie automobile...) ayant cependant une économie post-industrielle et un secteur tertiaire représentant 70% des emplois et du PIB
- l'UE dispose d'atouts dans le domaine **des technologies de pointe**. C'est le cas en particulier de l'aérospatiale avec Ariane Espace et de l'aéronautique avec Airbus.

*Gros plan sur une entreprise florissante* : Airbus est, avec Boeing, le premier producteur aéronautique mondial. L'Airbus A380, lancé en octobre 2007 est le premier avion gros porteur disposant de deux ponts et capable de transporter 850 passagers. Il constitue une vitrine du savoir-faire européen à l'exportation.

### • **B. Un pôle émetteur et récepteur de flux invisibles**

Outre la domination dans le commerce de marchandises, l'UE **polarise également les échanges de services et les flux invisibles de capitaux et d'informations**. Elle occupe la première place dans les échanges internationaux de services avec 25% de ces échanges (contre 22% pour les États-Unis). Ce secteur est essentiel puisqu'il représente **70% de la richesse totale produite** (PIB) et révèle une économie postindustrielle, capable de s'adapter aux exigences de la concurrence internationale.

Là aussi, l'UE dispose de secteurs clés comme (1) le tourisme, (2) les banques et les assurances, ou encore (3) les télécommunications.

(1) l'UE est en effet le **premier pôle touristique mondial** avec notamment ses grandes capitales (voir la fiche *La France, pôle touristique mondial*). (2) Elle représente le premier pôle émetteur et récepteur d'**investissements** (IDE : investissements directs à l'étranger) dans le monde. De plus les grandes places boursières européennes parmi les plus importantes à l'échelle mondiale. L'UE figurent est donc une importante **puissance financière**, cette puissance est matérialisée par le statut de l'euro qui facilite et dynamise les échanges, et rivalise avec le dollar. (3) Enfin, L'UE dispose de l'un des meilleurs taux de connexion des ménages à Internet. Sa production cinématographique et télévisuelle s'exporte partout dans le monde.

*Les IDE (investissements directs à l'étranger), désignent les capitaux placés à l'étranger et investis dans la création d'entreprises ou toutes formes d'activités économiques (le stock d'IDE, entrants et sortants, approche les 7500 milliards de dollars en 2010).*

### • **C. De nombreux flux humains**

Malgré un contrôle plus strict sur ce type de flux depuis la mise en place de l'**espace Schengen**, l'UE demeure le premier foyer d'immigration devant les États-Unis. Cette immigration est avant tout une immigration du travail, qui témoigne de l'attractivité du centre européen. Les flux humains englobent également des déplacements temporaires comme les flux touristiques : **60% des flux touristiques internationaux sont à destination de pays de l'UE**.

*Remarque* : L'Espace Schengen a été créé en 1995 et compte 26 états membres. Il s'étend sur plus de 4 millions de kilomètres carrés. Schengen désigne la ville du Luxembourg où a été signée la convention du même nom.

## • **2. Les acteurs de la mondialisation**

### • **A. Les populations, premiers acteurs de la mondialisation**

Troisième foyer de peuplement mondial, l'UE compte **plus de 500 millions d'habitants**. Cette population dispose d'un niveau de vie élevé et donc, malgré d'évidentes inégalités, d'un pouvoir d'achat élevé (le PIB moyen est de 25



000 € par habitant). Le centre européen est par conséquent un des marchés de consommation les plus importants et les plus attractifs de la planète.

Les sociétés étrangères n'hésitent guère à investir en Europe : haut **niveau de formation** (réel savoir-faire et assurance d'une production de qualité), grande **stabilité de ces sociétés** (sécurité d'investissement), etc.

## · **B. Les différents acteurs du secteur privé**

### • **Les entreprises internationales**

Les **FMN** (firmes multinationales) ou **FTN** (firmes transnationales) sont des entreprises implantées dans de nombreux pays (au moins 6) et elles réalisent une grande partie de leur chiffre d'affaire hors de leur pays d'origine. Ces entreprises, de par leur rayonnement, sont au cœur du processus de mondialisation. **L'UE représente l'une des plus fortes concentrations de FMN à l'échelle mondiale**, avec un peu plus de 30% des 500 premières FMN à cette échelle.

*Remarques* : Les secteurs d'activités sont très divers. Les FMN européennes dominent en particulier dans **le secteur de l'énergie et du pétrole** : la Royal Deutch Shell (conglomérat anglo-néerlandais) se classe au 2<sup>e</sup> rang mondial. British Pétroleum (BP) au 4<sup>e</sup>. On trouve également 6 firmes européennes parmi les 22 premières se situant dans le **secteur des services**.

Ces firmes concentrent leurs sièges sociaux, leurs centres de recherches dans les grandes métropoles du continent tandis que les activités productives se délocalisent de plus en plus vers les pays où la main d'œuvre coûte moins chère (chômage européen).

### • **Les organisations non gouvernementales**

Parmi les acteurs privés, les **ONG** (organisations non gouvernementales<sup>133</sup>) jouent un rôle majeur dans la production de services ou dans la circulation d'informations. Certaines de ces ONG européennes ont une action essentielle permettant **le rayonnement d'un modèle européen**. Médecins sans frontières, la Croix Rouge internationale œuvrent ainsi dans un souci d'action humanitaire et contribuent à **renforcer l'image** d'une UE soucieuse du **bien-être** des populations et du **respect des droits fondamentaux**.

## · **C. Les Etats**

Réunis au sein d'une UE qui est aussi politique, **les États impulsent des choix de politique économique**. Ils favorisent, plus souvent qu'ils ne restreignent, la libération économique et donc les échanges. D'un point de vue mondial, de nombreux membres jouent un rôle important dans de grandes institutions internationales :

- Le **G20**, qui rassemble 19 pays en plus de l'UE, joue un rôle majeur. Il est le symbole de la gouvernance économique mondiale, puisqu'il représente 85% du commerce mondial et 90% du produit mondial brut (somme des PIB de tous les pays).
- Les États européens sont également représentés au sein de **l'OMC** (organisation mondiale du commerce) par le commissaire européen au commerce, qui parle en leur nom.
- le **Fonds Monétaire International** (FMI), autre pôle de gouvernance mondiale, est dirigé en 2011 par une française, Christine Lagarde.

## · **3. Les faiblesses de l'UE dans la mondialisation**

### • **Des faiblesses politiques**

<sup>133</sup> Les ONG sont des organisations d'intérêts publics qui ne relèvent ni de l'état ni d'institutions internationales.



Si l'influence économique de l'UE est certaine, **son poids politique demeure faible face aux puissances mondiales que sont les États-Unis et la Chine**. Elle a du mal à être identifiée comme un ensemble cohérent, aux choix politiques convergents : politique extérieure, choix en matière de sécurité et de paix (les Européens ont été très divisés sur la question de l'intervention militaire en Irak). Enfin, les intérêts nationaux **prévalent sur la logique communautaire** et la dépendance vis-à-vis des États-Unis demeure forte dans ce domaine.

- **Des handicaps sociaux et économiques**

Si l'on se projette sur un long terme, le **vieillesse démographique** rapide du continent est inquiétant quant à la capacité de l'UE à maintenir son statut de puissance. La faible natalité (de moins de 10 pour mille par an en moyenne) et l'espérance de vie élevée posent le problème de l'augmentation de certains déficits (dépenses de santé, financement des retraites). Ce problème peut réduire l'attractivité de l'UE face à la concurrence de nouvelles puissances, au dynamisme de pays émergents.



**Les inquiétudes sont aussi économiques** : retards dans le domaine de la Recherche et Développements ; **fuite des cerveaux** vers les États-Unis ; un niveau d'endettement important qui fragilise aussi les économies. Enfin la crise financière mondiale survenue en 2010 impacte fortement certaines économies (Grèce, Espagne) et menace la stabilité de la zone euro.

#### 4.5. Vocabulaire d'annotation utilisé dans les documents des évaluations

Pour les expérimentations 2 et 3, nous avons formalisée la MFM de deux documents expositifs, qui était annotée au fichier XML représentant le document d'origine, afin de pouvoir évaluer différentes stratégies de restitution sur la base de ces documents. Leur MFM était relativement simple (voir section précédente), et comportait seulement quelques objets textuels particuliers que l'on retrouve dans le vocabulaire XML utilisé pour l'annotation.

Classe d'objet de MFM	Balises XML correspondantes	Attributs des balises et commentaires
Regroupement	<regroup>	id identifiant role le rôle est un attribut générique dont la valeur dépend de l'objet auquel il est appliqué (les autres classes héritent de ces attributs)
Page	<page>	numéro
Bloc de texte	<bdt>	-
Paragraphe	<par>	-
Preview	<preview>	description le texte de la vue d'ensemble décrivant le ou les objets subsequents
Emphase	<emph>	-

Mot-clé	<keyword>	Aucun attribut. Objet particulier retrouvé dans les documents du corpus.
Marqueur d'ordre	<mq>	type_numerotation système de numérotation utilisé par le marqueur
Syntaxe	<syntax>	description Ces objets textuels sont réalisés par des marques lexico-syntaxiques et correspondent à des utilisations des parenthèses ou des guillemets particulières, qui sont explicitées verbalement par cet attribut.
Titre	<titre>	niveau
Section	<section>	split attribut permettant de savoir si la section est divisée, étalée sur plusieurs pages
Item d'énumération	<item>	-
Enumération	<enum>	nb_items Nombre d'items de l'énumération nature_items Nature des items de l'énumération
Image	<img>	Description Description du contenu de l'image

#### Vocabulaire XML utilisé pour annoter la MFM des documents utilisés pour les expérimentations

Les relations suivantes étaient prises en compte dans l'annotation du document : composer, intituler, ordonner, hiérarchiser, labelliser, illustrer, numéroter et associer. Chaque relation était annotée au document sous la forme : `<rel verbe="id_relation" origines="id_objet1" cibles="id_objet2 ;id_objet3" />`. On autorise une relation à avoir plusieurs origines et/ou plusieurs cibles. Par exemple pour la relation de composition, plusieurs objets (origines) vont composer un seul objet cible, comme une section.

## 4.6. Consignes et questions des expérimentations 2 et 3

[CONDITIONS EXPERIMENTALE : PAS DE TELEPHONE, PAS DE DISTRACTION  
PENDANT 1'XP

• Discours présentation générale + signature du formulaire de consentement : Il est important de garder à l'esprit que j'évalue le système et non pas vos capacités ; ce qui ne doit pas vous empêcher d'essayer de suivre la consigne au mieux.

Date & Heure de début : .....

Tranche d'âge : .....

Niveau scolaire / profession : .....

Déficience (cécité / gros caractères / moyens / léger... de naissance ?) : .....

Adresse Mail : (pour bon FNAC) .....

• Informatique : (quotidien, N semaine, 1 semaine, 1 mois, moins 1 mois)

- Fréquence d'utilisation de l'ordinateur + lecteur d'écran ? lesquels ?.....
- Fréquence d'utilisation Smartphone ?.....
- Fréquence d'utilisation de la TTS ?.....

• Déroulement de l'expérimentation :

- Tu vas parcourir un document audio avec le logiciel READ. Tout d'abord vous allez écouter les consignes vous expliquant comment fonctionne le logiciel. [*document exemple dans la bonne condition*]
- Maintenant vous allez lire un document pédagogique, qui aborde des notions de <\_\_\_\_>. Votre objectif sera d'essayer **de comprendre au mieux le contenu du document ainsi que sa structure**. Vous êtes libre de naviguer dans le document comme vous le souhaitez. Vous avez un temps limité mais largement suffisant pour prendre connaissance du texte et de naviguer ; si vous avez terminé avant que le logiciel n'annonce la fin, vous pouvez arrêter.

-----EXPERIMENTATION-----

## Charge cognitive

1 2 3 4 5 6 7 8 9 (*très peu important.....très important*)

- (globale CC) : De manière générale, La compréhension du document et de sa structure m'a demandé un effort mental (échelle de 1 à 9) .....
- (charge intrinsèque) : Le contenu (domaines abordés, concepts) était (1-9 très facile...très complexe) .....
- (charge extrinsèque) : Comprendre la façon dont la synthèse prononçait les mots du texte m'a demandé un effort mental ? (1 à 9) .....
- (charge extrinsèque) : Comprendre la façon dont le texte était présenté à l'oral m'a demandé un effort ... ? (1 à 9) .....
- (germane) : Jusqu'à quel point étais-tu concentré pendant la lecture (1 à 9).....
- (ouverte) Est-ce que votre niveau de concentration a évolué pendant la lecture ? .....

## Désorientation

1 2 3 4 5 : *pas du tout d'accord ..... à tout à fait d'accord*

• Evaluer les propositions suivantes :

- J'arrivais à bien me repérer dans le document .....
- Je me sentais perdu (Ahuja 2001, désorientation) .....
- C'était difficile de retrouver un endroit dans le document auquel je pensais.....
- La navigation était difficile .....
- Je me sentais désorienté.....
- **J'arrivais à discerner ce qui était important du reste.....**

## Utilisabilité [questionnaire SUS]

1 2 3 4 5 : *pas du tout d'accord ..... à tout à fait d'accord*

Questions posées d'un ordre aléatoire.

- J'aimerais utiliser ce système fréquemment .....
- Je trouve ce système inutilement complexe.....
- Je pense que ce système est facile à utiliser .....

- Je pense que j'aurais besoin de l'aide d'un technicien pour être capable d'utiliser ce système ? .....
- J'ai trouvé que les différentes fonctions du système étaient bien intégrées ? .....
- Je pense qu'il y a trop d'incohérences dans ce système.....
- J'imagine que la plupart des gens seraient capable d'apprendre à utiliser ce système rapidement ?.....
- J'ai trouvé ce système très lourd à utiliser ?.....
- Je me sentais très en confiance en utilisant ce système ?.....
- J'ai eu besoin d'apprendre beaucoup de choses avant de pouvoir utiliser ce système ? .....

Demander au sujet : commentaires / Remarques, et condition préférée.

## **Questionnaires et consignes pour le document portant sur l'inflation (économie)**

### • Évaluation des connaissances :

- Évaluation de tes propres connaissances en économie (1 à 5) : .....
- Praticqué au collège / lycée ? .....
- Praticqué en post-bac ? .....
- À un niveau personnel, déjà renseigné sur des notions de macro-économie ? .....
- (après passation) Évaluation de tes propres connaissances vis-à-vis du contenu du document (1 à 5, 1 = rien, 5 = tout)

### • Retour en arrière (ré-ouvrir le document ; ils font comme ils veulent, inclus la VG) :

- (le plus vite possible) Peux-tu naviguer dans le document jusqu'à l'endroit qui traite d'une **période d'inflation élevée ?**
- (le plus vite possible) Peux-tu naviguer dans le document jusqu'à l'endroit qui traite des **causes de l'accroissement des coûts de production ?**

• Restitution du PLAN : Pouvez-vous me redonner le plan du texte en étant le plus précis et complet possible ?

### • Questions de compréhension (préciser SUR LA BASE DU TEXTE) :

- [section 1.B] Pouvez-vous me dire comment a évolué l'inflation depuis la période d'après-guerre ?
- [section 2] Pouvez-vous me donner différents grands types de causes à l'inflation ?
- [section 3] Pouvez-vous me citer un effet positif et un effet négatif de l'inflation ?
- [section 0] pouvez-vous me redonner les notions importantes liées au niveau des prix ?
- [section 2.C] Pouvez-vous me donner des exemples de causes structurelles à l'inflation ?
- [section 1.A] Pouvez-vous me donner des exemples de causes de hausse du coût de production ?
- [intégration] Quels sont les choix politiques (gouvernement) qui peuvent influencer sur l'inflation ?
- [intégration] Est-ce que la mesure de l'inflation est fiable ? pourquoi ?

## Questionnaires et consignes pour le document sur l'union européenne

### • Évaluation des connaissances :

- Évaluation de tes propres connaissances en macro-économie (marchés mondiaux) ? (1 à 5) : .....
- Pratiqué au collège / lycée ? .....
- Pratiqué en post-bac ? .....
- À un niveau personnel, déjà renseigné sur des notions de macro-économie ? .....
- (après passation) Évaluation de tes propres connaissances vis-à-vis du contenu du document è

### • Retours en arrière (ré-ouvrir le document ; ils font comme ils veulent, inclus la VG) :

- Peux-tu naviguer dans le document jusqu'à l'endroit qui traite des **ONG dans la mondialisation** ?
- Peux-tu naviguer dans le document jusqu'à l'endroit qui traite des **points fort de l'UE dans le domaine commercial** ?

### • Restitution du PLAN : Pouvez-vous me redonner le plan du texte en étant le plus précis et complet possible ?

### • Questions de compréhension (préciser SUR LA BASE DU TEXTE):

- [section 2.] D'après le document, quels sont les différents types d'acteurs dans la mondialisation ?
- [section 1.] Quels sont les points forts de l'UE au niveau mondial ?
- [section 1.B] sont les flux invisibles qui transitent par l'UE ? les grands types
- [section 2.A] Pourquoi les populations de l'UE attirent les investisseurs et constituent un acteur majeur dans la mondialisation ?
- [section 3] D'après le document, quels sont les faiblesses économiques de l'UE ?
- [section 2.C] D'après le texte, quels sont les différents organismes internationaux dans lesquels sont impliqués des membres de l'UE ?
- [intégration] En quoi l'aspect communautaire de l'UE peut être une force comme une faiblesse ?
- [intégration] Quels sont les effets pervers de la mondialisation ? et à contrario les effets positifs ?

# Table des matières du manuscrit

Sommaire .....	6
Préambule et indices de lecture .....	1
Introduction.....	2
<i>Chapitre I - Aspects théoriques de la mise en forme des textes, et bilan empirique de ses rôles pour le lecteur .....</i>	<i>7</i>
<b>1. La mise en forme des textes.....</b>	<b>8</b>
1.1 Définitions et bref historique .....	8
1.2 Le rôle de la mise en forme dans la compréhension des textes .....	9
<b>2. Le modèle d'architectures textuelles (MAT) .....</b>	<b>10</b>
2.1 Présentation .....	10
2.2 La Mise en Forme Matérielle (MFM) .....	13
2.2.a Définitions et formalisation .....	13
2.2.b Les principes de contraste et d'identité régissant l'utilisation des marques de MFM .....	17
2.2.c Les deux dimensions de la MFM pour les lecteurs.....	18
2.2.d L'aspect fonctionnel de la MFM & le système visuel humain.....	19
2.3 Analyse des intentions auteuriales.....	22
2.4 Application du MAT : exemples de modélisations d'architectures textuelles .....	23
<b>3. Conclusion.....</b>	<b>25</b>
<i>Chapitre II - Accessibilité des contenus textuels pour les déficients visuels : l'enjeu de la restitution de la mise en forme.....</i>	<i>26</i>
<b>1. Introduction .....</b>	<b>28</b>
<b>2. Définir l'accessibilité pour mieux cerner ses enjeux .....</b>	<b>31</b>
2.1 Le parti pris de l'accessibilité numérique : une accessibilité de moyens universelle.....	31
2.2 De l'accessibilité à l'utilisabilité .....	32
<b>3. Comment les déficients visuels perçoivent et traitent l'information .....</b>	<b>34</b>
3.1 Conséquences cognitives de la déficience visuelle .....	35
3.2 Caractéristiques des modes audio et tactile comparativement au mode visuel : impact sur la cognition .....	36
3.3 Quelles solutions pour pallier aux contraintes des modes audio et tactile ?.....	40
3.4 Conclusion et implications pour l'accessibilité des contenus numériques .....	41
<b>4. Approches actuelles pour l'accessibilité des contenus numériques.....</b>	<b>44</b>
4.1 Approches générales pour la création de contenus accessibles .....	44
4.2 Panorama des formats numériques pour documents textuels.....	45
4.2.a Niveaux de description du contenu : de la présentation à la sémantique.....	46
4.2.b Langages de markup et formats courants .....	46
4.2.c GeM : Un langage de description unifié .....	48
4.3 Les normes pour l'accessibilité des contenus numériques .....	48
4.3.a Portée et limites des « Web Content Accessibility Guidelines ».....	48

4.3.b Des normes basées sur l'annotation sémantique du contenu : « Accessible Rich Internet Applications ».....	50
4.3.c Vérification de la conformité des documents aux normes du W3C.....	52
4.4 Conclusion .....	53
<b>5. Les technologies d'assistance pour l'accès non-visuel aux contenus numériques : approches et limites</b>	<b>54</b>
5.1 Périphériques et modalités d'entrée et de sortie : les bases de l'accès (non-visuel) au numérique .....	54
5.1.a Modalités audio en sortie .....	55
5.1.b Terminaux braille .....	57
5.1.c Dispositifs haptiques particuliers .....	59
5.1.d Périphériques d'entrée .....	60
5.2 Les lecteurs d'écran.....	61
5.2.a Présentation générale des outils existants .....	61
5.2.b Accès aux contenus numériques par les lecteurs d'écran .....	64
5.2.c Accessibilité et utilisabilité des lecteurs d'écran .....	65
5.2.d Les navigateurs Web « audio » .....	68
5.3 Blocs-notes braille .....	69
5.4 Format et lecteurs DAISY.....	70
5.5 Conclusion .....	71
<b>6. Vers la restitution des architectures textuelles pour l'accessibilité des contenus numériques</b> .....	<b>73</b>
6.1 Prise en compte de la MFM par les lecteurs d'écran .....	73
6.2 Exemples empiriques d'apports de la MFM à l'accessibilité des contenus.....	75
6.3 De l'intérêt de la restitution de la sémantique de la MFM : exemples d'architectures textuelles non-restituées.....	76
<b>7. Étude en situation : accès et accessibilité de l'information pédagogique pour les élèves déficients visuels</b> .....	<b>78</b>
7.1 Contexte général : institutions, acteurs et moyens.....	78
7.2 Méthode et données d'investigation .....	79
7.3 Résultats : accès et accessibilité à l'information pédagogique .....	81
7.3.a Documents Braille .....	82
7.3.a.1 Pertes d'informations dans les documents braille .....	82
7.3.a.2 Problèmes d'accessibilités de l'information en braille .....	84
7.3.b Discours oral du professeur .....	85
7.3.c Tableau noir.....	86
7.3.d Bloc-notes braille .....	86
7.4 Récapitulatif des problèmes de disponibilité et d'accessibilité de l'information.....	86
7.5 Discussion.....	87
<b>8. Conclusion et problématique générale</b> .....	<b>89</b>
 <i>Chapitre III - Méthodologie de formalisation de la Mise en Forme Matérielle .. 90</i>	
<b>1. Introduction</b> .....	<b>92</b>
<b>2. Socle théorique pour la formalisation de la MFM</b> .....	<b>93</b>
2.1 Les différents niveaux de structuration des textes écrits.....	93
2.2 Analyse automatique du contenu des documents.....	95
2.3 Des bases théoriques du MAT vers une formalisation de la MFM.....	96
<b>3. Segmentation des objets textuels des documents</b> .....	<b>98</b>
3.1 Segmentation visuelle : une première approche .....	99
3.2 Le cas d'objets textuels imbriqués .....	101



3.3 Segmentation d'objets textuels sur la base d'indices lexico-syntaxiques .....	102
3.4 Exemple de segmentation .....	104
3.5 Conclusion pour nos travaux .....	105
<b>4. Définition et identification des conventions partagées de MFM .....</b>	<b>105</b>
4.1 De l'existence de conventions de MFM partagées.....	105
4.2 Modélisation des conventions de MFM partagées universellement .....	107
4.2.a Classes d'objets textuels .....	107
4.2.b Relations conventionnées entre objets textuels.....	112
4.3 Exemple de formalisation des architectures textuelles d'un document sur la base des conventions partagées de MFM .....	114
4.4 Conclusion .....	115
<b>5. Modélisation des conventions locales de MFM .....</b>	<b>115</b>
5.1 Méthode d'identification des conventions locales .....	116
5.2 Exemple d'annotation des conventions locales .....	118
5.3 Limites de l'approche .....	118
<b>6. Vue d'ensemble du processus d'annotation et opérationnalisation technique .....</b>	<b>119</b>
6.1 Schéma global .....	119
6.2 Implémentation informatique du processus de formalisation.....	120
<b>7. Vers une formalisation automatique de la MFM par apprentissage automatique.....</b>	<b>123</b>
<b>8. Conclusion.....</b>	<b>126</b>
<i>Chapitre IV - Restituer la Mise en Forme Matérielle à l'oral : conception et opérationnalisation.....</i>	<i>128</i>
<b>1. Introduction .....</b>	<b>130</b>
<b>2. Typologie des informations de MFM à restituer.....</b>	<b>131</b>
<b>3. Vers une restitution optimisée de la MFM : étude globale de l'espace de conception .....</b>	<b>133</b>
3.1 Que restituer de la MFM : filtrage des informations.....	134
3.2 Comment restituer les informations de MFM : espace de conception.....	137
3.2.a Reformulation discursive .....	138
3.2.b Modulation de la ou des voix de synthèse.....	140
3.2.c Modulation du son : icônes audio et son spatialisé .....	143
3.2.d Modalités tactiles.....	145
3.2.e Restitution à travers la définition de commandes utilisateur.....	147
3.2.f Comment combiner les modalités de restitution .....	147
3.3 Quand restituer les informations de MFM ? .....	148
3.4 Restitution de la dimension fonctionnelle de la MFM .....	150
<b>4. Proposition d'une stratégie de restitution de la MFM .....</b>	<b>151</b>
4.1 Restitution des informations locales de MFM.....	152
4.2 Restitution des informations transverses de MFM .....	152
4.3 Commentaires sur la stratégie de restitution.....	153
<b>5. READ : une plateforme expérimentale pour implémenter et évaluer des stratégies d'accès non-visuelles aux documents .....</b>	<b>155</b>
5.1 Introduction : pourquoi cette plateforme ? .....	155
5.2 Présentation générale de READ .....	156
5.2.a Fonctionnement et objectifs.....	156

5.2.b Données techniques sur le système et son fonctionnement .....	161
5.3 Formats et modalités utilisées pour restituer la MFM .....	162
5.4 Perspectives et limites .....	164
<b>6. Conclusion .....</b>	<b>164</b>
<i>Chapitre V - Évaluations de la restitution orale de la mise en forme matérielle</i>	
.....	167
<b>1. Introduction .....</b>	<b>168</b>
<b>2. Concepts clés .....</b>	<b>170</b>
<b>3. Évaluation de la mise en saillance des titres à l'oral auprès de sujets non-voyants .....</b>	<b>172</b>
3.1 Hypothèse éprouvée .....	172
3.2 Méthode .....	172
3.2.a Méthodes de restitution des titres .....	172
3.2.b Participants .....	173
3.2.c Matériel, tâches et consignes .....	173
3.3 Résultats .....	174
3.3.a Effet de la stratégie de restitution sur la rétention du plan .....	176
3.3.a.1 Identification et rétention des thèmes du plan .....	176
3.3.a.2 Ordre des thèmes dans le plan rappelé .....	177
3.3.a.3 Distance entre le plan rappelé et le plan d'origine .....	178
3.3.b Effet de la stratégie de restitution sur les scores de compréhension du contenu .....	179
3.3.c Effet de la stratégie de restitution sur les scores de charge cognitive globale .....	180
3.4 Discussion .....	181
<b>4. Évaluation de 4 stratégies de restitution de la MFM auprès de sujets non-déficients .....</b>	<b>182</b>
4.1 Présentation des hypothèses .....	183
4.1.a Hypothèses : contenu et méta-informations, le « quoi » .....	183
4.1.b Hypothèses : stratégies de restitution de la MFM, le « comment » .....	184
4.1.c Hypothèse : stratégies de restitution de la mise en forme, le « quand » (non évaluée) .....	185
4.2 Méthode .....	185
4.2.a Participants .....	185
4.2.b Matériel .....	185
4.2.c Stratégies de restitution évaluées .....	187
4.2.c.1 Condition « sans mise en forme » (N) .....	188
4.2.c.2 Condition « lecteur d'écran » (J) .....	188
4.2.c.3 Condition « optimale » (O) .....	189
4.2.c.4 Condition exhaustive (X) .....	191
4.2.d Procédure de test .....	191
4.3 Résultats attendus .....	193
4.4 Résultats .....	194
4.4.a Caractéristiques de la population .....	195
4.4.a.1 Usage des technologies .....	195
4.4.a.2 Connaissances antérieures .....	196
4.4.b Effet de la stratégie de restitution sur la rétention du plan du texte .....	197
4.4.b.1 Identification et rétention des thèmes du plan .....	197
4.4.b.2 Ordre des thèmes dans le plan rappelé .....	198
4.4.b.3 Distance entre le plan rappelé et le plan d'origine .....	199
4.4.b.4 Similitude entre la structure du plan restitué et la structure du plan d'origine .....	199

4.4.c	Effet de la stratégie de restitution sur les scores de compréhension du texte.....	201
4.4.d	Effet de la stratégie de restitution sur la charge cognitive .....	202
4.4.d.1	Charge cognitive globale .....	202
4.4.d.2	Charge intrinsèque liée au contenu.....	203
4.4.d.3	Charge cognitive liée à l'intelligibilité de la synthèse vocale.....	204
4.4.d.4	Charge extrinsèque liée à la présentation du contenu.....	205
4.4.d.5	Charge cognitive liée à l'apprentissage .....	206
4.4.e	Effet de la stratégie de restitution sur l'efficacité.....	207
4.4.f	Effet de la stratégie de restitution sur la désorientation.....	209
4.4.g	Effet de la stratégie de restitution sur les tâches de recherche d'information .....	210
4.4.g.1	Temps moyens mis pour trouver les emplacements cibles.....	210
4.4.g.2	Réussites de la recherche d'information .....	211
4.4.h	Effet de la stratégie de restitution sur l'utilisabilité subjective .....	212
4.4.i	Effet de la stratégie de restitution sur l'utilisation des commandes .....	213
4.4.i.1	Commandes de navigation .....	214
4.4.i.2	Commandes spécifiques au mode optimal.....	216
4.4.j	Effet de la stratégie de restitution sur les temps de parcours du document .....	216
4.4.k	Retours qualitatifs .....	217
4.4.l	Synthèse des résultats .....	218
4.5	Discussion .....	219
<b>5.</b>	<b>Évaluation de 2 stratégies de restitution de la MFM auprès sujets non-voyants .....</b>	<b>224</b>
5.1	Présentation .....	224
5.2	Méthode.....	225
5.2.a	Participants .....	225
5.2.b	Matériel.....	225
5.2.c	Stratégies de restitution évaluées.....	226
5.2.d	Procédure de test.....	226
5.3	Résultats .....	226
5.3.a	Connaissances antérieures des sujets.....	227
5.3.b	Effet de la stratégie de restitution sur la rétention du plan du texte .....	228
5.3.b.1	Identification et rétention des thèmes du plan.....	229
5.3.b.2	Ordre des thèmes dans le plan rappelé.....	230
5.3.b.3	Distance entre le plan rappelé et le plan d'origine .....	230
5.3.b.1	Similitude entre la structure du plan restitué et la structure du plan d'origine.....	231
5.3.c	Effet de la stratégie de restitution sur les scores de compréhension du texte.....	232
5.3.d	Effet de la stratégie de restitution sur la charge cognitive .....	233
5.3.d.1	Charge cognitive globale .....	233
5.3.d.2	Charge intrinsèque liée au contenu.....	233
5.3.d.3	Charge intrinsèque liée à l'intelligibilité de la synthèse vocale.....	234
5.3.d.4	Charge extrinsèque liée à la présentation du contenu.....	235
5.3.d.1	Charge cognitive liée à l'apprentissage .....	236
5.3.e	Effet de la stratégie de restitution sur l'efficacité.....	237
5.3.f	Effet de la stratégie de restitution sur la désorientation.....	238
5.3.g	Effet de la stratégie de restitution sur les tâches de recherche d'information .....	239
5.3.g.1	Temps moyens mis pour trouver les emplacements cibles.....	239
5.3.g.2	Réussites de la recherche d'informations.....	240
5.3.h	Effet de la stratégie de restitution sur l'utilisabilité subjective .....	241
5.3.i	Effet de la stratégie de restitution sur l'utilisation des commandes .....	241
5.3.i.1	Commandes de navigation .....	241

5.3.i.2 Commandes spécifiques au mode optimal.....	243
5.3.j Effet de la stratégie de restitution sur les temps de parcours des documents .....	244
5.3.k Retours qualitatifs .....	245
5.3.l Synthèse des résultats .....	246
5.4 Discussion .....	247
<b>6. Conclusion sur l'ensemble des résultats.....</b>	<b>250</b>
<i>Conclusion générale et perspectives.....</i>	<i>256</i>
<b>7. Synthèse des contributions .....</b>	<b>256</b>
7.1 Contributions théoriques .....	256
7.2 Implications et contributions pratiques .....	258
7.3 Limites des travaux.....	260
<b>8. Perspectives : futurs travaux .....</b>	<b>261</b>
<i>Références.....</i>	<i>264</i>
<i>Annexes.....</i>	<i>279</i>
<b>1. Exemple de document dont la MFM a été formalisée .....</b>	<b>279</b>
<b>2. Étude en situation .....</b>	<b>282</b>
2.1 Questionnaires ayant guidé les entretiens de l'étude en situation .....	282
2.1.a Questionnaire destiné aux éducateurs spécialisés .....	283
2.1.b Questionnaire destiné aux élèves déficients visuels.....	284
2.1.c Questionnaire destiné aux professeurs des établissements publics.....	285
2.1.d Questionnaire destiné aux transcrip-teurs braille professionnels de l'institut .....	286
2.2 Grille d'observation utilisée pour l'étude en situation.....	287
2.3 Extraits des entretiens réalisés pour l'étude en situation .....	287
<b>3. Extraits de documents formalisés automatiquement par apprentissage automatique.....</b>	<b>289</b>
<b>4. Expérimentations sur la restitution de la MFM .....</b>	<b>293</b>
4.1 Retours qualitatifs des sujets voyants.....	293
4.2 Retours qualitatifs des sujets déficients visuels .....	296
4.3 Fonctions R utilisées pour les tests statistiques .....	298
4.4 Documents utilisés pour les expérimentations 2 et 3 .....	299
<i>Table des matières du manuscrit.....</i>	<i>310</i>
<i>Glossaire .....</i>	<i>316</i>
<i>Abréviations .....</i>	<i>317</i>
<i>Table des figures .....</i>	<i>317</i>
<i>Liste des tableaux.....</i>	<i>320</i>
<i>Liste des publications .....</i>	<i>321</i>
<i>Résumé de la thèse en Anglais – Abstract .....</i>	<i>322</i>

# Glossaire

- § Architectures textuelles : objets textuels et relations entre objets textuels rendus perceptibles par des marques de mise en forme.
- § Auteuriale : de l'auteur (d'un texte en l'occurrence). Par exemple les intentions *auteuriales* sont les intentions de l'auteur.
- § Bloc-notes braille : ordinateur portable dédié à des utilisateurs déficients ; comprend habituellement un terminal braille, une plage braille, une sortie vocale, et des touches fonctionnelles.
- § Boucle phonologique : composante de la mémoire de travail dédiée au traitement de l'information de nature verbale.
- § Calepin visuo-spatial : composante de la mémoire de travail dédiée au traitement de l'information de nature visuelle et spatiale.
- § Charge cognitive : fait référence aux ressources mentales engagées en mémoire de travail pour une activité cognitive donnée.
- § Contenu primaire : message que l'auteur d'un texte souhaite communiquer aux lecteurs ; représente un ensemble d'idées et de concepts pouvant faire référence aux objets et événements du monde.
- § Désorientation : dans nos travaux, fait référence au sens de l'orientation dans un document accédé dans des modes non-visuels : savoir où l'on est, d'où l'on vient, et où l'on va.
- § Discursif : qui se rapporte au discours (insiste sur l'aspect énonciatif).
- § Expositif : texte ayant pour but d'exposer, de faire comprendre de manière objective des idées ou des concepts.
- § Haptique : (du grec *haptomai* qui signifie *je touche*), désigne tout ce qui est relatif au sens du toucher c'est-à-dire au sens tactile (cutané), et au sens kinesthésique qui désigne ce que l'on ressent par les muscles, les tendons et les articulations.
- § Lecteurs d'écran : systèmes logiciel permettant d'utiliser un ordinateur ou un dispositif mobile sans la vue. Ces systèmes font l'interface entre le système d'exploitation et les applications d'un côté, et les périphériques et médias utilisables par les déficients visuels, principalement les synthèses vocale et les plages braille.
- § Lexico-syntaxique : fait référence à la fois au lexique et à la syntaxe (du texte).
- § Marques de réalisation : marques de mise en forme appliquée au texte exprimant les intentions de l'auteur décrites dans le métatexte ; représente donc la réalisation du métatexte.
- § Métaphore : phrases relevant du métatexte.
- § Métatexte : texte faisant référence au texte lui-même.
- § Mise en forme matérielle : ensemble des propriétés signifiantes de réalisation du métatexte dans le texte. La mise en forme matérielle constitue la réalisation du métatexte par des marques de mise en forme.
- § Modalités : structures spécifiques d'une information pour un mode donné. Par exemple pour le mode audio, on peut citer plusieurs modalités : parole, icônes audio, extrait musical, etc.
- § Modèle d'Architectures Textuelles : modèle logico-linguistique de la mise en forme des textes. Postule que la mise en forme signifiante des textes est issue de la réalisation visuelle de métatexte portant sur le texte lui-même et son organisation.

- § Mode : Un mode fait ici référence à un mode perceptif de l'information chez l'humain ; un mode correspond donc à un sens perceptif.
- § Objets textuels : segments de texte rendus perceptibles, discernables du reste du texte, par un jeu de contraste.
- § Tactile : fait référence au sens du toucher, à la composante cutanée du sens haptique.
- § Terminal braille : aussi appelée plage braille, désigne une matrice de picots piézoélectriques (i.e. déformables sous l'action d'un champ électrique) permettant de former des caractères braille en plaçant chaque picot en position haute ou basse par rapport à une grille. Les terminaux braille permettent de représenter une ligne de texte en braille allant jusqu'à 80 caractères.
- § Typo-dispositionnel : fait référence à la fois à la typographie et à la disposition (du texte).

## Abréviations

- § ARIA : *Accessible Rich Internet Applications*. Normes d'accessibilité développées par le W3C.
- § DAISY : *Digital Accessible Information SYstem*. Standard pour la création et la publication de livres audio.
- § HTML : *Hypertext Markup Language*. Sous-langage du XML utilisé pour représenter des pages Web.
- § JAWS : *Job Access With Speech*. Lecteur d'écran populaire auprès des utilisateurs déficients visuels (voir glossaire à *lecteurs d'écran*).
- § MAT : *Modèle d'Architectures Textuelles* (voir glossaire).
- § MDT : *Mémoire de Travail*. Module de la mémoire gérant le traitement cognitif des informations perçues, voir Chapitre II3.2 pour plus de détails.
- § MFM : *Mise en Forme Matérielle* (voir glossaire).
- § SE : *Système d'Exploitation*. Système logiciel permettant à un utilisateur d'exploiter le matériel informatique sur lequel il est installé, à travers plusieurs applications.
- § USB : *Universal Serial Bus*. Standard pour la connexion physique des périphériques informatiques.
- § UT : *Unité Textuelle* (voir glossaire).
- § W3C : *World Wide Web Consortitum*. Groupement d'experts œuvrant à la standardisation du web pour assurer son développement et son évolution sur le long terme.
- § WCAG : *Web Content Accessibility Guidelines*. Normes d'accessibilité développées par le W3C.
- § XML : *eXtended Markup Language*. Langage de représentation de document.

## Table des figures

Figure 1 – Vue d'ensemble des concepts du MAT (adapté de [Lemarié et al. 2008]) .....	10
Figure 2 - Un prototexte et deux exemples de formes dérivées mises en forme (adapté de [Lemarié et al. 2008]).....	12
Figure 3 - Exemples de phénomènes de mise en forme équivalents.....	12

Figure 4 – Schéma des concepts importants du MAT : seconde version.....	14
Figure 5 - Schéma syntaxique général des métaphrases .....	15
Figure 6 - Exemple de modélisation de la MFM sous forme de métaphrases.....	17
Figure 7 - Premier exemple d'architecture textuelle modélisable par le MAT .....	23
Figure 8 - Deuxième exemple d'architecture textuelle modélisable par le MAT.....	24
Figure 9 - Troisième exemple d'architecture textuelle modélisable par le MAT .....	24
Figure 10 – Alphabet braille (gauche) et figure en relief sur papier thermogonflable (droite).....	36
Figure 11 - Schéma de la mémoire de travail adapté de [Baddeley 2003] .....	38
Figure 12 - Terminal braille ( <i>image libre de droits</i> ).....	58
Figure 13 - Le PHANTOM Omni de Sensable.....	60
Figure 14 - Clavier QWERTY (gauche) ; clavier braille et caractère braille (droite) ( <i>images libres de droits</i> ) .....	61
Figure 15 – Bloc-notes braille « Apex BT 32 » de <i>HumanWare</i> .....	69
Figure 16 - Lecteur de livres audio DAISY physique (à gauche), et logiciel avec le lecteur AMIS (à droite).....	70
Figure 17- Exemple de relation de composition (gauche) et de relation de labellisation (droite) .....	74
Figure 18 – Schéma résumant les sources informationnelles à disposition des élèves déficients visuels en classe .....	82
Figure 19 –Exemples d'objets textuels de manuels de mathématiques non retranscrits.....	83
Figure 20 –Exemple de disposition spatiale de texte non retranscrite.....	83
Figure 21 – Illustration de la loi de clôture du regroupement visuel (Gestalt) .....	100
Figure 22 – Illustration de la loi de familiarité du regroupement visuel (Gestalt).....	101
Figure 23 – Extrait d'un manuel de mathématiques : illustration de segments visuels imbriqués ....	101
Figure 24 – Exemple de segmentation des unités visuelles d'une page d'un manuel de Mathématiques .....	104
Figure 25 – Manuscrit Arabe des mille-et-une nuits datant du XIV <sup>ème</sup> siècle ( <i>image libre de droits</i> ). 106	
Figure 26 – Pyramide des couches de conventions de MFM.....	107
Figure 27 - Hiérarchie d'héritage des classes de conventions partagées de MFM .....	110
Figure 28 - Exemple d'annotation des statuts de MFM d'après les conventions partagées .....	114
Figure 29 – Exemple d'annotation de la relation <i>intituler</i> .....	115
Figure 30 – Exemple d'annotation des conventions locales .....	118
Figure 31 - Schéma global du processus d'annotation de la MFM d'un document.....	119
Figure 32 - Extrait du code source d'un document du corpus GEOP .....	124
Figure 33 – Les différentes formes possible pour la reformulation des métaphrases .....	138
Figure 34 - Illustration de la restitution de la fonction d'anticipation à l'aide de son spatialisé .....	150
Figure 35 - Capture d'écran d'un site web journalistique .....	154
Figure 36 - Capture d'écran illustrative du logiciel READ .....	155
Figure 37 - Schéma du fonctionnement global de READ .....	158
Figure 38 - Extrait d'un fichier de configuration .....	159
Figure 39 - Illustration du résultat de la configuration de la lecture d'un objet textuel.....	160
Figure 40 - Rétention des thèmes du plan en fonction de la stratégie et du document : expérimentation sur la restitution des titres .....	177
Figure 41 - Ordre de restitution des titres en fonction de la stratégie et du document : expérimentation sur la restitution des titres .....	178

Figure 42 - Distance au plan en fonction de la stratégie et du document : expérimentation sur la restitution des titres.....	179
Figure 43 - Scores de compréhension en fonction de la condition : expérimentation sur la restitution des titres.....	180
Figure 44 - Charge cognitive globale en fonction de la stratégie et du document : expérimentation sur la restitution des titres.....	181
Figure 45 - Distributions des réponses aux questions qualifiant l'usage des technologies des sujets	195
Figure 46 - Distributions des réponses aux questions de connaissances.....	196
Figure 47 - Scores de rétention des thèmes du plan.....	197
Figure 48 - Ordre des thèmes restitués relativement à l'ordre d'origine.....	198
Figure 49 –Distributions des distances au plan en fonction de la condition expérimentale.....	199
Figure 50 - Structure hiérarchique restituée par les sujets.....	200
Figure 51 – Scores de compréhension ; (gauche) droite de Henry vs Quantiles ; (droite) moyennes et intervalles de confiance.....	201
Figure 52 - Scores des réponses à la question 1 sur la charge mentale globale.....	202
Figure 53 - Charge cognitive liée au contenu.....	204
Figure 54 - Charge cognitive induite par la compréhension de la voix de synthèse.....	205
Figure 55 - Charge cognitive liée à la présentation du contenu.....	206
Figure 56 - Concentration des sujets pendant la tâche.....	207
Figure 57 - Scores d'efficacités calculés sur la base des scores de compréhension.....	208
Figure 58 - Scores de désorientation.....	209
Figure 59 – Temps moyens cumulés des deux tâches de recherche d'information.....	211
Figure 60 – Réussites aux tâches de recherche d'information.....	212
Figure 61 - Scores d'utilisabilité du système en fonction de la condition.....	213
Figure 62 – Report quantitatif de l'utilisation des commandes de navigation.....	214
Figure 63 - Utilisation des commandes de navigation avant (à gauche) et après (à droite) la fin du document.....	215
Figure 64 - Report quantitatif des commandes du mode optimal utilisées.....	216
Figure 65 – Pourcentage du temps total imparti utilisé pour le parcours du document.....	217
Figure 66 - Connaissances antérieures des sujets déficients-visuels.....	228
Figure 67 - Rétention des thèmes du plan chez les sujets déficients visuels; (gauche) quantiles et droit de Henry ; (droite) moyennes et intervalles de confiance.....	229
Figure 68 - Ordre des thèmes restitués (sujets déficients-visuels).....	230
Figure 69 - Distance au plan (sujets déficients-visuels).....	231
Figure 70 - Structure restituée (sujets déficients visuels).....	231
Figure 71 – Scores de compréhension (sujets déficients-visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance des scores.....	232
Figure 72 - Charge cognitive globale (sujets déficients visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance.....	233
Figure 73 - Charge cognitive liée au contenu (sujets déficients visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance.....	234
Figure 74 - Charge intrinsèque liée à la modalité orale (sujets déficients visuels).....	235
Figure 75 - Charge extrinsèque (sujets déficients visuels).....	236
Figure 76 - Effet de l'ordre de présentation sur la charge cognitive liée à l'apprentissage.....	237



Figure 77 – Efficience (sujets déficients visuels); (gauche) quantiles et droit de Henry ; (droite) moyennes et intervalles de confiance.....	238
Figure 78 - Scores de désorientation (sujets déficients-visuels) .....	239
Figure 79 - Durées cumulées des tâches de navigation (sujets déficients visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance .....	240
Figure 80 - Échecs lors des tâches de navigation (sujets déficients-visuels).....	240
Figure 81 - Utilisabilité du système en fonction de la stratégie et du document abordé (sujets déficients-visuels).....	241
Figure 82 – Commandes de navigation utilisées (sujets déficients visuels) ; (gauche) quantiles vs droite de Henry ; (droite) moyennes et intervalles de confiance .....	242
Figure 83 - Commandes de navigation utilisées avant (gauche) et après (droite) la fin du document (sujets déficients visuels).....	243
Figure 84 - Commandes du mode optimal utilisées (sujets déficients-visuels) .....	244
Figure 85 - Temps total de parcours des documents en fonction de la stratégie de restitution et de l'ordre de présentation (sujets déficients-visuels).....	245

## Liste des tableaux

Tableau 1 – Extrait des commandes listées dans le manuel de JAWS .....	63
Tableau 2- Études sur l'utilisabilité des lecteurs d'écran .....	67
Tableau 3 - Exemples d'architectures textuelles non-restitués par les technologies de lecture.....	77
Tableau 4 – Description des types de données recueillies pour l'étude en situation.....	80
Tableau 5 – Résumé des problèmes de disponibilité et d'accessibilité de l'information pédagogique.....	87
Tableau 6- Exemples de métaphrases.....	97
Tableau 7 – Illustration de segmentation visuelle basée sur l'arrière-plan .....	100
Tableau 8 - Exemples de réalisations d'un métatexte par différentes marques de MFM.....	103
Tableau 9 - Classes d'objets relatives aux conventions partagées de MFM .....	109
Tableau 10 - Métaphrases dérivées des classes d'objets de MFM .....	112
Tableau 11 - Relations entre segments de MFM d'après les conventions partagées.....	114
Tableau 12 – Marques de MFM et autres indices utilisés pour la labellisation des objets textuels (repris de [Fauconnier et al. 2014]).....	125
Tableau 13 - Performances pour la labellisation des objets textuels et la reconnaissance des relations de composition (repris de [Fauconnier et al. 2014]).....	125
Tableau 14 – Avantages et inconvénients de la reformulation discursive.....	140
Tableau 15 – Avantages et inconvénients des différentes modulations des voix de synthèse pour la restitution.....	143
Tableau 16 - Avantages et inconvénients des icônes audio et du son spatialisé pour la restitution .	145
Tableau 17 - Avantages et inconvénients de modalités tactiles pour la restitution.....	147
Tableau 18 - Avantages et inconvénients des commandes utilisateur pour la restitution.....	147
Tableau 19 – Déroulement de l'expérimentation sur les titres .....	174
Tableau 20- Résultats du test du $\chi^2$ : questions sur l'usage des technologies .....	196
Tableau 21 - Résultats du test du $\chi^2$ : questions de connaissances .....	197
Tableau 22 – Résultats du test Post-Hoc de Kruskal-Wallis : thèmes restitués .....	198

Tableau 23 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée .....	199
Tableau 24 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée .....	201
Tableau 25 – Résultats du test Tukey-HSD (ANOVA post-hoc) : scores de compréhension .....	202
Tableau 26 - Résultats du test Post-Hoc de Kruskal-Wallis : charge cognitive globale .....	203
Tableau 27 - Résultats du test Post-Hoc de Kruskal-Wallis : charge cognitive liée à la présentation du contenu .....	206
Tableau 28 - Résultats du test Post-Hoc de Kruskal-Wallis : scores d'efficacité basées sur la compréhension .....	208
Tableau 29 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée .....	210
Tableau 30 - Résultats du test Post-Hoc de Kruskal-Wallis : structure hiérarchique restituée .....	212
Tableau 31 - Synthèse des remarques qualitatives des sujets voyants .....	218
Tableau 32 - Synthèse des résultats : expérimentation sur sujets voyants .....	219
Tableau 33 – Contrebalancement des conditions et documents : expérimentation avec sujets déficients-visuels .....	225
Tableau 34- Résultats du test du $\chi^2$ : connaissances antérieures des sujets déficients-visuels, effet du document .....	228
Tableau 35 - Synthèse des résultats : expérimentation sur sujets déficients-visuels .....	247

## Liste des publications

### ► Journaux internationaux

[2015] **Sorin, L.**, Lemarié, Dupuis, A., J., Aussenac, N., & Mojahid, M. How educational information access can impact learning in class for blind students. **Submitted June 2015** to *Universal Access in the Information Society*. Springer Berlin Heidelberg.

### ► Actes de conférences internationales

[2015] **Sorin, L.**, Lemarié, J., Mojahid, M., Aussenac-gilles, N. READ : a (research) platform for evaluating non- visual access methods to digital documents. **To be published** In *ASSETS 2015 – Posters and Demonstrations*. ACM Press.

[2014] **Sorin, L.**, Lemarié, J., Aussenac-gilles, N., Mojahid, M., & Oriola, B. Communicating text structure to blind people with Text-To-Speech. In *ICCHP 2014 - Computers Helping People with Special Needs*. Springer-Verlag Berlin Heidelberg.

[2013] **Sorin, L.**, Mojahid, M., Aussenac-gilles, N., & Lemarié, J. Improving the accessibility of digital documents for blind users : contributions of the Textual Architecture Model. In C. Stephanidis & M. Antona (Eds.), *Universal Access in Human-Computer Interaction UAHCI 2013, Held as Part of HCI International 2013* (pp. 423–431). Las Vegas, Nevada: Springer-Verlag Berlin Heidelberg.

### ► Actes de conférences nationales

[2014] Fauconnier, J., **Sorin, L.**, Kamel, M., Aussenac-gilles, N., & Mojahid, M. Détection automatique de la structure organisationnelle de documents à partir de marqueurs visuels et lexicaux. In *TALN -RECITAL 2014, Traitement Automatique du Langage Naturel*.

# Résumé de la thèse en Anglais – *Abstract*

Digital accessibility plays a crucial role for the education, the social inclusion and the autonomy of impaired people. This work focused on a universal component of digital documents: text formatting. Our aim was to provide a functional access to text formatting semantics. Far from being a simple text ornament, text formatting underlying semantics allows authors to build actual text architectures.

We conducted an exploratory survey with visually impaired students working in inclusive settings, which aimed at assessing documents accessibility issues and needs in this context. The results showed visually impaired students encounter issues when accessing educational textual content, which can lead to situations of cognitive overload. Those issues relate to the inaccessibility of textual architectures which allow readers to build a mental representation of documents global organization. Besides, the literature shows that assistive technologies used by visually impaired people to access digital contents ignore most of documents textual architectures. Yet numerous works show the importance of text formatting and its semantics for content comprehension, in the case of visual reading.

Thus, we aimed at finding ways of rendering digital documents textual architectures in non-visual ways, and also to evaluate to what extent this could be beneficial to visually impaired users. Our work comprised three main steps: producing a formal methodology to extract text formatting semantics, designing multimodal restitution strategies to allow access to textual architectures conveyed by text formatting semantics, and finally evaluating the impact of those restitution strategies on content comprehension. We adopted a holistic approach for the design of restitution strategies, taking into account the characteristics of text formatting, of the comprehension task and the cognitive characteristics of visually impaired users.

The developed methodology for extracting text formatting semantics is based on text typography, disposition, lexical and syntactic features in texts. It relies on the theoretical framework defined by the Textual Architectures Model [Virbel 1985; Virbel 1989], and is able to model and formalize any textual architecture. We produced a first implementation of the methodology using machine learning techniques, which showed promising results for the automatic modeling of simple textual architectures.

We designed and evaluated several restitution strategies of textual architectures with 60 sighted and 12 blind volunteers; our strategies used several audio modalities to convey information. Those two experimentations showed that textual architectures restitution can improve content comprehension and lower cognitive load; besides, using non-verbal modalities can help users' orientation for non-visual documents access. Overall, those results show that digital documents accessibility can still be improved for visually impaired users, in terms of performance, satisfaction and efficiency (regarding the cognitive demand). To conclude, the study of textual architectures conveyed by text formatting, and the design of functional equivalents, constitute a new relevant approach to improve digital documents accessibility.