

UNIVERSITE TOULOUSE III – PAUL SABATIER
U.F.R. Math Informatique Gestion (M.I.G.)

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE TOULOUSE III
DISCIPLINE : INFORMATIQUE

Par

Fabrice Maurel
le 6 décembre 2004

Titre :

Transmodalité et multimodalité écrit/oral :
modélisation, traitement automatique
et évaluation de stratégies de présentation
des structures « visuo-architecturales » des textes

Directeurs de thèse : Nadine Vigouroux et Jean-Luc Nespoulous

JURY

Jean Caelen (Rapporteur)	Directeur de Recherche au CNRS-CLIPS
Denis Maurel (Rapporteur)	Professeur à l'Université de Tours
Régine André-Obrecht (Examinatrice)	Professeur à l'Université Toulouse III
Jacques Virbel (Examineur)	Chargé de Recherche au CNRS-IRIT
Nadine Vigouroux (Directrice)	Chargée de Recherche au CNRS-IRIT
Jean-Luc Nespoulous (Directeur)	Professeur à l'Université Toulouse II

Institut de Recherche en Informatique de Toulouse (IRIT)
Université Toulouse III – Paul Sabatier – 118, route de Narbone 31062 Toulouse cedex

*Amoureusement, à Isabelle avant tout,
tendrement, à Julien et Samuel bien sûr,
toué y mank a moin bokou, Marion, zoli ti mamzel zoreol,
à notre chère Ima...*

Remerciements

J'adresse mes plus sincères remerciements à ma famille et à mes amis pour leur soutien permanent et sans faille, ainsi qu'à :

- *Nadine Vigouroux pour le passionnant sujet qu'elle m'a permis d'explorer et pour la qualité du suivi de mon travail, à la fois assez rapproché pour suggérer et assez distant pour ne pas imposer ;*
- *Jean-Luc Nespoulous pour la confiance qu'il a toujours bien voulu m'accorder et la constante clairvoyance de ses conseils tout le long de ce travail ;*
- *Régine André-Obrecht pour avoir accepté de présider le jury de ma thèse ;*
- *tout particulièrement, Jean Caelen et Denis Maurel qui m'ont fait l'honneur d'être rapporteurs de mon travail.*

Je tiens à exprimer à Jacques Virbel toute ma reconnaissance pour la qualité scientifique de ses travaux et l'enthousiasme qu'il a su me communiquer pendant ces années, qui resteront pour moi une richesse « architecturante ».

Je tiens spécialement à remercier Mustapha Mojahid pour ses relectures avisées et ses précieux conseils.

J'adresse également mes remerciements les plus amicaux à tous les membres de l'équipe DIAMANT ainsi qu'à ceux du projet « Serveurs multimodaux destinés aux personnes âgées : méthode de conception et étude de leur utilisabilité », en particulier à Jean-Léon Bouraoui pour le rôle ingrat de correcteur orthographique qu'il a eu la gentillesse de jouer, ainsi qu'à Farida Aouladomar et à Julie Lemarié dont la collaboration au cours de leur stage a été une chance à la fois scientifique et humaine.

De tout cœur merci à mes parents, qui m'ont appris mes premiers mots puis fait aimer ceux des autres.

Résumé

Actuellement, une valeur de plus en plus grande est donnée à la structure informationnelle des documents qui est prise en charge par la morpho-disposition des textes. Longtemps considérée comme ornementale, cette spécificité du texte écrit est abordée comme un objet d'étude à part entière dans diverses disciplines telles que la linguistique, la psycholinguistique, la psychologie cognitive ou l'informatique. Une telle recherche, à caractère fortement pluridisciplinaire, s'avère nécessaire pour que cette composante soit intégrée dans les applications en traitement automatique des langues. En particulier, nous nous intéressons à l'utilité de la structure visuelle des documents et, le cas échéant, à son utilisabilité, dans le cadre de la transposition à l'oral par des systèmes de synthèse de parole à partir de textes (Text-To-Speech ou TTS). Dans l'objectif de fournir des solutions qui permettent de réagir efficacement à cette « inscription morphologique » des textes, nous proposons la synoptique d'un système d'oralisation de documents « tout-venants » qui conduit à une représentation du texte interprétable par un TTS. Nous avons modélisé et partiellement réalisé le module spécifique aux stratégies d'oralisation, afin de rendre « articulables » certaines parties signifiantes des textes, souvent « oubliées » par les systèmes actuels de synthèse de parole. Des spécifications formelles ont été précisées et partiellement intégrées par un partenaire industriel. Des hypothèses prédictives quant à l'impact sur la mémorisation/compréhension de deux stratégies (discursive et prosodique), issues de notre **Modèle d'Oralisation par Reformulation des Textes Écrits** pour être **Lus Silencieusement (MORTELS)**, ont été formulées et testées. Ce travail a permis de montrer que certaines fonctions cognitives de la morpho-disposition des textes semblent perdues. Des prototypes, exploitant la notion d'Image De Page (IDP), ont été imaginés à travers des interfaces dans lesquelles la multimodalité a pour rôle de combler cette lacune.

Sommaire

PROBLÉMATIQUE ET MISE EN PERSPECTIVE.....	11
TRANSMODALITE ET MULTIMODALITE.....	13
LA LECTURE SILENCIEUSE.....	21
RELATIONS ENTRE L'ECRIT ET L'ORAL.....	23
PARTIE A VALEUR ET PRISE EN COMPTE DE LA MORPHO-DISPOSITION	29
INTRODUCTION	31
CHAPITRE I. SYNTHESE DE PAROLE ET TYPOGRAPHIE	33
CHAPITRE II. DU POINT DE VUE DES PSYCHOLINGUISTES.....	39
CHAPITRE III. DU POINT DE VUE DES LINGUISTES	53
CONCLUSION	87
PARTIE B ARCHITECTURE, MODELE ET OPERATIONNALISATION	93
CHAPITRE IV. UNE ARCHITECTURE ADAPTEE A L'ORALISATION DES TEXTES PAR « OT »	95
CHAPITRE V. MODELE DES STRATEGIES D'ORALISATION	101
CHAPITRE VI. OPERATIONNALISATION DES STRATEGIES	105
PARTIE C ÉVALUATION COGNITIVE	149
CHAPITRE VII. HYPOTHESES ET VARIABLES	151
CHAPITRE VIII. <i>DESIGN</i> FINAL	161
CHAPITRE IX. ANALYSE DES PREMIERS RESULTATS	167
PARTIE D IMAGES DE PAGE ET MULTIMODALITE	177
CHAPITRE X. <i>FIRST GLANCE</i> ET IMAGES DE PAGE	179
CHAPITRE XI. PERSPECTIVES A COURT TERME.....	193
POINT SUR NOTRE CONTRIBUTION ET PERSPECTIVES A MOYEN TERME	201
(1) IHM : VERS UNE PRESENTATION ADAPTABLE.....	205
(2) VERS UN SYSTEME DE « TRANSMODALISATION » DE L'ARCHITECTURE TEXTUELLE	206
(3) COGNITION.....	207
TABLE DES FIGURES	229
TABLE DES TABLEAUX.....	229
TABLE DES MATIERES.....	231
ANNEXE	235
1. REPRESENTATION DES STRUCTURES ENUMERATIVES	237
2. STRATEGIES D'ORALISATION DES STRUCTURES ENUMERATIVES	239
TITRE ET RESUME EN ANGLAIS.....	247

Problématique et mise en perspective

A giant panda goes into one of those expensive and pretentious restaurants serving French/Asian fusion cuisine and takes a table for one. The surprised waiter for that table explains unctuously that his name is Marcel, he will be your server tonight, and we 'ave a number of specials (he is French), etc., etc. The panda listens impassively to the list of \$27 chili-pepper encrusted swordfish specials and so on, and then orders a delicately flavored dish of young bamboo tips and mixed greenery served with steamed jasmine rice. On finishing his meal, the panda gets up, reaches into his fur for a handgun, brings down the waiter with one shot, and calmly heads for the door.

The head waiter is near the door and exclaims in shock, "Oh, monsieur, what 'ave you done? You 'ave killed Marcel! Why 'ave you done zis, monsieur? You 'ad some problem? Ze service was not acceptable?" The panda scowls at him and says, "I'm a fucking panda. Go look it up." He stalks out into the night.

The baffled staff huddle round the compact encyclopedic dictionary that they keep on the premises, and turning to Panda, giant, they read this:

Panda, giant. Large bear-like animal, Ailuropoda melanoleuca, with distinctive black and white markings, related to raccoon family. Rare; found only in bamboo forests of Tibet and western China. Eats shoots and leaves.

Extrait qui a inspiré le titre du livre de ponctuation humoristique *Eats, Shoots and Leaves* que Mark Liberman mentionna récemment dans *Language Log*.

Transmodalité et multimodalité

La communication humaine est par nature « multicanale » (Crosnier, 2003). Cette simple remarque soulève au moins deux questions importantes, relatives à la manière dont nous « optimisons » cette caractéristique, consciemment ou non, pour aller dans le sens dicté par nos intentions communicatives :

- comment sont traitées les informations qui transitent simultanément par plusieurs canaux sensori-moteurs, que ce soit pour la production de messages (geste, oral, écrit...) ou pour la perception d'informations (auditives, visuelles, tactiles, ...)? Cette question pose la problématique d'une gestion multimodale de la communication ;
- comment pallier une situation de communication dégradée par la défectuosité, temporaire ou permanente, d'un canal sensori-moteur particulier? Cette question pose la problématique de la transmodalité, c'est à dire de la transposition de l'information d'une modalité vers une autre.

Le travail de recherche qui a conduit à la rédaction de cette thèse aborde ces questions au travers de l'étude de la présentation « transmodale » et multimodale d'un document électronique.

La problématique de la multimodalité a connu, au cours des 10 dernières années, des acquis rapides en termes de domaines d'application et de démonstrateurs dans les laboratoires (Caelen, 2000), mais la question de la présentation multimodale de documents prend tout son sens aujourd'hui. Avec l'envahissement de nos lieux de vie par les technologies de l'information et de la communication (TIC), le développement accru des réseaux de la télématique et la profusion des documents électroniques, se pose le problème de la consultation de ces derniers sur de nouvelles interfaces, afin de pallier les difficultés (handicap, mobilité, taille d'écran, ...) pour rendre l'information accessible. Bien que les nouveaux outils télématiques et les technologies utilisées offrent de plus en plus de possibilités en termes de supports et de modes de présentation, cela signifie-t-il pour autant que les interfaces et les moyens d'interaction proposés soient efficaces, faciles d'utilisation, acceptés par tous et dans toutes les situations? Sont-ils par exemple capables de proposer une présentation de l'information qui favorise notre gestion multimodale naturelle de l'information ou, le cas échéant, de transformer une information depuis une modalité dans une autre afin de la rendre plus accessible?

Pour fournir quelques éléments de réponse, la pluri- multidisciplinarité s'avère un principe incontournable puisque sont mises en jeu des connaissances relatives au langage, au traitement cognitif, à l'ergonomie et l'automatisation de la présentation de l'information. Pour l'informaticien, l'ergonome, le (psycho)linguiste ou le psychologue de la cognition, l'enjeu scientifique de l'étude des présentations multimodales se trouve dans la modélisation :

- des mécanismes « machines » de fission des informations à présenter : une information d'alerte dans un cockpit d'avion pourra par exemple être présentée à la fois par un signal sonore indiquant le niveau de l'alerte, un indicateur lumineux permettant de localiser le (ou les) problème(s) et l'affichage d'un texte clignotant selon un code couleur pour exprimer le type de défaillance ;
- des mécanismes humains de fusion des informations présentées : les différentes informations reçues doivent être combinées et/ou hiérarchisées et le message global compris et mémorisé ;
- des relations de communication qu'entretiennent l'homme et la machine à travers ces deux dispositifs.

Un enjeu technologique subséquent à une telle recherche concerne l'apport de recommandations pour la conception d'interfaces/interactions entre l'homme et la machine, qui permettent une adéquation optimale des modalités aux tâches, aux profils de l'utilisateur, aux usages et aux diverses contraintes situationnelles.

Précisons tout d'abord en quelques lignes la notion de multimodalité telle que nous l'utilisons : nous nous focaliserons essentiellement sur la multimodalité en sortie n'impliquant que les sens de la vue, de l'ouïe et du toucher car ceux-ci sont, de manière générale, les plus utilisés dans le contexte de l'interaction machine → homme. Le canal de communication est défini comme le *medium*, un même *medium* pouvant servir à la présentation de l'information selon différentes modalités, c'est à dire à travers différentes formes du contenu communiqué de part sa nature linguistique (parlé, écrit, signé) ou graphique (diagrammes, dessins, ...).

Lorsque la multimodalité est abordée dans le cadre de la communication machine → homme, les modèles proposés ont le plus souvent pour objet l'architecture fonctionnelle d'un système multimodal ou l'inventaire exhaustif des possibilités qui peuvent être offertes à l'utilisateur. D'autres systèmes développés peuvent reposer sur une modélisation *a posteriori* en fonction de critères d'usage ou de résultats d'évaluations expérimentales exploratoires.

Bien que (Nigay, 1996) précise qu'il est possible d'analyser l'utilisabilité du système au regard du pouvoir d'expression des modalités impliquées, les résultats des évaluations de (Carbonell, 1994) montrent qu'il est difficile de faire un lien entre les possibilités offertes par le système interactif du point de vue de la multimodalité et de l'usage de cette dernière qui en est fait : ce n'est pas parce qu'un système interactif offre un type de multimodalité qu'elle sera forcément utilisée par l'utilisateur ; autrement dit, l'utilisabilité n'implique pas l'utilité. Aussi, malgré l'intérêt indéniable de ces travaux, la richesse expressive des différentes modalités et la complexité des relations qu'elles entretiennent entre elles, rendent bien difficile une modélisation *a priori* des conditions d'efficacité d'une présentation multimodale de l'information, modélisation qui permettrait pourtant de restreindre substantiellement le champ d'investigation expérimentale. Que le système multimodal à concevoir veuille intégrer différentes modalités selon un schéma d'interaction programmé ou préétabli (notion d'adaptabilité des interfaces), ou qu'il choisisse dynamiquement la manière d'utiliser les modalités en fonction du contexte d'interaction (notion d'adaptativité des interfaces) – (Truillet, 1999), (Thevenin, 2001) –, il convient dans un premier temps de se pencher sur les caractéristiques de chaque modalité de présentation à la fois du point de vue de leur pouvoir d'expression et de leur traitement cognitif.

La méthode qui semble souvent utilisée pour poser un tel cadre repose sur l'élaboration de taxonomies dont l'intérêt principal est de caractériser les modalités à travers des distinctions jugées utiles pour diriger la conception d'interfaces multimodales (au niveau du choix des modalités et de la compatibilité de leur combinaison). La plupart des approches de la multimodalité en terme de concept, de modèle et d'architecture (Bellik, 1995), sont proposées pour la gestion des modalités en entrée. De même, les classifications adaptées essentiellement à la multimodalité en entrée sont nombreuses (Buxton, 1983), (Foley, 1984), (Card, 1990). Nous décrivons succinctement ci-après quelques taxonomies capables de prendre en compte la multimodalité en sortie, largement moins étudiée.

Nous trouvons parmi ces tentatives de classification celle de Nigay (1994) qui distingue dans sa taxonomie M2LD, pour chaque modalité impliquée, sa correspondance avec les autres en terme de système représentationnel et de dispositif physique. Ainsi la multimodalité peut être totale lorsque des langages différents sont utilisés pour représenter des modalités utilisant des dispositifs physiques également différents. Dans le cas contraire la multimodalité est partielle, soit répartie sur plusieurs dispositifs physiques, soit utilisant divers langages de représentation.

Une autre taxonomie est celle de Frolich (1991). Afin de décrire un espace de conception adapté à la multimodalité en sortie (il décrit également un espace de conception adapté à la multimodalité en entrée sur ces mêmes bases), il s'appuie sur quatre critères de distinction :

- le mode langagier implique des interfaces langagières et le mode action des interfaces à manipulation directe ;
- 3 canaux de communication pour l'utilisateur : ouïe, vue, toucher ;
- le type de système représentationnel : parole, texte, geste (mode langagier) et son, graphique, déplacement (mode action) ;
- le style d'interface : langage de commande ou de programmation, langue naturelle, formulaire, menu (mode langagier) et environnement à fenêtre, iconique, formalisme graphique (mode action).

La taxonomie de Bernsen (2002) présente l'avantage de caractériser les modalités en fonction de critères exploitables pour la multimodalité en sortie. Dans son objectif de produire une « Théorie des Modalités », deux traits distinguent le sens humain impliqué dans la perception (sonore, graphique, tactile) et quatre critères d'oppositions :

- modalité dynamique contre statique : présence d'une dimension temporelle ou non de la communication ;
- linguistique ou non : présence d'un système structuré de symboles de communication ou non ;
- analogique ou non : ressemblance avec la réalité ou non ;
- arbitraire ou non : absence ou présence de lien entre le contenu et le contenant.

Bien que 28 classes de modalités « pures » cohérentes soient définies à partir de ces distinctions, une remarque importante est l'absence de critère relatif à la dimension spatiale de la modalité alors que cette caractéristique semble une distinction majeure entre, par exemple, l'écrit et l'oral.

Enfin les propriétés CARE (Complémentarité, Assignation, Redondance, Équivalence) identifiées par Martin (1994) et reprises par Coutaz (1995) à travers les notions d'état, de but, de modalité et de relation temporelle ont été utilisées principalement pour caractériser les systèmes multimodaux en entrée, mais peuvent être théoriquement appliquées aux interfaces en sortie.

Assignment	Équivalence	Redondance	Complémentarité
ED \xrightarrow{M} EA	ED $\xrightarrow[M2]{M1}$ EA	ED $\xrightarrow{M1 \text{ ou } M2}$ EA	ED $\xrightarrow{M1 \text{ et } M2}$ EA

Tableau 1: Propriétés CARE

Le tableau ci-dessus décrit comment s'appliquent les propriétés CARE en fonction des modalités M nécessaires pour opérer une transition (flèche) entre un état (propriétés mesurables à un instant donné) de départ (ED) vers un état d'arrivée (EA). Pour évaluer le pouvoir d'expression d'une modalité, une fonction d'accessibilité (*Reach*) affecte la valeur *vrai* à deux états pour lesquels une modalité permet de passer de l'un à l'autre. Aussi, la notion de pouvoir d'expression d'une modalité est ici considérée en aval, c'est-à-dire relativement à d'autres modalités dans un système donné, alors que notre objectif est d'évaluer le pouvoir d'expression en amont de la conception d'un système quelconque.

Une autre tentative intéressante pour définir des propriétés de combinaison des modalités est davantage orientée sur l'usage. Les types de multimodalité en sortie sont définis par les propriétés CASE (Concurrence, Alternance, Synergie et Exclusivité – Nigay, 1993) :

- la multimodalité est dite *concurrente* lorsque plusieurs messages sont produits en même temps en utilisant des modalités différentes ;
- la multimodalité est dite *alternée* lorsqu'une seule modalité est utilisée à un instant donné mais que plusieurs sont disponibles pour la présentation d'un même message ;
- la multimodalité est dite *synergique* lorsque plusieurs modalités collaborent à la présentation d'un message ;
- la multimodalité est dite *exclusive* lorsqu'une seule modalité est utilisée à un moment donné et pour un message.

En résumé, ces différentes méthodes de conception basées sur des classifications des modalités permettant d'exprimer des distinctions *a priori* utiles, ne fournissent finalement pas d'éléments clairs pour identifier des règles de choix et de composition des modalités qui assurent une utilité de l'interface de présentation de l'information. Le travail présenté dans la suite de ce tapuscrit s'appuie sur une approche originale différente.

La méthode choisie dans notre travail pour capturer ces caractéristiques est celle de la **transmodalité**. Ce terme recouvre, à l'instar du terme *mapping* (Edwards, 2002), la conversion d'une modalité de communication vers un canal de communication lié à un des cinq sens du récepteur. La littérature nous indique que cette notion a été pour l'instant principalement appliquée à la réalisation d'outil d'aide aux personnes ayant un canal de communication déficient pour, par exemple, la représentation tactile de partitions ou de diagrammes (*ibid.*) ; cela dit, son utilité nous semble pouvoir être envisagée d'un point de vue méthodologique pour concevoir des interfaces multimodales pour tous (Maurel, 2003a).

L'étude de la transmodalité est en effet susceptible de définir un cadre argumenté pour une multimodalité utile et efficace :

- dans un premier temps, l'utilité de la multimodalité se mesurera proportionnellement à notre incapacité à transposer sans « pertes » une information d'une modalité originelle vers une modalité *a priori* plus adaptée à une situation de communication particulière ;
- dans un second temps, l'efficacité de la multimodalité se mesurera proportionnellement à sa capacité de réponse aux problèmes identifiés dans le point précédent.

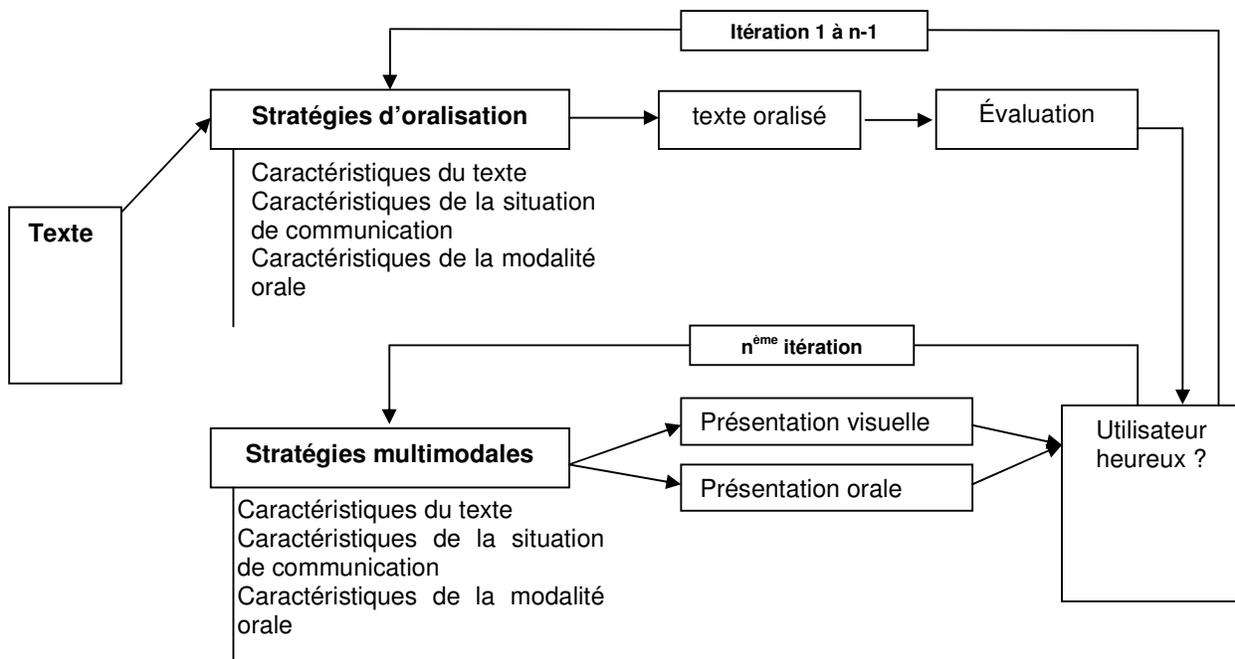


Figure 1 : Transmodalité (écrit→oral) pour une multimodalité utile et efficace

Nous avons adapté cette approche au cas de la relation écrit / oral (Figure 1). Les enjeux applicatifs d'une présentation transmodale du texte courent selon deux directions principales : la lecture sur écran de petite taille et l'accessibilité de l'information aux aveugles et aux mal voyants. Ces axes d'étude impliquent tous deux de pallier des difficultés (pour percevoir, (re)chercher une information, comprendre, mémoriser, etc.) lors d'une lecture visuelle d'un texte, gêné par des handicaps, qu'ils soient d'ordre sensoriels ou situationnels. Dans ce cadre, l'apport de la présentation orale paraît une solution intéressante.

Comme nous le dicte la démarche proposée, nous essaierons de caractériser les possibilités et les limites inhérentes à l'utilisation des modalités visuelles (textes écrits) et orales (parole), à travers l'évaluation de la transposition automatique à l'oral de textes dont la lecture a été anticipée par le rédacteur comme visuelle, c'est-à-dire qu'elle concerne des textes conçus pour être lus silencieusement et non pour être dits.

La lecture silencieuse

Il est parfois fait l'hypothèse que cette lecture silencieuse a eu un développement récent. Certains ont vu des raisons technologiques au déplacement d'une lecture orale vers une lecture silencieuse. L'idée que l'introduction de l'imprimerie conduit à un changement radical d'une culture orale à une culture visuelle, d'une lecture orale en groupe à une lecture silencieuse individuelle, a été suggérée par Chaytor (1945), popularisée par McLuhan (1962) et développée par Ong (1982). Pourtant la littérature indique (Gullick, 1986) que la notion de société médiévale purement orale est une simplification puisque des agencements de pages non linéaires complexes étaient courants. En fait le véritable élan de la lecture silencieuse viendrait plutôt selon Saenger (1982) des besoins fonctionnels de la croissance de la scolastique au XII^{ème} et XIII^{ème} siècle : la demande intellectuelle de la part des lecteurs pour des longs livres lourdement commentés et parfois contenant un grand nombre de schémas ne pouvait être satisfaite que par la technique relativement rapide de la lecture silencieuse. A partir de là, des procédés spécifiques ont pu se développer avec une relative autonomie par rapport à la langue parlée.

Sous cet éclairage, le défi que nous nous lançons en essayant de prendre en compte les spécificités de l'écrit lors de la transposition automatique d'un texte à l'oral prend un caractère quelque peu paradoxal : *il s'agit de revenir à la lecture orale d'un texte exploitant des procédés qui ont probablement émergés grâce au fait que la lecture orale n'était justement plus de rigueur !*

Cela dit c'est ce paradoxe qui nous donne quelques chances de relever la part de non transposable, de définitivement perdu par l'opération de transposition, et ainsi la possibilité d'une exploitation intéressante pour la mise en place de stratégies multimodales.

La dernière partie de cette mise en perspective aborde la problématique de la distinction entre linguistique de l'écrit et linguistique de l'oral afin de justement renforcer l'idée d'une relative autonomie de ces deux modalités de communication ; elle sera également l'occasion de présenter et de justifier le plan d'ensemble du manuscrit.

Relations entre l'écrit et l'oral

Parmi les différentes disciplines scientifiques, la linguistique pourrait être une des plus à même d'avoir à charge de distinguer et de comparer les modes de communication écrit et parlé. La distinction de deux supports de transmission d'une information en tant que critère de différenciation de deux objets d'études relativement autonomes n'a pourtant que peu été prise en compte. De manière générale, les facteurs graphiques ont reçu une attention limitée dans le champ de recherche circonscrit par la linguistique moderne. Cette vue provient directement de Saussure, usuellement considéré comme le père fondateur de la linguistique moderne et de la sémiologie, qui a placé l'écriture hors du domaine de la linguistique :

Langue et écriture sont deux systèmes de signes distincts ; l'unique raison d'être du second est de représenter le premier ; [...] le mot écrit se mêle si intimement au mot parlé dont il est l'image, qu'il finit par usurper le rôle principal ; [...] comme si l'on croyait que pour connaître quelqu'un, il vaut mieux regarder sa photographie que son visage. (Saussure, 1964 : 45)

Bien que depuis des temps récents un certain nombre de problèmes de terminologie montrent que les facteurs graphiques sont à la périphérie du champ de la linguistique, au moins deux limites anciennes sont encore présentes :

- la primauté de la parole : Bloomfield argue que l'écriture n'est pas un langage, mais seulement une manière d'enregistrer le langage par l'intermédiaire de marques visibles (Bloomfield, 1935). Le ton des défenseurs d'une primauté de la parole s'avère très incisif et autoritaire : Saussure parle de *tyrannie de l'écriture* et titre un de ses cours de linguistique générale *Influence de l'écriture ; raisons de son ascendance sur la forme parlée* ;
- la restriction à la phrase : Bloomfield accompagne sa préoccupation pour la parole de la restriction au niveau de la phrase pour la quête linguistique. Étant donné que la phrase semble être le plus haut niveau auquel les concepts de grammaticalité sont intuitivement admis par les utilisateurs du langage, les études « convenables » des linguistes se restreignaient à la phrase. La construction d'unités plus grandes, comme les paragraphes, est plus un genre de choix rhétorique que l'application de règles grammaticales.

Pourtant, beaucoup de ceux qui ont comparé la parole à l'écriture ont remarqué qu'alors que la cohésion de la structure est réalisée à l'oral par des aspects souvent considérés comme paralinguistiques, à l'écrit elles sont établies par une syntaxe plus

formelle et élaborée. Une remarque intéressante à ce propos concerne l'utilisation de ces deux modes de communication chez des enfants initiés à la lecture et à l'écriture :

L'utilisation de l'intonation par les enfants est essentielle et non une part d'arrière plan ou additionnelle de la charge de signalisation de l'information pour un message (Cook-Gumperz, 1981:101).

Ou encore :

Pour les enfants, le changement principal entre l'écrit et le parlé est le changement depuis la multimodalité de la parole vers des séquences discursives lexicalisées du langage écrit (Cook-Gumperz, 1981:101).

Les enfants compensent le manque de possibilités paralinguistiques et prosodiques de l'écriture par l'utilisation de moyens graphiques : ponctuation abondante et dramatique¹, libre mélange d'images et de mots, utilisation spontanée de l'espace et de la direction de l'écriture : les caractéristiques typographiques et spatiales semblent être plus naturelles que ce qui est habituellement pensé et enlèvent par là un peu de poids à l'idée de la primauté de la parole.

Quant à la restriction à la phrase elle doit être également minimisée pour toute étude qui veut prendre pour objet le texte écrit, en raison des problèmes soulevés par la notion de linéarité. La nature linéaire de la chaîne verbale est fondamentale pour Saussure, car à la base de ses deux catégories de relation linguistique : syntagmatique contre associative (paradigmatique) ; il précise d'ailleurs pour le signifiant auditif que

par opposition aux signifiants visuels (signaux maritimes, etc.) qui peuvent offrir des complications simultanées sur plusieurs dimensions à la fois, les signifiants acoustiques ne disposent que de la ligne du temps ; leurs éléments se présentent l'un après l'autre ; ils forment une chaîne (Saussure, 1964 :103)

et pour le signifiant graphique que

ce caractère (la linéarité) apparaît immédiatement dès qu'on le représente par l'écriture et qu'on substitue la ligne spatiale des signes graphiques à la succession dans le temps (ibid.).

Cependant, lorsque l'on considère le processus d'écriture, le texte, bien que segmenté (en mots, lignes, phrases, images, ...), peut voir sa linéarité cassée par l'utilisation des deux dimensions du support. Dans les termes de Twyman (1982), l'écrit offre la possibilité d'utiliser les caractéristiques intrinsèques du système de composition

¹ ce type de ponctuation se place selon la manière dont le texte doit être oralisé par opposition à la ponctuation dite grammaticale qui se place en fonction de la structure syntaxique de la phrase.

(gamme des caractères disponibles et variations stylistiques – gras, italique...) aussi bien que les caractéristiques extrinsèques (possibilités de manipulation de ces caractères – configuration, micro- et macro-espacement, ...). Dans ce cas, du sens peut être ajouté à la chaîne segmentée à travers la forme « analogique » de la disposition globale. De telles caractéristiques potentielles du texte (non linéaire et analogique) fournissent un défi aux modèles de la linguistique car

il n'y a aucune comparaison possible entre la linéarité temporelle pure à laquelle répond l'oral, et la linéarité à la fois temporelle et spatiale (donc sur laquelle on peut revenir), qui caractérise l'écrit (Gadet, 1996 : 19).

Remarquons que, à l'instar des images picturales, on peut aller plus loin en arguant de l'existence de trois dimensions dans une page typographique :

Le jeu relatif des corps, des graisses, des formes et des familles permet d'accentuer les différences de gris et de structures dans une troisième dimension de l'espace visuel de la page imprimée (Duplan, 1986).

Goodman a d'ailleurs essayé de déterminer si une approche linguistique doit être appliquée aux images visuelles (Goodman, 1976) ; mais selon lui seul l'écrit permet de distinguer les propriétés constitutives d'un travail (les mots utilisés sont ce qui constitue le travail proprement dit d'un écrit) des propriétés contingentes (dans lesquelles il range donc la typographie – exceptés les aspects spécifiés par le rédacteur) alors que la peinture par exemple ne le permet pas.

L'intérêt de la typographie est qu'elle s'agence quelque part dans cette dichotomie image/mot : quand les pages typographiques ajoutent des qualités spatiales et graphiques au langage verbal segmenté, elles démontrent l'utilisation simultanée de systèmes de symboles denses et articulés.

En conclusion, le signifiant auditif est

une linéarité qui n'offre aucun choix, puisqu'elle dépend d'une nécessité biomécanique (Harris, 1993),

alors qu'à l'écrit la linéarité est choisie : au niveau de la direction (sens d'écriture de droite à gauche dans les langues sémitiques) et des ruptures de la continuité (le boustrophédon commande un passage à la ligne « en zigzags » alternant à gauche et à droite). Les choix se traduisent également à travers des configurations spécifiques comme par exemple les énumérations dont Luc (2001) dit qu'elles **participent à la fois de structures entièrement visuelles qui n'ont pas de source dans l'oral et de structures entièrement discursives**. Gardes-Tamine (2004) résume ces différentes remarques en

précisant que **l'écrit implique un alignement et une** (ou plusieurs !) **direction(s) mais pas la linéarité.**

Cependant, quelle que soit la non linéarité du texte, il y a malgré tout une dimension temporelle de sa lecture tout comme il y a une dimension temporelle de la parole. Ainsi une idée intéressante à relever est proposée par Waller (1987) à travers l'hypothèse de deux types de syntagmes à l'écrit, celui du rédacteur et celui du lecteur, ce dernier représentant une entrée linéaire aux processus de cognition.

Par exemple, le Tableau 2 ci-après présente les règles de lecture de Robert Waller en regard avec les configurations typographiques dont Twyman (1979) propose un classement en fonction du degré de linéarité qu'elles génèrent (un document réel n'utilise pas, bien sûr, une seule configuration).

Ce type de tentative nous laisse entrevoir la possibilité d'une sorte de linéarisation au moins partielle du texte prenant en compte des procédés pourtant établis sur plusieurs dimensions. Savoir si, par la mise en place de ce type de stratégies, nous facilitons le déroulement des processus sensoriels (de perception du texte) et cognitifs (tels que la compréhension/mémorisation du texte, la recherche d'un élément, ...) reste une inconnue pour laquelle il s'agira d'apporter quelques éclaircissements.

Configuration de Twyman (1979)	Règles de lecture de Waller (1987)
Linéaire pur	Commencer au début et continuer jusqu'à la fin
Linéaire interrompu	Commencer au début et continuer jusqu'à la fin sans tenir compte des interruptions, qui sont arbitraires ; à chaque interruption, continuer de lire la ligne, colonne ou page suivante.
Liste	Si la liste est numérotée, commencer au début, prendre note des interruptions qui ont du sens et séparent les parties du tout. Si la liste n'est pas numérotée, les items peuvent être lus dans n'importe quel ordre.
Branchement linéaire	Commencer à un autre point pertinent, et laissez vous guider selon ce que vous lisez pour déterminer quelle partie de texte (connexe) lire ensuite.
Matrice	Sélectionner un titre de chaque axe de la matrice et regarder la cellule à l'intersection ; ou vice et versa. Ou comparer toutes les cellules d'une ligne ou d'une colonne particulière. Ou comparer les contenus de toutes les lignes et les colonnes.
Vue dirigée non linéaire	Commencer au point focal et continuer selon les instructions ou comme il paraît le plus raisonnable.
Plusieurs options non linéaires ouvertes	Faire ce que l'on veut.

Tableau 2 : Linéarité et règles de lecture selon Twyman et Waller

Dans le même ordre d'idée, nous pouvons également citer Richaudeau (1989) qui propose 6 principaux modes de lecture auxquels il associe 6 macrotypographies de base. Le tableau suivant résume ces 6 structures typographiques et montre l'importance de la prise en compte de la relation entre les modes de lecture potentiels et les types de non linéarité de la mise en page du texte (Tableau 3)

Comportement quantitatif		Comportement qualitatif	
		Intérêt égal	Intérêt variable
Lecture intégrale		Structure <i>uniforme</i> (roman, nouvelle ...)	Structure <i>hiérarchisée</i> (manuel scolaire, d'utilisation...)
Lecture partielle	Sélective	Structure <i>ponctué</i> (article, revue...)	Structure <i>modélée</i> (page de magazine...)
	De recherche	Structure <i>rythmée</i> (annuaire, bibliographie...)	Structure <i>mosaïque</i> (survol de la « une » d'un quotidien...)

Tableau 3 : Les 6 structures typographiques selon Richaudeau (1989)

Nous concluons cette introduction à notre problématique par une citation qui exprime clairement où se situe le défi que nous relevons puisqu'

on doit considérer que l'écrit n'est pas l'image de la parole mais l'image que la langue écrite, en tant que système autonome, donne de la parole. C'est cette autonomie de l'écriture qui fait qu'il n'existe aucun rapport d'imitation naturelle entre l'énonciation de vive voix et la ponctuation (Arabyan, 2001).

Même si, d'un point de vue naïf, la ponctuation, préoccupation centrale dans la citation précédente, pouvait sembler être la part typographique des textes la plus aisée à mettre en correspondance avec des phénomènes oraux, nous précisons que cette absence d'imitation naturelle peut également s'appliquer en général aux autres aspects typographiques et dispositionnels du texte.

Notre approche globale et les remarques précédentes expliquent le découpage de ce document en quatre parties.

La première partie concerne la mise en évidence d'une dimension propre aux textes écrits pour être lus silencieusement qui, bien que peu traitée par des logiciels de synthèse de la parole à partir de texte, nous semble incontournable pour une oralisation correcte en terme d'intelligibilité : l'inscription graphologique des textes (Maurel, 2003b). Après avoir exposé les limites des outils technologiques pour traiter ce phénomène, nous insisterons sur la lente prise de conscience des typographes de la part

langagière et de l'impact psychologique qui sont véhiculés par la mise en forme des textes. Nous justifions ainsi une approche revêtant un caractère fortement pluridisciplinaire et décrivons la prise en compte de cet aspect du texte à travers le point de vue des psycholinguistes et des linguistes.

La seconde partie s'inscrit dans la nécessité de se concentrer dans un premier à temps au premier visage de la transmodalité, celui de l'équivalence intermodale du contenu informationnel présenté. Pour répondre aux problèmes soulevés de ce point de vue par la première partie, nous proposons une architecture informatique basée sur une série de transformations des différents objets textuels qui composent le texte afin de fournir une représentation des textes directement oralisable par les logiciels de synthèse sans perte du point de vue du contenu informationnel. De là, nous nous focalisons sur un des modules de l'architecture en concevant un modèle qui décrive les stratégies d'oralisation par reformulation de textes écrits pour être lus silencieusement. Ce modèle prétend mettre en relation la liste exhaustive des stratégies de transposition orale prenant en compte les informations de nature sémantiques et pragmatiques véhiculées par la morpho-disposition des textes. L'opérationnalisation de ce modèle est partiellement démontrée à travers l'exemple des paragraphes, des structures énumératives et des titres.

La troisième partie s'intéresse à la seconde face impliquée par la transmodalité : l'équivalence intermodale en terme de traitement cognitif, entre l'information écrite et l'information orale présentée à travers le modèle décrit dans la seconde partie. Nous proposons l'évaluation de l'impact sur la mémorisation/compréhension de certaines stratégies du modèle. Les différentes caractéristiques d'un protocole expérimental sont justifiées, les résultats exposés et discutés.

La quatrième partie veut revenir sur notre approche globale : après avoir montré que si la deuxième partie permet d'envisager la résolution du problème de l'équivalence intermodale en terme de contenu informationnel, la démarche expérimentale de la troisième partie a mis au jour des limites de traitement cognitif, relatives au « premier regard » ou *first glance*, que nos efforts n'ont pu combler en nous appuyant sur le seul principe de la transmodalité tel que nous l'avons défini. Aussi, nous présentons le cadre théorique que cette constatation nous ouvre pour exploiter une certaine « multimodalité » du texte. Nous exploitons finalement ce cadre pour imaginer des solutions, en terme d'interactions et d'interfaces multimodales, en détournant un concept, évoqué au cours de la première partie pour représenter « l'architecture » du texte : les Images De Page (IDP). La plausibilité cognitive de ce concept et son intérêt pour une présentation bimodale de l'information sont validés par deux protocoles ; deux cadres applicatifs sont proposés pour son intégration en IHM.

« À qui appartient la ponctuation ? » Mais à tout le monde : auteurs, typographes, protes, lecteurs... Un patrimoine collectif de l'Humanité pensante. Puisse le « grand public » propriétaire ne jamais lui infliger les dégradations qu'il réserve volontiers à ses parcs, à ses jardins et à ses rivières ! »

Marc Wilmet, *À qui appartient la ponctuation ?*

Partie A Valeur et prise en compte de la morpho-disposition

Cette partie justifie la problématique de l'oralisation des structures visuelles des textes en comparant le niveau d'analyse de la typo-disposition utilisé par les systèmes de synthèse de la parole à partir de texte avec celui recommandé par les typographes, les linguistes et les psycholinguistes. Ce sera l'occasion de nombreuses discussions et de choix terminologiques qui nous guideront méthodologiquement dans la suite du tapuscrit. Nous terminons par une discussion générale au bout de laquelle est proposée une architecture plus adaptée à la perspective d'une transposition orale automatique des structures visuelles des textes écrits pour être lus silencieusement. Nous expliquons enfin les raisons qui nous ont conduits à focaliser notre propos sur le module de cette architecture, dit module de reformulation.

Introduction

Actuellement, des travaux au carrefour de plusieurs disciplines telles que la typographie, l'informatique, la linguistique ou encore la psycholinguistique, ont pour objectif la modélisation de l'organisation des textes ; et en particulier :

- intégrer la description d'un ensemble important de structures visuelles dans le cadre linguistique ;
- proposer des formalismes visant à représenter ces connaissances ;
- évaluer l'incidence des variations de l'inscription morphologique des textes procéduraux² sur la compréhension et la mémorisation.

Les résultats de ces études fournissent une base de travail pour l'identification automatique dans le texte des phénomènes structurels et rhétoriques à prendre en compte lors de la présentation du texte à l'oral.

Pourtant, les systèmes de synthèse de parole à partir de textes (TTS) tardent à intégrer réellement cette composante du texte. Pour ne parler que des marques ponctuationnelles, celles-ci (lorsqu'elles sont traitées) sont en relation quasi-bijective avec les effets que leur traitement produit : une marque entraîne toujours le même effet. Or, par exemple, un simple tiret ne devrait pas aboutir à une oralisation équivalente selon qu'il est utilisé :

- pour marquer le début d'un item d'énumération ;
- comme une marque d'incise – une parenthétique par exemple – ;
- en lieu et place d'un séparateur de type « , » - « ; » - ...

De manière générale, une marque ne peut être considérée isolément. Elle est un élément à part entière du texte, pouvant être lié à d'autres éléments par la constitution intentionnelle et signifiante d'une configuration particulière.

L'objet des quatre sous-sections suivantes est d'extraire de la littérature la place donnée à cette dimension du texte du point de vue des spécialistes de la synthèse automatique de la parole à partir de textes, des typographes, des linguistes et des psycholinguistes.

² Textes qui expliquent comment exécuter des procédures : modes d'emploi, notices d'utilisation, recettes de cuisine, etc.

L'ordre de présentation des recherches dans les deux dernières disciplines s'inscrit dans une perspective historique. En effet nous verrons que, s'il est évident que les premières réflexions ayant trait à la typographie nous sont proposées par les typographes, et que les considérations les plus récentes devraient provenir des recherches suscitées par l'apparition de technologies modernes, un certain nombre de psychologues se sont rapidement attelés à appliquer cette science humaine à la présentation typographique des documents écrits. C'est devant un échec relatif pour comprendre les mécanismes psychologiques qui conduisent à produire et comprendre cet aspect des textes qu'est apparue la nécessité de développer une approche plus linguistique pour exhiber les propriétés structurales langagières qui sous-tendent l'inscription du texte sur un support écrit. Ces psychologues précurseurs étaient donc des psycholinguistes avant l'heure puisqu'ils essayaient de rendre compte des processus cognitifs permettant le traitement de structures que l'on peut considérer aujourd'hui comme appartenant à la langue naturelle.

Aussi, bien qu'il soit important de garder à l'esprit qu'une étude linguistique se doit de précéder une étude psycholinguistique en raison des relations hiérarchiques particulières (Nespoulous, 1997) que ces disciplines entretiennent :

on peut faire de la linguistique descriptive sans s'intéresser le moins au traitement [cognitif] des structures linguistiques de la langue prise pour cible [...]. En revanche on ne peut faire de psycholinguistique sans de solides assises en linguistiques (Nespoulous, 2002).

nous avons malgré tout choisi de présenter les travaux des psychologues appliqués à la typographie et ceux plus récents issus de la psycholinguistique *avant* de considérer les avancées permises par différentes approches linguistiques de ce domaine.

Chapitre I. Synthèse de parole et typographie

Un système de synthèse de parole à partir de texte enchaîne de nombreuses procédures de calcul (Dutoit, 1997) dont la première consiste à rendre le texte prononçable en le transformant en phrases organisées en mots. Les concepteurs de ces systèmes ont rapidement pris conscience des difficultés de cette étape lorsqu'ils se trouvèrent confrontés à des sources de texte réelles, variées, voire entachées de fautes orthographiques ou syntaxiques : il est apparu qu'il était nécessaire de traiter un certain nombre de phénomènes tels que les sigles, les abréviations, les acronymes, les divers types de nombres et de chiffres ou encore les symboles non alphanumériques. Face à ces difficultés, afin d'obtenir une chaîne orthographique qui permette une transcription de graphèmes en phonèmes, divers prétraitements spécifiques sont le plus souvent appliqués à travers un module de prétraitement du texte.

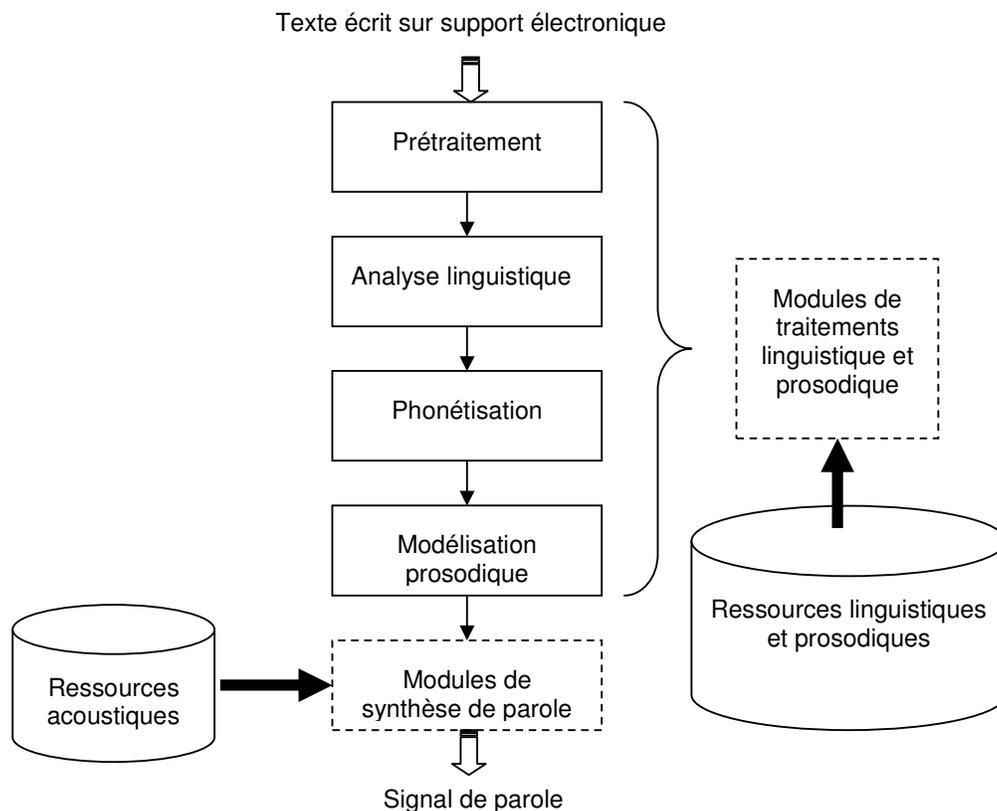


Figure 2 : Architecture « classique » d'un système de Text-To-Speech

Ce n'est qu'à partir de ce « nouveau texte » que sont envisagées la phase de phonétisation puis celle de synthèse acoustique de la suite de phonèmes. Enfin l'obtention d'une synthèse plus acceptable en terme de proximité à la voix humaine nécessite le calcul des paramètres accentuels et intonatifs adéquats (rythme, pauses, mélodie, intensité). Classiquement un organigramme des traitements pourrait être celui de la Figure 2 (selon les architectures, les modules de phonétisation et de calcul prosodique peuvent voir leur ordre interchangé ; le module de modélisation prosodique peut même être amené à disparaître dans les systèmes de synthèse récents qui s'appuient sur un principe de recopie³ de la prosodie).

Les premiers travaux qui se sont consacrés à l'étude et à l'amélioration des traitements linguistiques des textes en vue de leur transposition à l'oral recommandaient déjà un traitement particulier pour prendre en compte certains aspects dits métatextuels (ponctuation, inattendus, ...). L'objectif était de modifier la chaîne orthographique afin d'obliger le synthétiseur vocal utilisé à générer la bonne prononciation (Cotto, 1992).

Un exemple plus récent de prétraitement, dépendant de langue, est donné par Torzec (2001) pour le système de synthèse de parole multilingue développé par France Telecom R&D (CVOX) qui reformulera « Au V^{ème} siècle av. J.C., 1kg de plastique coûtait 30576\$ » en « Au CINQUIEME siècle AVANT JESUS CHRIST, UN KILOGRAMME de plastique coûtait TRENTE MILLE CINQ CENT SOIXANTE SEIZE DOLLARS. ».

Pour obtenir un tel résultat, le système s'appuie sur une analyse de surface du texte dirigée par les formes orthographiques et typographiques et des dictionnaires ; ceux-ci, à partir d'expressions régulières, permettent une segmentation du texte en entrées dites lexicales et extra-lexicales. Ces dernières sont ensuite converties dans une forme littérale appropriée à leur vocalisation. (Cherbonnel, 1996).

En ce qui concerne les pré-traitements qui sont classifiés par les concepteurs comme indépendants de la langue, nous retrouvons en particulier la tentative d'élimination ou de conversion d'informations dites métalinguistiques (mise en forme, mise en page, etc.) qui consiste à ne transcrire que ce qui semble « intéressant » ; cela dit aucune autre

³ Les voix sont basées sur de vrais voix enregistrées en lisant des textes variés pendant plusieurs heures. A partir de ces enregistrements, la base acoustique est construite. Les unités sonores les plus appropriées contenues dans cette base peuvent ensuite être sélectionnées et re-assemblées pour la restitution du texte à synthétiser en exploitant ainsi les intonations originelles du lecteur. Le signal est ensuite intégralement recréé à partir d'un modèle stocké dans la base. Ceci permet de modifier des caractéristiques du signal de parole, comme la fréquence fondamentale, la vitesse ou le timbre de la voix, sans aucun calcul supplémentaire.

précision ni référence n'est donnée sur les critères de repérage et les solutions envisagées pour la vocalisation de ces phénomènes.

D'autres exemples de pré-traitement peuvent être fournis pour des systèmes sensiblement différents. Kali, logiciel de synthèse vocale à partir de texte prétraitera « l'ADN des hommes célèbres intéresse le président. » en « l'a.d.n des hommes célèbres intéresse le président », préférant ne pas développer le sigle mais plutôt désambiguïser la nature du mot « ADN » (une syllabe /adn/ versus trois syllabes /a/de/en/) en le substituant par le sigle ponctué « a.d.n ». (Morel, 2001)

Le système industriel multilingue commercialisé en 2001 par Elan Speech gère, lors de la phase de prétraitement, un jeu de caractères (espace, tabulation, tiret, arobase, point, virgule, deux-points, slash et apostrophe) pour permettre la reconnaissance de motifs de différents types (abréviations communes, prénoms, titres honorifiques, sigles, dates, heures, durées, numéros de téléphone, monnaies, nombres, chiffres romains, URL, adresses électroniques, fractions, unités de mesure, expressions mathématiques, extensions de fichiers informatiques).

La synthèse de parole à partir de textes a ainsi connu une évolution remarquable au cours des 30 dernières années. Son intelligibilité et son naturel ont été améliorés au point que ces deux critères de jugement deviennent de plus en plus obsolètes. A partir de ce constat, nous pouvons distinguer, de notre point de vue, trois démarches qui continuent de stimuler l'investissement dans la recherche autour de ce type de technologie.

La démarche industrielle pure, par exemple, Tzoukermann (2001), poussée par la nécessité économique, a cherché principalement à minimiser les problèmes perceptifs en optimisant la qualité du timbre de la voix et en traitant un maximum de phénomènes lexico-syntaxiques tels que ceux énumérés dans les exemples précédents. Sur ce dernier point la phase de prétraitement du texte nécessite une « simple » reformulation, sous une forme discursive, des sigles, abréviations, unités de mesure, formules mathématiques, systèmes de numérotations ou d'adressage.

La démarche scientifique regroupe souvent deux catégories de chercheurs dont les motivations principales résident dans la réalisation d'outils pour (1) valider certaines techniques utilisées en traitement automatique du langage naturel ou (2) mettre à l'épreuve des théories et des modèles prosodiques. Dans ces deux cas, les prétraitements automatisés opérés sur le texte sont également minimaux puisque il est possible et parfois même nécessaire d'éviter sciemment les problèmes posés par la mise en forme des textes. Béchet (2001), qui relève de la première catégorie, propose une phase de

formatage pour segmenter le texte en mots, phrases et paragraphes. Remarquons qu'il rejoint partiellement nos considérations car la distinction de ce dernier objet textuel est utilisée principalement pour conserver la structure logique et isoler les titres afin de les traiter de manière spécifique. Quant aux recherches qui relèvent de la mise en œuvre de cadres théoriques de la prosodie, elles ne s'occupent généralement pas des textes mais des démarches méthodologiques pour élaborer le module prosodique du système de synthèse. Le système SYNTAIX (Di Cristo, 2001) en est un exemple ; citons également Mertens (2001) qui a pour objectif un système complet de synthèse de la parole à partir du texte de bonne qualité mais ne partant que de la structure syntaxique fournie par une analyse basée sur la théorie Chomskyenne des principes et paramètres (Chomsky, 1995). Ainsi ce système effectue une segmentation du texte brut en phrases démunies de la ponctuation originale. Ne seront alors traités spécifiquement que quelques phénomènes intra-phrastiques comme les sigles, les nombres ou les expressions numériques.

Certaines démarches hybrides rejoignent plus précisément notre problématique car, en alliant le souci d'efficacité pratique et la recherche de verrous scientifiques intéressants, elles abordent plus facilement la problématique qui nous préoccupe ici. Par exemple certaines synthèses issues de ce type de recherche disposent d'un module de prétraitement de mèl (Sproat, 1998), (Torzec, 2001), qui tient compte de certains phénomènes typographiques, spécifiques à ce mode de communication, se situant à un niveau plus complexe que celui de la lexico-syntaxe (par exemples : les *smileys* ou « émoticônes » ; la gestion de la succession des réponses et des transferts). D'ailleurs cette problématique émergente a été relevée, dans le cadre de la présentation du système de synthèse de la parole multilingue Elan Speech, comme une évolution intéressante qui nécessiterait

de déborder du cadre de la phrase, objet de la syntaxe (d'aller peut-être jusqu'au paragraphe, voire à une hiérarchisation du document, pour ne pas lire de façon linéaire, par exemple, les titres, les chapitres ou les notes de bas de page), et d'étudier dans quelle mesure il existe, dans les formes, des indices permettant de prendre en compte le sens (Boula de Mareuil, 2001 :248).

Malgré cela, lorsque nous étudions les systèmes récents (Alessandro, 2001), nous remarquons que la mise en forme du texte reste encore peu prise en compte.

Une des explications d'un traitement de la typographie à un niveau de prétraitement du texte, et pas au niveau de l'analyse linguistique, provient du fait de considérer davantage la typographie comme un canal de transmission du langage que comme faisant partie du langage même.

Nous retrouvons d'ailleurs cette idée depuis fort longtemps chez les typographes eux-mêmes, en atteste cet extrait d'un imprimeur anglais du XVII^{ème} siècle :

A good Composer is ambitious as well to make the meaning of his author intelligent to the Reader, as to make his Work shew graceful to the Eye, and pleasant in Reading. (Moxon, 1683/1962 :211).

La conception de la typographie comme relevant seulement d'une sorte de clarté esthétique a été souvent exprimée plus « contemporanément ». La métaphore dite de la coupe en cristal (*the crystal goblet metaphor*) en est un exemple : Warde (1955) affirme que la typographie devrait être assez transparente pour ne pas ajouter sa propre coloration aux intentions de l'auteur. Elle devrait être complètement naturelle et discrète, la forme des lettres parfaitement harmonieuses et sans maniérisme, les lignes et les colonnes d'imprimerie arrangées pour diriger les yeux sans à-coups qui entraîneraient distraction ou fatigue.

Pourtant, les typographes en sont venus peu à peu à considérer cet aspect des textes écrits pour être lus silencieusement sous un angle plus pluridisciplinaire. En effet, cette vision historique contraste avec ce qui est aujourd'hui communément appelé la typographie moderne et dont les origines remontent au moins à 1919 avec la création du *Bauhaus* qui étendit ses recherches à tous les arts majeurs et appliqués, en vue de les intégrer à l'architecture. Parmi les principes de la célèbre école allemande s'inscrivait l'adaptation aux nécessités du monde industriel moderne et donc la recherche d'une harmonie entre esthétique et technique. Le *Bauhaus* proposait ainsi non seulement des classes de sculpture, de peinture et d'architecture mais aussi des cours d'arts décoratifs ou appliqués à l'industrie et au commerce ou encore consacrés à la symbolique formelle et à la typographie. La typographie moderne a ainsi transposé vers l'édition le langage visuel, particulier aux nouveaux mouvements d'art (futurisme, dadaïsme, constructivisme, ...), en utilisant les principes de tension visuelle, de contraste et de rythme pour manipuler l'attention des lecteurs. Comme l'explique une figure majeure de la nouvelle typographie, professeur au *Bauhaus* :

Au contraire d'un équilibre statique-concentrique vieux de plusieurs siècles, on cherche aujourd'hui à produire un équilibre dynamique-excentrique. Dans le premier cas l'objet typographique est capturé d'un regard, avec tous les éléments centraux focalisés– en incluant ceux à la périphérie ; dans le second cas, l'œil est conduit pas à pas d'un point à un autre, de façon à ce que la conscience des relations mutuelles entre les éléments individuels ne soit pas affectée (Lazlo Moholy-Nagy cité par Kostelanetz, 1970 :80).

Dans un tel effort la police de caractère devait être claire et simple et présentée en gras avec une grande taille, les puces furent introduites et l'espace utilisée pour regrouper l'information. De tels traits ont été ensuite absorbés par le répertoire typographique général.

C'est à partir de ce courant que l'idée d'une typographie fortement liée au langage a pu véritablement apparaître et se développer. Karl Gerstner, issu de la typographie fonctionnelle suisse, est un représentant des premières véritables tentatives d'une telle connexion, basées sur l'analyse systématique et consciencieuse des propriétés expressives des polices de caractère et une exemplification se référant à des acteurs majeurs d'autres disciplines tels que des psychologues ou des linguistes :

Le mariage du langage et des caractères d'imprimerie a pour résultat une nouvelle unité, un tout supérieur. Texte et typographie ne sont pas tant deux processus consécutifs de différents niveaux que des éléments qui s'interpénètrent (Gerstner 1959 :66).

Une des dernières contributions influentes de la typographie suisse est le système de grille de maquette de la disposition de la page (Gerstner, 1974) pour répondre aux besoins techniques de publication de l'information dans un pays multilingue ; ce système a été ensuite largement utilisé de manière plus générale pour son utilité dans l'illustration des livres et des magazines.

Le rôle de ce très rapide retour sur l'histoire de la typographie contemporaine est de justifier les deux chapitres suivants dont le premier est consacré au point de vue des psychologues et plus particulièrement de la psychologie appliquée à la typographie et donc (et il s'agit déjà d'une prise de position) à la linguistique ; tandis que le deuxième se concentrera justement sur l'intérêt qu'ont montré les linguistes pour l'étude de typographie.

Chapitre II. Du point de vue des psycholinguistes

II.1 La Gestalt-Théorie

Le verbe allemand *gestalten* signifie « mettre en forme, donner une structure signifiante ». La Gestalt est ainsi une forme structurée à laquelle nous pouvons accorder un certain sens.

Selon les fondateurs de cette théorie psychologique, tels que Max Wertheimer ou Kurt Koffka (Ellis, 1938/1997), nos perceptions obéissent à un certain nombre de lois : ainsi, une totalité ne peut se réduire à la simple somme des stimuli perçus ; de même que l'air que nous respirons est autre chose que l'addition d'azote, d'oxygène, d'argon, de dioxyde de carbone, de néon, de méthane, d'hélium, de krypton, d'hydrogène et de xénon ; qu'une œuvre musicale est plus qu'une succession rythmée de notes ; le langage ne peut se réduire à l'empilement syntagmatique d'unités discrètes : le tout est différent de la somme de ses parties. La Gestalt-Théorie affirme également qu'une partie dans un tout est autre chose que cette même partie isolée ou incluse dans un autre tout, puisque sa place et sa fonction génère des propriétés spécifiques : être nu n'a pas le même sens dans son bain que sur la place du Capitole ! Un mot en *italique* par exemple ne subit probablement pas la même loi de perception que *le même mot, toujours en italique, mais dans un contexte phrastique et stylistique différent ...*

La compréhension d'une situation ou d'une information demande donc une capacité d'analyse qui s'appuie sur une vue synthétique, c'est-à-dire sur le contexte global de sa production.

Les psychologues de la Gestalt de l'Allemagne des années 20 ont dégagé ainsi un certain nombre de lois dont certaines peuvent apparaître pertinentes pour décrire la perception de la typographie. De manière générale, ces lois suivent le principe selon lequel si plusieurs structures alternatives sont possibles c'est celle qui démontre les meilleures qualités de stabilité et de simplicité qui sera sélectionnée. Ainsi, le principe de proximité explique la possibilité d'utiliser l'espace pour regrouper des composants du texte pour, par exemple, l'organiser en lignes et colonnes. Lorsque ce n'est pas l'utilisation de l'espace, ce peut être un principe de similarité d'éléments qui crée le regroupement donnant par exemple la possibilité de choisir une police de caractère pour signaler un fait textuel particulier. Les titres sont aussi des regroupements par similarité. Enfin le typographe peut s'appuyer sur le principe de fermeture qui exploite notre tendance à « boucher les trous » entre les différents éléments graphiques et à voir des

formes stables dès que c'est possible. Les grilles de maquette typographique sont un exemple qui est basé partiellement sur ce principe de fermeture.

A partir de là, un grand nombre d'études de l'effet de la typographie sur les lecteurs ont été publiés par les psychologues spécialisés dans divers « domaines ».

Nous reprenons en partie dans la sous-section suivante la description de certaines de ces recherches telles qu'elles sont présentées dans Waller (1987), c'est-à-dire à travers les trois « domaines » de la psychomotricité (les qualités « physiques » du texte), de l'affectif (l'appréciation esthétique et morale du texte) et du cognitif (les qualités cognitives du texte). La valeur principale de ce découpage réside dans le cadre de travail qu'il fournit pour une discussion autour de la psychologie appliquée à la typographie.

II.2 Psychologie appliquée et typographie

II.2.1 Rôle de la lisibilité comme qualité « physique » du texte

Les premières approches scientifiques appliquées à la typographie peuvent être attribués à Anisson, typographe, Leclerc, notaire, et Javal, ophtalmologiste. Le premier démontre en 1800 que lors d'une lecture de deux textes imprimés dans des polices de caractère différentes, Garamond et Didot, la distance physique du texte par rapport au lecteur ne produit pas les mêmes effets : le Garamond permettrait de percevoir et lire le texte plus longtemps que le Didot. Cette expérience est malgré tout tempérée par l'usage prépondérant du Garamond dans l'imprimerie de l'époque et la conclusion contemporaine est que le choix d'une police de caractère doit particulièrement s'inspirer des imprimés les plus communément utilisés. La seconde expérience, de 1843, permet de dégager 5 points (le lecteur devine plutôt qu'il ne lit, la moitié d'un mot suffit pour le deviner, il peut aussi être deviné par les deux mots qui l'encadrent, les lecteurs ne lisent que la partie supérieure des lignes et celle-ci suffit pour une lecture courante et rapide) et une conclusion : *la partie inférieure des lignes est inutile !* La troisième étude expérimentale est menée par Javal en 1878 et permet entre autres choses de montrer qu'en lecture normale, des secteurs de 10 lettres en moyenne sont fixés par le lecteur en « sautant » de l'un à l'autre par des saccades très vives.

Un nombre considérable d'études furent également publiées dans la première moitié du XX^{ème} siècle. C'est Legros (1922) qui conçoit une des premières listes de variables typographiques : taille des caractères, épaisseur des séparateurs, espace blanche entre les séparateurs, dissimilitude des caractères, longueur de la ligne, similarité et largeur des figures, variation de la hauteur des caractères, qualité, couleur et réfléchissement de la lumière du papier, couleur de l'encre, enluminure... Mais parmi les expériences les

plus marquantes, puisque encore citées, nous pouvons relever celles de Tinker depuis 1929 jusqu'à ses dernières publications (Tinker, 1955, 1963 et 1965). Cela dit elles ont été sévèrement critiquées dès Buckingham (1931). Les critiques principales reposent sur les implications de la méthode scientifique utilisée, du fait qu'elle reposait sur des modèles univariés :

- recommandations pas toujours en lien avec les données ;
- alors que les variables typographiques interagissent :
 - o les recommandations pour la longueur des lignes ne sont pas fonction de l'interlignage,
 - o les recommandations donnent la taille du corps sans expliciter la police de caractère utilisée ou l'interlignage ; le travail est largement gâté par le fait de travailler en fonction de la taille du corps de la police de caractère en nombre de points et pas en fonction de la taille mesurable visuellement.

D'autres critiques plus générales ont également été formulées :

- les imprimeurs perçoivent l'ignorance des expérimentateurs et ignorent de toute manière leurs résultats ; ce fait est parfois aggravé par les difficultés à les comprendre que peuvent éprouver les imprimeurs ou les maquettistes ;
- une étude complète d'un modeste échantillon de polices de caractère, tailles, longueurs de ligne et interlignage demanderait un effort que personne n'est préparé à fournir ;
- les études portent sur d'infimes détails plutôt que des problèmes de large envergure ;
- les recherches tendent à être déconnectées des préoccupations réelles des utilisateurs quand un choix doit être fait concernant une police de caractère ; les maquettistes voudraient plus de détails sur les caractéristiques de performance de chaque polices de caractère : par exemple, peuvent-elles être réduites ou photocopiées ?

Tinker se défendra dans ses dernières publications en arguant que malgré tout

avant la recherche scientifique, les imprimeurs et les concepteurs de polices de caractère étaient seulement concernés par l'apparence esthétique de la page imprimée. Cette préoccupation pour l'esthétique, accompagnée de considérations économiques et traditionnelles, ont dominé toute la typographie depuis les années 1920. Cette vision obstructive a engendré une lenteur dans le développement d'une

typographie scientifique. En fait, l'industrie de l'imprimerie continue de résister aux changements procéduraux suggérés par les conclusions expérimentales (Tinker, 1965 :115).

Malgré tout, l'addition de ces critiques à la capacité des imprimeurs à généralement prendre les bonnes décisions seuls, ont conduit la plupart des chercheurs à ne plus investir dans la recherche sur la lisibilité. Néanmoins, nous pouvons citer un regroupement marquant dans les années 1980, *The Readability of Print Research* situé au *Royal College Of Art*. Son point fort est qu'il répond à au moins la critique d'absence de cadre de dialogue entre imprimeurs et chercheurs ; en effet ce regroupement combine l'habileté des concepteurs et des psychologues pour étudier non seulement la lisibilité mais également la signalisation typographique et spatiale pour la disposition des bibliographies (Spencer, 1975) ou encore les problèmes de recherche d'information (un nom dans un index ou une bibliographie). Parallèlement s'est développée une autre équipe composée d'un typographe et d'un psychologue, Peter Burnhill et James Hartley, mais dont les travaux se sont rapidement déplacés depuis le champ de la lisibilité vers celui de l'information structurée dont nous reparlerons dans la suite. Une série de résultats intéressants provient également des travaux de Richaudeau (1984) sur la lisibilité dans lesquels il démontre que la lettre n'est pas l'unité de lecture mais bien le mot, et souvent le groupe de mots.

Bien sûr, il est un domaine pour lequel la lisibilité d'un message est une qualité primordiale : la publicité. L'expérience dans ce domaine de Ogilvy (1977) nous montre par exemple que le titre est plus lu que le corps du texte, les légendes des photos que les articles, qu'une grande lettrine en début de texte augmente le taux de lecture de 13%, que plus larges sont les colonnes moins nombreux sont les lecteurs, qu'une suite de faits indépendants les uns des autres nécessite une numérotation simple, ...

Enfin, les nouvelles technologies d'affichage qui apparaissent dès 1970 ainsi que l'apparition de dispositifs techniques tels que les appareils de suivi du regard ont suscité un regain d'intérêt pour ce domaine de la psychologie, en particulier pour les chercheurs en « facteurs humains » ; cet effort se poursuit par exemple aujourd'hui pour le problème de la présentation de l'information sur petits écrans (Kamba, 1996), (Rahman, 1999), (Russell, 2002).

II.2.2 Les propriétés expressives de la typographie

Berliner (1920) se posait déjà la question de la sensibilité des lecteurs aux propriétés expressives des polices de caractère désirées par le typographe. Des expériences ont permis alors d'identifier la pertinence de polices de caractère en fonction du thème des textes sur lesquels elles étaient appliquées. De là étaient tirées des propriétés qui

pouvaient être utilisées pour « traduire » une police de caractère devenue obsolète par une autre, tout en conservant sa valeur expressive. D'autres études reliaient plus indirectement les polices de caractère au thème à travers différents contrastes du type dur/mou, actif/passif, etc. en faisant l'hypothèse qu'une police qui possède un certain pouvoir expressif devrait être utilisée pour un message qui possède les mêmes particularités (Ovink, 1938).

Cependant Zachrisson (1965) montre un biais expérimental pour la plupart de ces études en raison du manque général de contrôle de la variable liée à l'éducation littéraire et artistique des sujets (il démontre que l'habilité à discriminer des polices de caractère dans le cas des couvertures de livre peut être nulle pour le profane).

C'est à partir de la classification des caractères de Maximilien Vox en 1952, qu'un certain nombre de travaux se sont orientés sur une classification des « sensations typographiques » en fonction de trois registres : le jaillissement, l'objectivité et l'intégration.

Selon Duplan, le jaillissement

utilise la force plastique des contrastes, qui peuvent être de nature différente : contraste de forme, de valeurs, de couleurs, de matières, de directions, de surfaces, etc. (Duplan, 1986)

et entraîne une certaine violence visuelle. L'intérêt pour les représentants de cette famille de police de caractère serait d'être d'abord vus et perçus et de n'être lus et compris qu'ensuite. Elle est très utilisée, pour cette capacité à capter l'attention, dans la communication de masse telle que l'affiche ou l'annonce publicitaire.

La sensation d'objectivité est produite par

une typographie visible, scientifique et sert à transmettre le plus facilement possible au lecteur, un message textuel le plus souvent complexe : la structure du message est rendue évidente. (*ibid.*)

Enfin, l'intégration utilise

un registre de sensations calmes dans une tonalité harmonique qui va du détail à l'ensemble ou inversement. (*ibid.*)

Ce type de classification donne l'opportunité à un rédacteur de choisir une police de caractère selon que le texte doit être appréhendé sur le mode de la lecture continue ou de la consultation.

Rajoutons également qu'en matière d'expressivité une certaine part d'invariabilité des formes typographiques est susceptible de s'ajouter aux significations des autres facteurs : droit contre italique, capitale contre bas de casse, maigre contre gras. Dans cet

éventail de choix successifs, il est à noter que les typographes attribuent généralement au style droit l'expression d'une certaine rigidité plastique, de stabilité ; tandis que l'italique signale de la souplesse et dénote des relations plus intimes, secrètes ou sensibles avec le lecteur. Sa nature instable invite à une lecture plus rapide et accélérée. La Capitale quant à elle porte une certaine emphase en valorisant le texte ; le bas de casse véhicule la notion de simplicité et s'éloigne, par comparaison à la Capitale, de la notion d'autorité. Le jeu des graisses permet d'affirmer plus ou moins la présence d'un mot. Ces différentes valeurs pourront signifier différents contrastes, comme par exemple une expression chuchotée, parlée ou criée.

Nous pouvons également évoquer l'expressivité dégagée par le choix des différentes proportions pour la mise en page du texte : une exploitation connue de cette fonction est liée au nombre d'or pour construire une mise en page idéale basée sur le rectangle d'or :

L'emploi de la section dorée donne une impression de repos, de sécurité, de constance dans un rythme indéfiniment continué (Bouveresse, 1976)

Enfin, l'expressivité peut être conduite à son paroxysme par l'utilisation de la typographie expressive qui consiste à additionner par le seul fait typographique du sens aux mots ; ainsi les

mots deviennent images du mot, du son, d'un gestuel que le mot propose (Duplan, 1986 :163)

Nous pouvons donc conclure que

l'étude des différentes familles de caractères permet de mettre en évidence à la fois la plastique des lettres, leur lisibilité mais également une dimension intellectuelle spécifique qu'elles confèrent aux textes qu'elles signifient (Duplan, 1986 :87).

C'est sur ce dernier aspect que nous nous focaliserons principalement avec la sous-section suivante, à travers l'impact de la structuration typographique de l'information sur son traitement cognitif.

II.2.3 Influence de la typographie sur le traitement cognitif

Dans cette section, nous présenterons des travaux ayant étudié l'effet du marquage typographique sur la lecture, la mémorisation et la compréhension.

II.2.3.1 La place de la mise en forme des textes sur le plan empirique

En ce qui concerne l'effet du marquage typographique sur le traitement cognitif du texte, la plupart des études⁴ a montré une amélioration de la mémoire des informations marquées typographiquement (Glynn, 1979 ; Lorch, 1995). Malgré l'observation de cet effet robuste, rares sont les travaux qui se sont intéressés à l'effet des variations typographiques sur l'attention ou la compréhension si ce n'est Glynn (*op. cit.*) qui remarque que l'amélioration de la qualité du rappel des idées marquées typographiquement se fait au détriment des idées non marquées. De plus l'amélioration est facilement montrée en terme de rappel immédiat mais pas de rappel différé. A ce propos il est intéressant de noter que ce dernier fait n'est pas inattendu puisque la signalisation typographique permet au sujet « de laboratoire » d'anticiper pendant la lecture les questions qui vont lui être posées. D'ailleurs, dans une étude de Coles (1975) pour laquelle des étudiants n'avaient pas été informés qu'ils seraient testés sur le matériel marqué qu'ils utilisaient, a été constaté un échec de la signalisation typographique utilisée. Les auteurs précisent que cette condition expérimentale avait pu entraîner la perception de la typographie comme déroutante voire même distrayante plutôt qu'utile.

Dans une perspective d'efficacité cognitive, les précurseurs d'expérimentations sur des formats innovants peuvent être représentés par Hartley et Burnhill, souvent cités pour leur recherche en psychologie sur la typographie. Leurs premières publications concernent les problèmes de lisibilité tandis que les dernières abordent une sphère plus large de problématiques autour de la typographie (Hartley, 1979) : l'information structurée, les textes à consignes et le processus de lecture. Ils ont ainsi dirigé la publication d'un certain nombre d'études dans un numéro spécial de *Visible Language* (Hartley, 1981) avec la consigne originale donnée aux contributeurs de pratiquer pour la rédaction de leur article justement ce qu'ils recommandaient ; les lecteurs pouvaient ainsi juger par eux-mêmes des nouveaux formats proposés. Cela dit la réussite de ces recherches n'est pas évidente comme en attestent les trois exemples suivants extraits de Waller (1987).

On trouve dans ce numéro spécial un article de Jewett (1981) qui propose plusieurs niveaux d'indentation pour indiquer les niveaux hiérarchiques du raisonnement. L'article est imprimé au format paysage pour permettre une indentation relativement généreuse sur trois niveaux. Ironiquement cela entraîne des lignes excessivement longues qui entravent la recherche rapide d'information alors que le système

⁴ Notons tout de même que ce fait a parfois été mis en défaut (Garcia-Debanco, 2001)

hiérarchique est censé la faciliter. De plus les paragraphes de plus haut niveau font références à des informations contenues dans les niveaux inférieurs ce qui ne permet pas un parcours de l'article en ignorant ces derniers. L'argument de l'auteur est que ce format permet d'absoudre le rédacteur de la responsabilité de verbaliser la structure hiérarchique, cependant cela n'a visiblement pas suffi à ce que ce procédé se développe...

Un autre article (Shebilske, 1981) distingue trois sortes de contenu : du texte non marqué, du texte mis en gras pour indiquer les idées importantes et la mise entre crochets pour indiquer l'essentiel de chaque idée. La lecture de l'article dans le sens auquel l'auteur aspirait, se révèle quasi impossible : la mise en gras correspond à l'usage conventionnel mais l'utilisation des crochets pour mettre en valeur l'essentiel d'une idée va à l'encontre du sens habituel véhiculé par le parenthésage qui est en général d'intercaler du texte peu important : nous ne nous attendons pas naturellement à voir chaque phrase ou paragraphe comme contenant un noyau de mots isolés typographiquement comme « l'essentiel ». Profitons de cet exemple pour signaler qu'une étude de Hershberger (1965) montre, par la comparaison de différents niveaux de complexité pour la signalisation (jusqu'à 5 niveaux dans les expérimentations), qu'aucun avantage n'était acquis lorsqu'il était distingué plus de deux niveaux.

Quant à Hartley et Burnhill, ils préconisent trois principes pour la prise de décision typographique : l'utilisation d'une taille de page standard (A4), d'une grille de maquettage pour pré-organiser les pages et d'une utilisation pertinente de l'espace pour transmettre la structure de l'information. Cette dernière recommandation semble pourtant difficile à justifier car l'utilisation de l'espace plutôt que d'autres procédés typographiques n'est défendue par aucune explication en terme de perception, de traitement cognitif, de mémoire de travail ou quelque autre modèle psychologique. Pour une réflexion sur l'impact de la disposition dans les textes narratifs nous renvoyons à Terrier (2000).

En fait, en raison d'une certaine forme de naïveté typographique (volonté réformatrice d'un système jugé à tort trop irrationnel) et linguistique (alors que beaucoup d'aspects peuvent être captés par un lecteur habile sans avoir recours à la typographie mais directement par la structure classique des différents éléments verbaux du discours) de certains auteurs, l'intérêt de ces études consiste surtout dans le déplacement de la recherche en psychologie appliquée à la typographie, telle qu'elle était généralement menée jusque là, vers les champs délicats de la cognition et de la sémantique. Nous aurons l'occasion de revenir sur le dernier point dans la sous-section suivante consacrée à une « linguistique de la typographie ».

En ce qui concerne les aspects cognitifs, d'autres travaux ont été menés par Lorch autour de la notion de « signaux » du texte, c'est-à-dire des dispositifs écrits qui mettent l'accent sur certains aspects du texte sans y ajouter de contenu (Lorch, 1989). Ils s'agit des marques, aussi bien typographiques que discursives, qui s'appliquent par exemple aux titres, aux résumés ou aux plans initiaux.

Lorch (1986) étudie l'effet des signaux numériques sur la lecture et la mémorisation des textes. Les résultats obtenus indiquent que le marquage numérique dirige l'attention sur les phrases marquées (temps de lecture allongé) et influence le processus de rappel (meilleur rappel de l'organisation du texte). Selon les auteurs ces résultats œuvrent en faveur de l'hypothèse « attentionnelle » (Loman, 1983) ; d'après cette hypothèse, le marquage des informations amène les lecteurs à accorder plus d'attention à l'information marquée et donc à mieux l'encoder. Cette hypothèse explicative des effets du marquage typographique porte donc sur des bas niveaux de traitement et concerne l'étape d'encodage du texte.

Gaonac'h (1990) examine l'effet du marquage de l'« importance » sur la lecture. Les résultats révèlent que lorsque l'importance est marquée par un procédé typographique de type soulignement, le temps de lecture sur la phrase précédant la phrase soulignée diminue et le temps de lecture sur la phrase soulignée augmente tandis que lorsque l'importance est marquée discursivement, une diminution du temps de lecture de la phrase suivant la phrase importante est observée. Ainsi, il semble que le marquage de l'importance affecte les temps de lecture de façon anticipée lorsqu'il est typographique et de façon ultérieure sinon. Les résultats obtenus par Passerault (1991) vont dans le même sens ; les auteurs étudient les temps de lecture en fonction de la présence ou non d'une marque de paragraphe et de la position du segment lu par rapport à cette marque. Les résultats mettent en évidence un effet d'intégration : le temps de lecture du segment qui précède une rupture de paragraphe est allongé. Les auteurs attribuent cet allongement du temps de lecture à la mise en œuvre par le lecteur de traitements intégratifs des informations du paragraphe. Le lecteur repèrerait donc cette rupture de paragraphe de façon anticipée et l'interpréterait comme « une instruction de clôture ».

Ces deux derniers résultats ont ceci en commun qu'ils mettent au jour, même indirectement, une propriété essentielle de la dimension morpho-dispositionnelle pour saisir son impact sur les traitements cognitifs. La typo-disposition est appréhendable par le lecteur de façon anticipée et globale. C'est ce que nous appellerons la propriété « first glance ». Bien que cette propriété n'ait jamais été directement étudiée, plusieurs arguments empiriques jouent en sa faveur. Ainsi, Lorch (1995) explore l'effet du marquage typographique sur la lecture et le rappel de texte. Dans une première

expérience, il montre que les phrases soulignées sont mieux rappelées que les phrases non soulignées seulement lorsque le marquage reste léger. La procédure de présentation utilisée est la suivante : les sujets disposent du texte écrit en entier et lisent le texte à leur propre rythme. Dans une seconde expérience, ils testent l'effet de la mise en majuscules de certaines phrases du texte avec une procédure de présentation légèrement différente : les sujets lisent le texte présenté sur ordinateur phrase par phrase. Les résultats sont identiques à ceux observés dans la première expérience mais l'ampleur de l'effet est moitié moindre dans le cas de la présentation phrase par phrase. Pour les auteurs, cette différence peut être liée à une moindre efficacité du marquage par majuscules par rapport au soulignement ou au changement de procédure de présentation opéré par rapport à la première expérience : la présentation phrase par phrase rendrait impossible la vue complète de la page, ce qui, selon les auteurs, pourrait interférer avec l'effet de marquage. Cette remarque est capitale de notre point de vue puisque elle signifie que l'effet du marquage typographique serait en partie médiatisé par la possibilité d'appréhender globalement le texte à des fins d'anticipation.

Les travaux de Schmid et Baccino (Baccino, 1999 ; Schmid, 2001) plaident également en faveur de l'existence de cette propriété « *first glance* ». Ces auteurs ont montré que les « indices visuels de structuration » exercent un effet sur les stratégies d'encodage du texte. Dans une expérience, les auteurs comparent les stratégies de lecture mises en œuvre lors de la présentation d'un texte avec ou sans indices visuels. Les résultats obtenus indiquent que la structuration visuelle du texte engendre une stratégie de lecture de type hiérarchique, tandis que la structuration par l'utilisation de marques linguistiques entraîne une stratégie de lecture temporelle qui gère l'encodage des informations au fil du texte. Un texte exploitant la typo-disposition permettrait au lecteur de mettre en œuvre un pré-traitement global de la structure du texte. Nespoulos (2002) va même jusqu'à avancer la notion d'« avenance textuelle » comme une transposition de la notion d'« affordance » au texte ; la notion d'affordance renvoie au fait que ce sont les propriétés de l'environnement qui conditionnent notre traitement de l'information (Gibson, 1977). Le texte posséderait, à des degrés divers et à travers l'utilisation des marques typographiques, une affordance conditionnant son traitement. Ainsi, le fait que la dimension morpho-dispositionnelle du texte soit appréhendable globalement conditionnerait le traitement des informations du texte. Cette idée trouve un écho tout particulier dans les travaux de Goody pour qui l'apparition de l'écriture a profondément modifié la parole et surtout les processus cognitifs. Sa conception s'oppose à l'attitude saussurienne selon laquelle l'écriture ne serait qu'une transposition directe du langage oral :

l'écriture a une importance décisive, non seulement parce qu'elle conserve la parole dans le temps et dans l'espace mais aussi parce qu'elle transforme le langage parlé [...]; ainsi la communication par l'œil engendre des possibilités cognitives nouvelles par rapport à celles qu'offre la communication par la voix. (Goody, 1979)

Pourtant les principales modélisations psycholinguistiques de la compréhension du texte manquent singulièrement de considération pour le support matériel sur lequel il s'inscrit. Nous allons le montrer dans la sous-section suivante.

II.2.3.2 Typo-disposition et modèle propositionnel de la compréhension du texte

Une première remarque qui met en évidence la non prise en compte de la modalité verbale est faite par Kintsch (1978) lorsqu'il précise que le modèle propositionnel de compréhension du texte peut s'appliquer aussi bien à la compréhension du texte écrit qu'à celle du discours oral. Selon les auteurs, il est possible de s'affranchir des propriétés spécifiques du texte écrit ou du discours oral pour modéliser la compréhension. Notre position est que la façon dont le texte exploite son inscription sur un support matériel exerce une influence sur les traitements cognitifs du texte, qu'il s'agisse de traitement de bas niveau (prise d'information) ou de haut niveau (compréhension, mémorisation). Étudions maintenant plus précisément quel rôle jouent les caractéristiques morpho-dispositionnelles dans la compréhension du texte d'après le modèle propositionnel.

Pour ce modèle, l'unité cognitive de traitement et de représentation du texte en mémoire est une structure propositionnelle du type prédicat-argument. La compréhension du texte renvoie à la construction de trois niveaux hiérarchisés de représentation du texte : le code de surface, la base de texte et le niveau du modèle de situation. Le code de surface est une représentation des propositions du texte qui conserve le lexique et la syntaxe empruntés par le texte. Cette représentation est très peu pérenne en mémoire de travail et elle ne porte que sur les propositions les plus récemment lues. La base de textes est une représentation mentale du sens du texte, constituée d'un ensemble de propositions organisées en une structure cohérente. La base de texte est structurée selon deux niveaux : la micro-structure et la macro-structure. Le niveau micro-structural renvoie à la cohérence locale, c'est-à-dire inter-propositionnelle de la représentation. Le niveau macro-structural réfère à la cohérence globale qui relie les propositions au thème principal de la séquence de texte traitée. Les propositions sont traitées et représentées selon leur ordre d'apparition dans le texte ; ainsi, leur organisation dans la base de textes respecte la façon dont l'auteur a structuré le texte (Macnamara, 1996A). Le micro-traitement consiste à relier les propositions en cours de traitement avec celles antérieurement traitées et stockées temporairement en mémoire de

travail. Le lien est établi si la nouvelle proposition partage un ou des arguments avec celles présentes en mémoire de travail. S'il n'y a pas de recouvrement d'arguments, un processus de recherche des propositions stockées en mémoire à long terme est mis en œuvre en vue d'un processus inférentiel qui assure l'établissement d'une relation entre les propositions. Cette phase aboutit à un graphe de cohérence. Parallèlement à ce micro-traitement, la base de textes fait l'objet d'un macro-traitement. Il consiste à relier les propositions au thème de la séquence lue et ceci par l'application de macro-règles qui suppriment, sélectionnent, généralisent des propositions du graphe de cohérence ou qui en construisent de nouvelles pour aboutir à la formation de macro-propositions. Selon Van Dijk (1977), l'application des macro-règles peut être facilitée par des indices textuels de sélection macro-propositionnelle tels que les titres, les résumés initiaux ; le lecteur interpréterait ces indices comme des indications fournies par l'auteur concernant la signification globale du texte. Il évoque également le rôle des marqueurs thématiques dont le lecteur se servirait pour justifier sa stratégie de sélection des propositions. Ce même auteur signale l'utilisation du gras ou de l'italique comme des dispositifs équivalents aux marqueurs thématiques mais il les qualifie de « moins subtils ». Enfin, parmi les indicateurs textuels, la structure conventionnelle du texte (celle du récit, par exemple) peut participer à la construction de la macro-structure. Le modèle de situation est une représentation cognitive des éléments du monde évoqués par le texte ; c'est le niveau de représentation le plus élevé. Il est construit par inférences, sur la base d'interactions entre le texte explicite et les connaissances antérieures du lecteur. D'après les auteurs, ce qui est crucial dans la compréhension du texte, c'est la rencontre entre les connaissances du lecteur et les informations dérivées du texte.

Dans ce modèle, les marques typographiques s'apparentent aux traits de surface du texte qui n'interviendraient que très tôt dans le traitement pour donner lieu à une représentation du texte de bas niveau et peu persistante en mémoire (le code de surface). De ce fait, leur fonction dans le modèle propositionnel de compréhension du texte serait réduite à une fonction quasi ornementale.

Peu de travaux testant l'implication potentielle des marques morpho-dispositionnelles dans la compréhension du texte ont été menés dans le but précis d'intégrer ces résultats dans un modèle plus large de la compréhension (Schmid, 2001), (Lorch, 1989). Pourtant, de nombreuses études menées à la suite des propositions théoriques de Kintsch se sont centrées sur les interactions entre les caractéristiques du texte et les caractéristiques du lecteur avec une attention particulière portée au rôle du niveau des connaissances en mémoire chez le lecteur (Kintsch, 1988), (Kintsch, 1993), (Kintsch, 1999), (Macnamara, 1996B). Or, il nous semble que l'effet de la morpho-

disposition sur le traitement cognitif du texte peut tout à fait s'étudier dans la perspective interactionniste du modèle. En effet, bien qu'il puisse être argué que les marques morpho-dispositionnelles ou plus précisément le contraste qu'elles entretiennent avec le reste du texte exercent des effets sur l'attention du lecteur et que ce sont ces effets qui médiatisent des effets sur des traitements de plus haut niveau (Lorch, 1989), il est nécessaire de considérer l'intervention des connaissances du lecteur pour rendre compte des effets de la morpho-disposition. En effet, pour déceler l'intention sous-jacente à l'utilisation des tels procédés typographiques ou dispositionnels de la part du scripteur et les interpréter de façon adéquate, le lecteur doit faire appel à la fois à ses connaissances sur les textes en général et sur le type de texte qu'il est en train de lire ainsi qu'à la représentation qu'il a du rédacteur et de ses objectifs de communication. Ainsi, lorsque le lecteur rencontre une configuration de marques indiquant un titre de rubrique (numérotation, augmentation de la taille des caractères et éventuellement grasse, disposition particulière dans la page), il ne suffit pas que les marques attirent son attention par rapport au reste du texte, il faut également que le lecteur mobilise ses connaissances conventionnelles sur le texte pour identifier que l'objet textuel est un titre de rubrique. Nous pensons donc qu'alors que

la plupart des modèles actuels de compréhension partent du principe que [la] représentation [mentale] du contenu est indépendante de la forme avec laquelle le texte est présenté [, et que cela] revient à nier le rôle des effets de mise en forme, d'enrichissements graphiques ou de procédés d'organisation dans la manière avec laquelle le lecteur construit cette représentation mentale (Baccino, 2004:169),

la morpho-disposition doit trouver sa place au sein d'un modèle de compréhension du texte (Lemarié, 2004), (Maurel, 2004).

Chapitre III. Du point de vue des linguistes

De manière générale, même lorsque la linguistique devient textuelle elle s'attaque le plus souvent à la description du texte à partir d'une définition « amodale » et « désincarnée » de celui-ci. Ainsi, les différents niveaux classiques de la linguistique (lexical, syntaxique, sémantique et pragmatique) ne prennent que peu en compte les aspects spécifiques au support du texte : les modèles s'abstraient considérablement de la modalité de présentation de l'information pour expliquer les structures pragmatiques et rhétoriques du discours/texte. Toutefois, cette remarque s'applique moins au discours oral depuis qu'ont émergé la notion de structure prosodique et l'étude des relations qu'elles entretiennent avec les autres niveaux linguistiques (années 1970-1980). En revanche, peu de tentatives ont été faites pour essayer de considérer, dans le champ circonscrit par les recherches en linguistique textuelle, l'analyse de procédés spécifiques à l'écrit, pourtant tout à fait comparables aux phénomènes intonatifs : ceux relatifs à la morpho-disposition des textes. Cette sous-section s'attachera en particulier à mettre en valeur les exceptions notables que nous avons pu relever dans la littérature.

III.1 Positionnement terminologique et taxonomie

La place du traitement graphique en linguistique souffre dès le départ d'une absence de consensus au niveau de la terminologie utilisée pour décrire le champs d'étude qu'elle circonscrit.

Pei parle au choix de graphémique ou graphonomie dont il donne la définition générale

d'étude des systèmes d'écriture et de leur relation avec les systèmes linguistiques. (Pei, 1954)

Hartman définit sous les termes graphémique ou graphologie

l'étude des signes graphiques utilisés dans un langage particulier (Hartman, 1972),

et les termes de graphétique, graphonomie, grammatologie ou art graphique comme

l'étude de la substance graphique et des formes des signes écrits sans considération pour un langage particulier ou un système d'écriture. (ibid.)

Nous trouvons également Crystal (1980) qui clarifie le terme de graphétique en incluant les propriétés du support écrit telles que la couleur, la taille de la police de caractère et l'interlignage ; et Derrida (1967) qui reprend Gelb (1963) pour définir l'étude des systèmes d'écriture en général, comme relevant de la grammatologie.

Nous ne garderons dans la suite que les termes de graphologie et de graphétique tels qu'ils sont définies par Crystal, et repris par Waller :

au niveau graphologique, on met l'accent sur les contrastes qui peuvent être fait dans le système linguistique, plutôt que le système lui même, qui est étudié au niveau graphétique (Crystal, 1969).

C'est à dire que

prises individuellement, les techniques visuelles telles que la conception de la forme des lettres, les symboles, [...] doivent être vues comme graphétiques ; mais quand elles sont utilisées ensemble pour structurer un texte entier, nous voyons un système graphologique au travail (Waller, 1987).

La graphétique, au contraire de la graphologie, ne contribue pas à notre compréhension de la manière dont les procédés graphologiques participent à l'inscription du raisonnement du texte (indentation, capitalisation, ponctuation, disposition, ...). Remarquons toutefois que le choix d'une police de caractère par exemple peut être de nature graphétique ; mais s'il s'agit d'un poème dans une édition ou un livre choix était proposé à ce niveau, la sélection d'une police de caractère particulière devient de nature graphologique.

De plus, les trois principaux types de variations morphologiques habituellement définis en linguistique façonnent le mot dans sa forme en opérant des flexions interne ou externe, en le combinant avec d'autres mots réservés ou encore en l'agglutinant à des affixes particuliers. Dans ce cadre nous pourrions considérer la typographie des textes comme relevant d'une dérivation morphologique se réalisant non pas au niveau du mot mais en deçà ou au delà. De la même manière les marques dispositionnelles (blancs horizontaux ou verticaux) peuvent être traitées comme combinant des marques de ponctuation particulières (espace et retour à la ligne) et assimilées avec elles à des dérivations morphologiques de niveau textuel, en ce sens qu'elles participent également à lui donner une empreinte, un « relief » spécifique. La caractéristique commune à cet ensemble de marques est donc qu'elles ne se situent pas nécessairement à un des deux niveaux d'articulation souvent décrit dans la littérature (niveaux du phonème et du mot).

Ajoutées aux premières, ces considérations selon lesquelles

le texte moderne, prétendument réduit à la seule écriture, n'a pu se développer que par son inscription dans un espace graphique, espace noté implicite parce qu'il était visuel et non verbal (Laufer, 1979)

nous amènent à regrouper dans un premier temps cet ensemble de marques de nature typographique ou dispositionnelle comme relevant d'une inscription graphologique du

texte, c'est à dire comme autant de procédés permettant de (conce)voir une part non directement « articulable » de sa structure informationnelle.

La distinction de Twymann (1982) entre traits du langage graphique intrinsèques (gamme de caractères disponibles pour un système de composition donné ainsi que les variations stylistiques de ces caractères) et extrinsèques (manipulation de ces caractères), nous laisse présumer qu'un raffinement de la notion d'inscription graphologique du texte peut être proposé.

En effet, en plus du marquage des divisions dans un texte (point en fin de phrase, « blanc » entre les mots, retour à la ligne en fin de paragraphe, ...), cette inscription graphologique peut être également utilisée pour distinguer des segments spécifiques (sauts de ligne inattendu, mise en items d'une énumération, ...). Elle peut encore provenir non pas, à l'instar des deux précédentes, des possibilités intrinsèques offertes par le médium utilisé, mais des contraintes extrinsèques qu'il impose (retour à la ligne pour ne pas sortir du support, choix d'une police de caractère, ...). Enfin, l'utilisation de procédés morfo-dispositionnels peut être de nature purement stylistique ou esthétique.

Aussi, nous proposons la taxonomie suivante pour décrire plus finement l'inscription graphologique des textes :

- **inscription graphologique démarcative** pour définir l'ensemble des marques qui permettent de découper le texte en unités appréhensibles visuellement ;
- **inscription graphologique distinctive** pour définir l'ensemble des marques qui permettent de donner une granularité à la saillance de ces unités ;
- **inscription graphologique imposée** pour définir l'ensemble des marques qui sont inhérentes au support et donc subies par le rédacteur (et les autres acteurs textuels) ;
- **inscription graphologique stylistique** pour regrouper l'ensemble des marques liées au style du rédacteur, aux choix esthétiques, ...

Les différents choix du rédacteur dans la gestion de ces marques peuvent influencer sur l'efficacité communicationnelle du texte mais également son sens.

Bien sûr, aucun de ces quatre ensembles de marques pris deux à deux n'ont d'intersection vide : selon son contexte d'utilisation, une même marque peut couvrir l'une ou l'autre de ces catégories, voire plusieurs en même temps. Prenons l'exemple de deux retours à la ligne successifs entre deux paragraphes justifiés, alors que ces derniers ne sont pas séparés par une telle marque dans le reste du texte ; les marques de retour à la ligne participent à l'inscription graphologique du texte : le premier retour de

séparation contribue au processus d'inscription graphologique démarcative (puisque suffisant pour « ponctuer » le paragraphe) et le second au processus d'inscription graphologique distinctive (puisque donnant un statut spécifique à ce paragraphe) ; les retours en bout de ligne à l'intérieur du paragraphe seront quant à eux considérés comme des processus d'inscription graphologique imposée, tandis que leur justification pourra dénoter un simple choix esthétique⁵.

III.2 Justifications préliminaires

Twymann (*op. cit.*) a directement abordé la relation entre typographie et contenu langagier. Un de ses exemples significatifs montre la version imprimée d'un document et le manuscrit original dont celle-ci est issue, chacun datant de 1473. Une première différence est que le manuscrit présente des largeurs de caractère parfaitement ajustées pour réussir une justification du texte sur un double colonnage, alors que l'imprimeur a, lui, besoin d'une seule colonne. Mais la différence essentielle réside dans la signalisation graphique d'une division majeure du texte : la version manuscrite utilise deux couleurs d'encre différentes tandis que la version imprimée introduit plutôt de l'espace supplémentaire. En conclusion, la possibilité de telles équivalences met en évidence une part langagière de la communication graphique. Celle-ci devrait être étudiée aussi sérieusement que le langage verbal.

Avant de décrire les recherches en linguistique du texte pertinentes pour notre étude, nous avons voulu justifier également, à notre manière, cette relation entre inscription graphologique du texte et linguistique classique en évaluant dans quelle mesure il est possible de considérer les moyens matériels d'expression écrite à travers les différentes notions (extraites de Ducrot (1995) et Moeschler (1994)) proposées par ce domaine de recherche.

Pour cela, définissons en premier lieu les unités minimales qui composent les objets graphologiques manipulés par un rédacteur : il s'agira de l'ensemble composé des caractères alphanumériques et typographiques (composant le domaine d'étude de la graphétique selon notre choix terminologique). Alors que les études linguistiques proposent comme unité minimale le phonème/graphème, l'étude de l'écrit implique l'enrichissement de l'ensemble regroupant les graphèmes qui ont une correspondance phonémique directe, par ce nouveau matériel. Cette extension engendre la possibilité

⁵ Cet exemple est à visée exemplificatrice car une telle inscription graphologique distinctive entre deux paragraphes est en fait la « quantité » de blanc vertical ; les deux retours à la ligne sont en fait les marques qui permettent de percevoir l'inscription (il aurait aussi bien pu s'agir d'un interlignage plus fort).

d'exploiter une nouvelle combinatoire. Comment, à partir de là, peut-on utiliser les concepts de mot, de phrase ou d'énoncé ?

De nouveaux objets graphiques seront constitués par des combinaisons des unités minimales ainsi définies : un premier niveau supérieur de composition sera celui des mots graphologiques ; il comprendra toutes les séquences de caractères alphabétiques auxquelles sont agrégées d'autres marques typo-dispositionnelles, au minimum une espace :

le blanc seul institue un véritable équivalent alphabétique du zéro. Il est dépourvu de référent phonétique. Par là, il introduit dans le système alphabétique un second niveau d'abstraction, qui ne renvoie plus à la parole (Laufer, 1979).

Ces mots graphologiques pourront à leur tour être combinés pour constituer des objets plus complexes et ainsi de suite jusqu'à produire l'objet graphologique de plus haut niveau : le texte inscrit. Les marques utilisées à ces différents niveaux de composition graphique sont celles définies dans la sous-section précédente comme relevant de l'inscription graphologique démarcative.

La perception de la syntaxe est particulièrement affectée en terme de règles de gestion de la place et de l'ordre de ces nouvelles unités de traitement, spécifiques à l'écrit. Par exemple, l'interprétation de la ponctuation de la phrase graphologique peut s'avérer difficile : certains disent qu'une virgule a une valeur purement syntaxique, d'autres ne le contestent pas mais y ajoutent un rôle rythmique et respiratoire tandis que pour d'autres encore elle marque uniquement l'endroit où le lecteur peut reprendre son souffle ! Cependant on peut rapprocher de la notion de règle syntaxique certains principes qui régissent la place des objets graphologiques (une note de bas de page se trouve le plus souvent ... en bas de page) ou l'ordre dans lequel nous les observons habituellement (un titre de chapitre précède toujours ce dernier). Nous observons également des phénomènes identifiables à des réactions, en ce sens que les variations d'un objet graphologique peuvent imposer des variations à un autre (la numérotation des titres de différents niveaux par exemple).

A ce stade nous pouvons dire qu'une étude exhaustive de cette « syntaxe » des objets graphologiques fournirait quelques pistes de recherche pour repérer dans le document écrit les objets sans contrepartie phonémique directe et dont la transposition à l'oral semble problématique (bien que, comme nous le verrons, la grande variété d'objets et les nombreuses variations internes qu'ils peuvent subir rendent difficile d'envisager à court terme l'automatisation d'une telle tâche pour un texte « tout venant » !). De plus, transformer ces objets graphologiques en objets « oralisables » nécessite d'être capable

de leur affecter une sémantique précise et de comprendre les intentions communicationnelles qu'ils révèlent.

Affirmons tout d'abord que l'inscription graphologique participe effectivement à l'organisation sémantique des énoncés, puisque l'assemblage des marques en configurations particulières peut ajouter du sens à la chaîne alphabétique ciblée en l'encadrant, en s'y insérant ou en se superposant à elle. Un argument en faveur de cette hypothèse vient de la possibilité de montrer que la non prise en compte de ces marques est susceptible de produire des contresens :

(1) Ce sont les imbéciles, qui aiment les militaires qui boivent trop.

Seul les imbéciles aiment les militaires qui boivent trop :

$$\forall x \left(\exists y \left(\text{militaire}(y) \wedge \text{boit_trop}(y, \text{alcool}) \wedge \text{aime}(x, y) \right) \rightarrow \text{imbécile}(x) \right)$$

(2) Ce sont les imbéciles qui aiment les militaires, qui boivent trop.

Seuls les imbéciles qui aiment les militaires boivent trop :

$$\forall x \left(\text{boit_trop}(x, \text{alcool}) \rightarrow \exists y \left(\text{militaire}(y) \wedge \text{aime}(x, y) \wedge \text{imbécile}(x) \right) \right)$$

**(3) Ce sont les imbéciles
qui aiment les militaires,
qui boivent trop.**

Seuls les imbéciles aiment les militaires et boivent trop :

$$\forall x \left(\exists y \left(\text{militaire}(y) \wedge \text{aime}(x, y) \wedge \text{boit_trop}(x, \text{alcool}) \right) \rightarrow \text{imbécile}(x) \right)$$

Une simple différence ponctuationnelle entre (1) et (2) conduit à deux interprétations sémantiques différentes. De la même manière l'interprétation de (3) est modifiée par rapport à celle de (2) alors que ces deux objets ne diffèrent que par leur spatialisation sur le support.

La question que nous nous posons alors est de savoir si la description du sens véhiculé par les énoncés graphologiques peut être rapprochée de certains des principes d'organisation souvent utilisés en linguistique pour la description sémantique.

Si l'on considère la distinction entre sujet logique et prédicat, le sujet (objet auquel est attribuée une propriété) serait la chaîne alphabétique support de l'inscription graphologique et le prédicat (indicateur de la propriété) serait signifié par la configuration graphologique. Par exemple, la mise en parenthèse conférerait le statut de sujet logique au texte encadré et les deux symboles prédiqueraient ce sujet logique en

énonçant sa propriété d'être « supprimable sans altérer la cohérence globale ». En revanche, contrairement à beaucoup de langues, cette distinction n'est pas reflétée dans la structure syntaxique des objets graphologiques puisque aucun groupe grammatical spécifique (en particulier le sujet grammatical) n'est préféré comme objet de l'attribution.

Nous pouvons élargir cette analyse en nous référant aux concepts de relation et d'arguments. En effet, une configuration peut conférer une propriété à plusieurs objets en établissant entre eux une relation particulière. Par exemple une configuration de type « titre » met en correspondance une chaîne alphabétique (le contenu lexical du titre) et les autres objets graphologiques qui composent avec le titre, par exemple un chapitre. Dans le cas de cette section, nous pourrions ainsi dire que l'inscription graphologique « **X** **XXXXXXXXXX** » affirme la relation *titrer au niveau 2* du couple d'arguments (*De point de vue des linguistes, De manière générale [...] soulever l'oralisation des titres*). De la même manière l'inscription graphologique de la parenthèse dans la séquence de la troisième phrase de ce paragraphe : « **XXXXXXXXXXXXXXXXXXXXX (XXXXXXXXXXXXXXXXXX) XXXXX** », affirme la relation *préciser* du couple d'arguments (*le contenu lexical du titre, une chaîne alphabétique*). Précisons que dans ces deux cas, la relation est dissymétrique car « orientée » et que le premier argument est toujours la chaîne alphabétique sur laquelle s'applique l'inscription graphologique.

De la même manière que l'inscription graphologique démarcative permet de repérer les objets graphiques, c'est elle qui organise leur logique sémantique telle qu'elle vient d'être rapidement décrite.

Cependant, nous pouvons également considérer d'autres notions qui se fondent non pas sur le fonctionnement logique du langage mais sur une distinction, d'ordre psychologique, entre thème et propos. Classiquement (mais de manière simplificatrice), ce qui permet de déterminer le thème (ce dont parle le locuteur), c'est la question à laquelle l'énoncé répond. Le propos (ou rhème) est l'information que le locuteur entend apporter relativement à ce thème. Le thème est donc un objet extérieur auquel l'énoncé fait allusion. Il semble envisageable de rapprocher de ces notions une certaine part de l'inscription graphologique des textes en considérant qu'elle peut être utilisée pour exhiber au lecteur ce type de structure informationnelle, c'est à dire thématiser ou rhématiser des segments de texte. Les marques utilisées dans ce but sont celles définies dans la sous-section précédente comme relevant de l'inscription graphologique distinctive. En effet, elles distinguent un segment de texte en lui conférant le statut de réponse à une question que le rédacteur croit présente dans l'esprit de ses lecteurs. Écrire « **Max** est venu » peut généralement être interprété comme ayant pour fonction

d'inscrire que l'on répond à une question potentielle du type : « qui est venu ? », d'où une thématization de « est venu » et en conséquence une rhématisation de « Max » (on parle de *qui est venu* et on en dit que c'est *Max*). A l'inverse, la dislocation « Max, il est venu » thématise « Max » et donc rhématise « il est venu » (on parle de ce *qu'a fait Max* et on en dit qu'*il est venu*).

Une discussion s'impose ici autour de l'exemple du titre pour réduire ce que l'on pourrait considérer d'un premier abord comme paradoxal : d'un côté nous classons l'inscription graphologique « thématizante » et « rhématizante » comme distinctives, et d'un autre côté, l'inscription graphologique d'un titre peut n'être que démarcative même si sa fonction est d'exprimer le thème de la section qu'il précède. Il n'y a pas là, de notre point de vue, matière à paradoxe, puisque ce n'est pas l'inscription graphologique du titre qui thématise un segment de texte (elle ne fait en effet que le démarquer) mais plutôt le choix d'un contenu discursif particulier qui confère au titre le statut de thème ! Dans le cas de la mise en gras de notre exemple, c'est le choix de cette sorte de mise en relief qui confère à son support le statut de rhème, tandis que dans le cas du titre de la partie 1 de ce document c'est le choix du matériel lexical particulier « valeur et prise en compte de la morpho-disposition des textes » qui donne le statut de thème à ce titre. D'ailleurs un matériel lexical d'une autre nature peut engendrer des titres non thématiques : « Chapitre 1 » par exemple. De la même manière les titres d'œuvres littéraires peuvent ne pas être thématiques : le sujet du livre de Sartre est-il réellement « Huis clos » ou encore le rouge et le noir sont-ils de manière évidente les thèmes de Stendhal ? Bien que ces deux titres entretiennent avec le thème un rapport que l'on pourrait qualifier de métonymique pour le premier et de métaphorique pour le second, ils ne constituent pas littéralement les thèmes de ces livres.

Notons enfin qu'une autre distinction d'ordre psychologique est parfois faite dans l'étude de l'organisation sémantique des énoncés : celle du posé et du présupposé. Le présupposé d'un énoncé diffère du posé de ce même énoncé en ce sens :

- qu'il est maintenu lorsque l'énoncé est nié ou l'objet d'une interrogation ;
- que sa prise en considération est secondaire lors de l'enchaînement avec un autre énoncé ;
- qu'il est donné comme allant de soi et donc ne peut être mis en doute.

Existe-t-il une part de l'inscription graphologique qui implique du présupposé ? Peut-être s'agit-il de celle qui a une fonction d'ordonnement : un titre de type « 2.2 Xxxxx » présuppose l'existence passée d'un « 1. Xxxxx » et d'un « 2.1 Xxxxx », une note « ² » présuppose l'existence passée d'une note « ¹ ». (remarquons en revanche que

si le titre « 1. Xxxx » présuppose bien l'existence future d'un titre « 2. Xxxxx », une note « ¹ » ne présuppose pas l'existence future d'une note « ² » !). Dans le même ordre d'idées, le choix dans une structure énumérative d'une notation alphabétique dans un certain format pour marquer le premier item, impose au rédacteur, donc permet au lecteur de présupposer, à la fois une notation de même nature pour tous les autres items et l'utilisation de l'ordre alphabétique (ceci même s'il n'y a pas d'ordre logique apparent dans la présentation des items). Un autre exemple pourrait être celui des renvois aux références bibliographiques dans un texte scientifique : la présence de marques spécifiques du type « [Xxxxx, XXXX] » présuppose l'existence d'une référence complète généralement en fin de texte (tandis que l'absence de ces marques : « Xxxxx a dit en XXXX que ... », ne l'impose pas !).

Après cette tentative rapide de rapprochement entre l'inscription graphologique des textes et les concepts traditionnels qui ont traités à l'organisation syntaxique et sémantique des énoncés, il faut encore s'intéresser aux relations d'une part entre les énoncés et le texte qu'ils composent et d'autre part entre les énoncés et le monde extérieur au texte. Il s'agit pour cela de s'ouvrir à la notion d'énonciation, c'est à dire aux conditions de production d'un énoncé, et à celle de pragmatique, c'est à dire aux facteurs des échanges linguistiques entre interlocuteurs. En effet, les objets graphologiques de niveau supérieur participent à l'énonciation en ce sens qu'ils sont des réalisations signifiantes particulières, proposées par un sujet « écrivain » déterminé, qui se situe à un endroit et à un moment donné. Aussi, de la même manière qu'il est possible d'attribuer à une configuration strictement phonologique (mot ou phrase « classique ») une signification qui fasse référence à un type d'événement énonciatif, il en est de même pour certains des objets inscrits sur un support, mais de par leur nature graphologique.

Les phénomènes d'organisation de l'information qui ont été abordés plus haut autour des notions de thème et propos sont un des aspects que l'énonciation met en jeu ; en effet, ce type de structure suppose un choix du rédacteur en fonction de la façon dont il se représente les attentes de son interlocuteur. A la signification de ces marques, que nous avons catégorisées comme distinctives, peuvent s'ajouter d'autres valeurs énonciatives. Par exemple, la mise en gras peut revêtir une valeur focalisatrice et insister sur l'importance en contexte du segment distingué même si l'on ne suppose pas une interrogation du lecteur à son sujet. Ou encore il peut s'agir d'une emphase pour attirer l'attention sur un élément à des fins de contraste, d'opposition.

Par exemple, « **Max** était là » peut être interprété en fonction du contexte (ou du co-texte) comme :

- la réponse à la question potentielle « qui est venu ? » que le lecteur est supposé se poser (Cf. *supra*) ;
 - *je suis allé à la soirée. Max était là. Pierre également.*
- une insistance sur l'importance de cette information et réalisant ainsi un focus ;
 - *je suis allé à la soirée. Tout le monde était là. Max était là.*
- l'établissement, de manière emphatique, d'une insistance non pas sur une partie de l'information elle-même, mais sur une opposition, un contraste (dans l'exemple, entre les efforts faits par « Pierre » et « Max » pour aller à la soirée).
 - *Pierre n'est pas venu à la soirée. Max était là. Lui.*

Malgré leur proximité et, souvent, la difficulté de les distinguer, il nous semble que chacune de ces interprétations pourrait potentiellement conduire à une oralisation différente de cet énoncé.

Remarquons que nous pouvons rapprocher le dernier exemple de la notion de modalisation linguistique en ce sens qu'il exprime le jugement du rédacteur sur l'événement qu'il relate, son attitude vis-à-vis de celui-ci. Cet aspect de l'énonciation est décrit en linguistique à travers les notions de *dictum*, contenu proprement dit de la proposition, et de *modus*, qui traduit le jugement que le locuteur porte sur le *dictum*. Dans la littérature sont envisagés pour le *modus* des moyens d'expression variés mais qui se situent essentiellement à des niveaux grammaticaux, propositionnels, ou verbaux. Nous pensons qu'une certaine part de l'inscription graphologique distinctive (mais également à l'oral une part de l'intonation, des mimiques ou des gestes) peut être considérée comme un moyen supplémentaire pour le rédacteur d'exhiber une attitude particulière vis-à-vis du contenu sémantique de son énoncé.

La section précédente avait pour objectif de montrer que l'étude de la mise en forme des textes n'était pas vouée à une marginalisation systématique puisque beaucoup de ses dimensions peuvent être rapprochées des notions classiques de la linguistique.

Bien que cette discipline ne s'est que rarement attachée à tenir compte des implications liées au choix du support physique sur lequel la chaîne linguistique s'inscrit, nous décrirons dans la suite de cette sous-section les approches de quelques précurseurs pour l'étude des relations entre typographie et langage, puis trois études linguistiques qui se centrent principalement sur un sous-ensemble des phénomènes morpho-dispositionnels – la ponctuation (Catach, 1994), (Nunberg, 1990)

(Power, 2003) – et, enfin, nous nous focaliserons sur une des seules propositions, à notre connaissance, qui revêt un pouvoir explicatif plus global de la sémantico-pragmatique de l'inscription graphologique du texte, à travers la notion d'architecture textuelle (Virbel, 1989), (Pascual, 1996).

III.3 Les précurseurs

L'examen du style anglais dans Crystal (1969) se répartit pour moitié sur l'étude de l'oral et pour moitié sur celle de l'écrit. Ces recherches proposent 5 niveaux de description dont trois de haut niveau (grammaire, syntaxe et sémantique) et deux de bas niveau (phonétique/graphétique et phonologie/graphologie). Mais dans la pratique ces deux derniers niveaux ont été principalement étudiés dans leur versant relatif à la parole, en particulier pour se doter d'un système de transcription des traits saillants de l'oral (fréquence fondamentale, rythme, emphase, ...) qui permette une véritable analyse. Pour l'écrit peu d'efforts ont été fournis pour distinguer les traits linguistiques des traits non linguistiques. Ceci s'explique par l'apparente simplicité de l'écrit en terme de segmentation qui laisse présumer que la graphologie serait théoriquement moins intéressante que la phonologie qui, par contraste, admet *a priori* plusieurs niveaux supra-segmentaux à la fois relativement indépendants les uns des autres et partageant des frontières segmentales complexes. Nous pouvons relever malgré tout une remarque intéressante qui procède d'un découpage des unités de l'écrit en 14 niveaux qui sont mis en relation avec les unités correspondantes des niveaux phonologique, grammatical et sémantique. Les résultats qui ressortent de ces travaux peuvent se résumer de la manière suivante :

- le statut de la ligne graphologique est le seul qui soit entièrement un artefact du processus d'écriture et cet objet constitue un pivot particulier entre les niveaux qui lui sont inférieurs et ceux qui lui sont supérieurs ;
- les bas niveaux, situés sous celui de la ligne, présentent tous des liens triviaux entre graphologie, phonologie et grammaire : peu de choses peuvent donc être gagnées par le choix d'études de données écrites plutôt qu'orales ;
- les niveaux au dessus de celui de la ligne n'indiquent que des relations vagues avec la structure de l'information : pour ces chercheurs, la graphologie ne concerne plus de manière centrale le linguiste, mais plutôt le contexte, le thème de la communication et les facteurs de la rhétorique que la linguistique « pure » a été longtemps heureuse d'ignorer.

Vachek (1948/1967, 1959, 1973) a toujours prôné l'autonomie du langage écrit par rapport au langage oral. Il se définit par son appartenance aux linguistes de Prague et partage donc avec eux une perspective fonctionnaliste dans le sens où les caractéristiques du langage proviennent de ses fonctions dans la communauté des usagers de la langue. Par exemple étant donné que poser des questions, faire des constatations et donner des ordres sont des utilisations universelles du langage, la grammaire se dote de formes interrogatives, déclaratives et impératives dans la plupart des langues. On ne peut non plus s'attendre à trouver le mot *neige* dans le vocabulaire d'un habitant des tropiques. Il rejette donc par là la distinction saussurienne entre langue et parole, et précise que

La norme parlée du langage est un système d'éléments langagiers qui peuvent se manifester phonétiquement et dont la fonction est de réagir à un stimulus donné (urgent) de manière dynamique, c'est-à-dire d'une manière prompte et immédiate, en exprimant non seulement l'aspect purement communicatif mais aussi émotionnel de la démarche du locuteur.

La norme écrite du langage est un système d'éléments du langage qui peuvent se manifester graphiquement et dont la fonction est de réagir à un stimulus donné (non urgent) de manière statique, c'est-à-dire d'une manière conservable et facilement observable, en se concentrant particulièrement sur l'aspect purement communicatif de la démarche du locuteur. (Vachek 1973 :15-16)

La conclusion de cette approche est que l'on ne doit pas espérer trouver un parallèle exact entre graphologie et phonologie. Par exemple, la nature statique, observable de l'écrit suggère un rôle pour la typographie relativement distinct du rôle émotionnel et immédiat de certaines caractéristiques de la parole.

Nous tirons de Waller (1987) une autre distinction intéressante proposée par Vachek entre le marqué et le non marqué. Le marquage d'un élément est dû à un ajout et des fonctions asymétriques : *chat* est non marqué (on peu parler sans choquer de chat mâle ou de chat femelle) alors que *chatte* l'est (on ne peut parler de chatte mâle sans être contradictoire ou de chatte femelle sans entrer dans la sphère de la tautologie). A partir de là, Vachek définit la norme écrite comme l'élément marqué d'une opposition dont l'élément non marqué serait la forme parlée correspondante : Vachek (1979) fournit une liste des fonctions pour lesquelles les ensembles marqués de symboles graphiques (comme l'italique) peuvent être utilisés pour distinguer les caractéristiques du texte nécessitant de l'emphase ou une stylisation de la norme non marquée (comme la police de caractère Roman). Il montre d'ailleurs que l'impression de préfaces en italiques dans la tradition allemande ou tchèque oblige à marquer l'emphase par du droit, entraînant un

effet peu convaincant pour le lecteur : dans l'opposition entre caractère italique et caractère droit, *l'italique* peut être reconnu comme du « non-droit » mais *le droit n'est reconnu comme du non italique que plus difficilement.*

La réponse définitive de Vachek aux tenants de la primauté de la parole qui argumentent que le langage parlé est universel alors que seulement une partie des communautés langagières connaissent l'écriture est que

le but vers lequel le développement du langage a été dirigé dans toute communauté est la communication linguale la plus hautement efficace possible et le développement maximal de la sphère fonctionnelle. [Et en outre que] les optimaux du langage ne devraient pas moins primer par leur importance que les universaux du langage. (Vachek, 1973 :17)

Bernhart (1985) a essayé d'étudier un aspect de la constitution des textes, auquel la taxonomie des relations cohésives (Halliday, 1976) ne répondait pas bien, en listant un large échantillon de techniques utilisées par les rédacteurs pour lier les composants textuels : comment, dans l'usage réel, un auditeur ou un rédacteur choisit-il ces techniques cohésives ? Pour cela il a comparé plusieurs textes sur le même sujet mais écrits dans des buts et des intentions différents et ses conclusions précisent :

dans mon essai pour expliquer les patterns de stratégie rhétorique et les réalisations conséquentes de cohésion en considérant le contexte de la situation, il est vite devenu apparent que la conception graphique doit figurer de manière proéminente dans l'analyse des patterns de structuration cohésive. (Bernhart, 1985 :18)

Waller (1987) fait l'hypothèse d'un modèle des textes organisés typographiquement qui fournit les bases pour une conversation entre un rédacteur et un lecteur, dans laquelle le contrôle s'échange entre les participants. Cette idée d'un dialogue textuel se trouve déjà chez nombre de chercheurs issus de disciplines variées telles que le dialogue (Gumperz, 1982), (Coulthard, 1985), la linguistique (Hoey, 1983), la sémiotique (Eco, 1981), la psychologie de la cognition (Wright, 1978), (Nystrand, 1986). L'idée est que les rédacteurs s'adressent à un lecteur imaginaire dont les caractéristiques et les attitudes seront, si l'habileté du rédacteur est suffisante, facilement endossés par le lecteur réel. Le lecteur imaginaire a des questions et des attentes qui doivent être traitées par le rédacteur. Le modèle veut mettre en lumière les contraintes fonctionnelles qui gouvernent le rôle du typographe dans la communication textuelle : trois structures sous-jacentes sont associées à différents genres de texte (Figure 3).

Le modèle se définit donc par sa nature dialogique qui sépare le texte du rédacteur de celui du lecteur. Chacun d'eux présente la même structure artefactuelle mais le rédacteur organise sa structure thématique tout en proposant à un lecteur imaginaire une

structure d'accès aux informations ; tandis que le lecteur organise son accès à l'information en parcourant et en anticipant la structure thématique d'un rédacteur imaginaire du texte. La combinaison de ces différentes structures permet selon l'auteur de dégager une définition des genres typographiques des textes. En pratique les trois premières catégories théoriques se combinent et peuvent même coïncider (un nouveau chapitre comme point d'accès et frontière thématique par exemple) ; c'est à la quatrième que revient le rôle de la gestion de ces combinaisons et la principale raison de ce découpage est de servir de base à la description du genre des textes organisés typographiquement. Quand les technologies de présentation de l'information changent, il devient nécessaire de séparer alors ces trois catégories des impératifs fonctionnels et de les rassembler pour tenir compte des nouvelles contraintes subies et possibilités offertes.

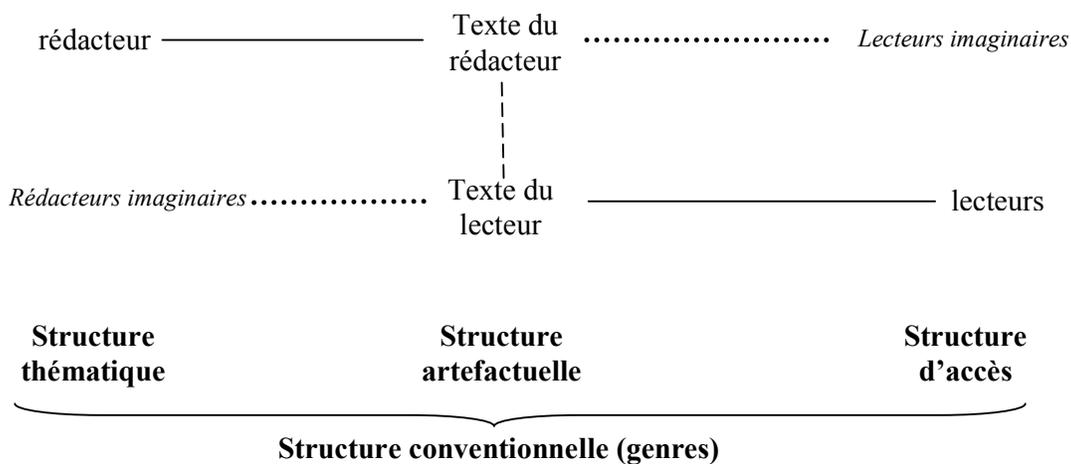


Figure 3 : Modèle du genre textuel (Waller, 1987)

Toujours pour l'auteur, en ce qui concerne la structure thématique, il existe un potentiel considérable pour son renforcement par l'arrangement typographique. Celui-ci se baserait pour cela sur un ensemble d'instances de métaphores visuelles utilisées en rhétorique ou en sémantique : les notions rhétoriques d'étape, d'empilement, de chaînage ou de pondération (Nash, 1980) pourront être signalées aussi bien par des adjoints conjonctifs (par exemple, respectivement : premièrement, en outre, donc, néanmoins) que par la typographie ; celle-ci pourra également mettre en valeur des relations sémantiques dans le vocabulaire (Lyons, 1977) telles que les contrastes d'opposition binaire (chaud/froid, homme/femme, haut/bas, etc.) et d'ensembles non binaires (mauvais...moyen...bon, caporal...sergent...capitaine, etc.) ou encore les hiérarchies (hyperonymie, synecdoque, ...).

La structure artefactuelle jouera quant à elle sur la gestion des ruptures, qu'elles soient arbitraires ou signifiantes (nouveau paragraphe, nouveau thème, liste, tableaux, ...).

Un exemple de la variation rédactionnelle des structures d'accès en fonction du lecteur imaginé est montrée par Waller (1987) à travers la multitude des formats de la Bible (13 formats depuis le traditionnel jusqu'à celui pour enfants avec images et notes explicatives ; en passant par l'édition en lettres rouges dans laquelle des mots sont mis en valeur pour une lecture dévotionnelle et un apprentissage par cœur, la version de poche qui présente l'intérêt d'être assez peu coûteuse pour être distribuée gratuitement, l'allégée qui ne sert qu'à des occasions spécifiques, la version informatique qui permet de faciliter des recherches ou encore le format polyglotte qui propose la mise en regard de plusieurs traductions...).

Des études plus récentes menées par des linguistes se sont centrées sur le phénomène typographique tant au niveau historique que dans la perspective d'en exhiber les aspects littéraires et linguistiques. C'est sur ce dernier point que nous focaliserons notre propos dans les sous-sections suivantes, en isolant certaines des recherches qui nous semblent les plus abouties pour nous éclairer sur le rôle potentiel de l'inscription graphologique dans ces diverses tentatives de définition d'une linguistique de l'écrit.

III.4 L'approche de Nina Catach

Du point de vue de son histoire, l'évolution de la ponctuation moderne depuis le XVI^{ème} siècle peut être décrite comme étant au départ

[...] un système basé sur l'oral dont le but est d'aider la lecture à haute voix, qui donne une ponctuation sobre et claire. Puis, avec le développement de l'imprimerie, la domination autoritaire des imprimeurs et l'accroissement du nombre des lecteurs, une ponctuation très grammaticale, basée sur la logique, qui aboutit à surcharger les textes d'un grand nombre de signes. Au XX^{ème} siècle, les écrivains, prenant le pas sur les imprimeurs, donnent à la ponctuation une grande diversité et originalité, dans le but de faire passer dans le texte le mouvement des sentiments. La ponctuation tend à s'intégrer à tous les moyens typographiques et graphiques pour donner du texte (Lorenceau, 1980).

Les travaux de Catach (1994) sur la ponctuation montrent comment cette intégration, prise au niveau du texte, contribue à regrouper les phrases qui le composent en une « unité de sens total ».

Une distinction primordiale concerne la ponctuation générale (ou constructive) par rapport à la ponctuation séquentielle. La première concerne le dispositif permettant la construction d'unité de plus en plus complexe depuis la syllabe jusqu'au texte en passant par une unité centrale : la proposition. La ponctuation séquentielle est dépendante de la phrase dans laquelle elle se trouve obligatoirement de manière intégrée (virgules doubles, parenthèses, tirets, ...) ou non intégrée (discours rapporté par exemple).

Les signes constituant la ponctuation générale peuvent être considérés comme participant à une ponctuation de premier ordre (supérieure à la phrase), de second ordre (au niveau de la phrase) et de troisième ordre (inférieure à la phrase). Selon ce paramètre, l'auteur attribue ainsi aux différents signes un indicateur de force :

- le troisième ordre regroupe les forces 1 à 3 (respectivement la ponctuation au niveau de parties de mots, de mots, de parties de proposition) ;
- le deuxième ordre regroupe les forces 4 et 5 (ponctuation de la proposition et des phrases complexes) ;
- le premier ordre regroupe les forces 6 et 7 (ponctuation du paragraphe et de la période qui constitue une « unité de pensée totale »).

Une des distinctions les plus intéressantes pour notre propos affirme que la ponctuation d'une part permet une lecture orale en indiquant ou en complétant les informations de la langue orale (assertion, ironie, insistance, admiration, ...) et d'autre part autorise une lecture silencieuse où **seuls les dessins des signes [font] sens (Catach, 1994 :53)**.

Nous inférons déjà que la transposition automatique à l'oral se doit au mieux d'interpréter sans cesse la ponctuation ou à minima de lui choisir une interprétation par défaut probablement en fonction du genre du texte.

De manière générale les fonctions essentielles de la ponctuation sont pour la linguiste d'indiquer : la segmentation de l'énoncé, les pauses orales et écrites – majeures et mineures –, l'intonation, les modalités, la construction et les rapports des parties, l'opposition thème/propos, les rapports inter-locuteurs, l'opposition des sens partiels et du sens total, des oppositions grammaticales spécifiques, la clarté de l'énoncé.

De plus, la ponctuation est vue, à l'instar de la prosodie dans de nombreux travaux de ce domaine, comme suprasegmentale en ce sens que les signes dit doubles – qui peuvent être couplés (la majuscule et le point), inversés (certains guillemets) ou symétriques (parenthèses, crochets,...) – portent toujours un segment syntaxique en délimitant ses

bornes ; et en ce qui concerne les signes simples il n'est que culturel qu'ils soient placés dans la chaîne écrite à la manière des caractères alphabétiques puisque certaines langues les positionnent au dessus des segments qu'ils portent. Remarquons que la virgule doit être considérée à part – en atteste Drillon (1991) qui met en évidence 139 usages de ce signe de ponctuation – bien que celle-ci puisse être également à « visée bilatérale », d'une des manières décrites précédemment (on parle par exemple de virgule double pour isoler certaines appositions).

Enfin, et surtout, la discussion la plus intéressante de ses travaux sur la ponctuation pour notre étude se déploie autour de deux questions :

**a/ Peut-on, dans l'optique de rapports relativement autonomes entre l'oral et l'écrit, établir sur une base nouvelle les fonctions des signes écrits en correspondance avec les unités d'intonation ainsi dégagées ?
b/ Que deviennent ces fonctions en cas d'une lecture coupée de l'oral ?
(Catach, 94 :100)**

La pertinence de ces deux questions est justifiée par une autonomie entre l'écrit et l'oral soutenue avec elle par de nombreux spécialistes de la ponctuation. En atteste :

Que nous sachions, l'œil ne respire pas. Et s'il fait des pauses, c'est que son champs de vision couvre quelques centimètres à la fois, et qu'il se déplace ainsi, de bloc en bloc – sans tenir compte particulièrement de la position des virgules ou des points. (Drillon, 1991 :100)

Ou encore :

L'interprétation de la ponctuation comme moyen de marquer à l'écrit l'intonation de la phrase paraît insuffisante [...]. L'oral et l'écrit représentent deux systèmes dont les éléments correspondent par l'intermédiaire du signifié [...]. [ils] différent quant aux moyens de réaliser cette fonction. (Védénina, 1989)

Pour la seconde question l'auteur nous précise que même si certains arguent qu'une lecture, même silencieuse, est toujours accompagnée d'un phénomène dit de kinesthésie, c'est-à-dire de sensations internes qui font que l'oral reste premier, il est sûr qu'au moins dans le cas de la relecture, l'oral s'efface au profit d'une

élaboration seconde et quasi indépendante de corrections (et de ponctuations). (Catach, 1994 :98)

D'après Catach, une réponse partielle à ces questions doit être cherchée dans le fait que si les fonctions intonatives les plus complexes ne trouvent pas de contrepartie écrite, il n'en est pas de même pour les aspects prosodiques les plus linguistiques. Elle

conclut alors que grâce à un tel tri d'ordre métalinguistique que propose le processus d'écriture,

on aurait là [...] des traces précieuses de phénomènes qui nous restent en grande partie cachés. (Catach, 1994 :101)

Elle distingue enfin à partir de là les signes internes au texte des phénomènes de mise en page, car les premiers seraient susceptibles d'une approche linguistique à travers la notion de ponctème, dérivée de celle d'intonème à l'oral.

Dans sa vision globale du texte ce concept permet de saisir la ponctuation de façon non linéaire, dans sa dimension suprasegmentale. De la même manière que Delattre (1967) distingue déjà dix intonations de base pour le français (question, continuation majeure et mineure, implication, écho, parenthèse, finalité, interrogation, commandement, exclamation) dont il donne pour chacune la configuration idéale de l'évolution de la fréquence fondamentale lors de l'énonciation d'une phrase (les intonèmes), il s'agirait d'y faire correspondre des configurations ponctuationnelles (des ponctèmes) et d'étudier leur récurrence. Ce sont ces ponctèmes qui permettront par exemple d'admettre, à l'instar des intonèmes, l'intégration d'un mot isolé au rang de phrase. D'ailleurs, c'est déjà à partir de cette dernière constatation que Nunberg a proposé une autre approche de la phrase inscrite qui voulait résoudre ce type de « viol légal » des règles syntaxiques établies par la grammaire classique.

III.5 Les travaux de Nunberg, Power et Scott

« Il est entré dans le bureau. Désastre. Le coffre-fort était ouvert et l'argent avait disparu »

Nunberg (1990) remarque que l'étude classique de la structure syntaxique d'un texte en développant des grammaires (dites lexicales dans sa terminologie) composées de règles telles que $P \rightarrow SN + SV$, fait peu de cas du grand nombre de cas à l'écrit où ces règles ne sont pas respectés. Ainsi, dans la séquence de l'exemple du début de cette sous-section, la seconde phrase n'est pas syntaxiquement correcte. A partir de l'observation de deux sortes de phrases, selon qu'elles respectent ou non les règles d'une grammaire syntaxique, il propose un second type de description grammaticale dédié aux catégories textuelles particulières qui forment la structure du texte (à ne pas confondre avec la structure du discours d'un texte).

Les quatre catégories textuelles exhibées dans ces travaux ont des propriétés structurales abstraites en ce sens qu'elles peuvent se réaliser concrètement de plusieurs manières en fonction du contexte ou de conventions particulières. Sont distingués :

- la phrase textuelle (S_t pour *text-sentence*) : elle commence par une capitale et se termine par un point ;
- la proposition textuelle (C_t pour *text-clause*) : elle se termine par un point-virgule ;
- le paragraphe : il commence par une nouvelle ligne et une tabulation ;
- la section (sans précision supplémentaire sur ses caractéristiques).

La grammaire de texte permet alors de formuler une règle du type $S_t \rightarrow C_t^+$ exprimant qu'une phrase textuelle contient une ou plusieurs proposition textuelle.

Ainsi notre exemple est formé de trois phrases textuelles elles-mêmes composées respectivement de 1, 0 et 2 phrases syntaxiques (voire de 3 phrases syntaxiques dans la dernière phrase textuelle si le tout est considéré en plus des parties). Si la dernière phrase de cet exemple est transformée en « *Le coffre fort était ouvert ; l'argent avait disparu.* » nous aurions une phrase textuelle composée de deux propositions textuelles (le point-virgule virtuel de la seconde proposition textuelle étant, lors de la réalisation concrète, écrasé par la ponctuation de l'unité englobante : le point de la phrase textuelle).

Cette approche a été reprise et développée par Donia Scott pour générer plusieurs versions d'un même message dans différents styles (Power, 2003A ; Power, 2003B ; Scott, 2003). Dans un premier temps, une hiérarchie à six niveaux est proposée pour regrouper les différentes unités documentaires :

L_0 : syntagme textuel

L_1 : proposition textuelle

L_2 : phrase textuelle

L_3 : paragraphe

L_4 : section

L_5 : chapitre

Une règle : $L_N \rightarrow L_{N-1}^+$ ($N > 0$) exprime que les niveaux L_1 à L_5 sont composés d'un ou plusieurs sous-niveaux inférieurs, tous de même rang. Il est précisé qu'une description complète devrait tenir compte d'autres configurations (titres, énumérations/listes marquées par des puces, notes de bas de page, schémas, citations).

Pour la gestion du problème de l'indentation des unités documentaires (les structures indentées peuvent violer les règles précédentes : par exemple la structure parenthétique

que vous êtes en train de parcourir est de niveau L3 mais est contenue dans une unité de niveau L1), il est proposé l'ajout d'un trait supplémentaire dans leur définition : l'indentation comprise entre I_0 et I_{max} (du non indenté à l'indentation la plus profonde de l'auteur). Il est à noter que l'indentation prise en compte n'est pas graphique mais logique (réalisation concrète par de l'italique aussi bien que par une tabulation par exemple). La conséquence est que deux règles

Un constituant sans indentations devra respecter la règle : $[LN, IM] \rightarrow [LN-1, IM]^+$; tandis qu'un constituant avec indentation respectera la règle : $[LA, IM] \rightarrow [LB, IM+1]^+$. Une contrainte sur I_{max} pourra être rajoutée en imposant par exemple $B < 4$.

se substituent à la précédente.

Cette dernière phrase serait ainsi représentée comme indiquée dans la Figure 4 ci-dessous.

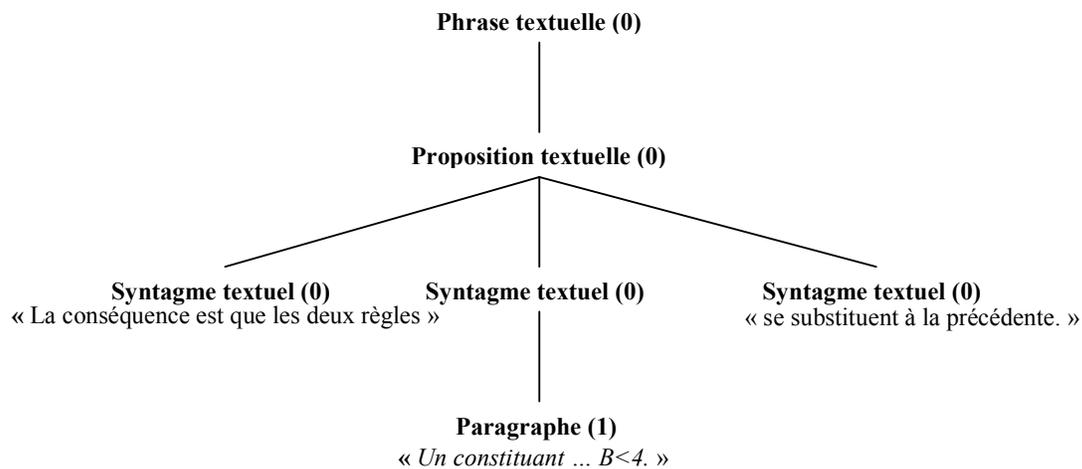


Figure 4 : Structure indentée d'un document

Ainsi :

- un niveau est composé d'un ou plusieurs sous-niveaux inférieurs et tous de même rang s'il sont avec lui de même niveau d'indentation.
- un niveau est composé d'un ou plusieurs constituants de même rang mais d'un niveau hiérarchique quelconque, à condition que ces derniers aient un niveau d'indentation supérieur.

Nous remarquerons que cette représentation ne prend pas en compte les intentions du rédacteur. Travaillant dans le domaine de la génération de texte, l'objectif était pour ces

auteurs de proposer un système qui permette de construire un maximum de messages dont la forme logique était acceptable et filtrer les structures de texte *a priori* improbables. Ce n'est donc pas à ce niveau que les intentions du rédacteur sont prises en compte pour trier, parmi toutes les présentations possibles d'un même contenu textuel, quelle présentation est la plus cohérente avec le message du texte. En fait, son système dispose en entrée de la représentation de la structure rhétorique du document à générer et propose en sortie toutes les structures de documents qui la réalisent correctement (avec certaines hypothèses simplificatrices sur la composition de la structure du document et les connecteurs du discours disponibles dans le lexique). Or, notre hypothèse est justement que nous devons considérer le sens du texte et les intentions du rédacteur rendus explicites par la mise en forme elle-même, c'est à dire de manière indissociable des aspects lexico-syntaxiques du contenu alphabétique impliqué. Nous avons donc besoin d'un modèle de représentation de la structure du texte qui repose sur l'analyse de cette relation lexico-syntaxe/structure logique et non d'un système qui sépare systématiquement le traitement syntaxique classique et le formatage du document. Un tel modèle permettrait potentiellement de rendre « articulable » lexico-syntaxiquement la structure logique du document. A notre connaissance, le seul modèle qui permette d'envisager une telle transformation du texte est décrit dans la sous-section suivante.

III.6 Le Modèle d'Architecture Textuelle

L'architecture de texte est une composante abstraite du texte et prend source dans la notion de **Mise en Forme Matérielle** (abréviation usuelle : MFM – Virbel, 1989) et dans l'hypothèse de l'existence d'une équivalence fonctionnelle entre des phénomènes typographiques, dispositionnels et lexico-syntaxiques. La MFM est un sous-ensemble de propriétés morfo-dispositionnelles du texte, propriétés possédant des équivalents langagiers. Ainsi, les constituants des textes, les objets textuels (définition, énumérations, parties, titres, etc.), sont perceptibles par le jeu de contrastes de la MFM et l'architecture de texte est l'ensemble des objets textuels et les propriétés qu'ils entretiennent entre eux.

L'approche générale de l'architecture textuelle est sous-tendue par les travaux de Z. Harris sur le métalangage (Harris, 1971, 1990). Harris pose que la langue peut être décrite par elle-même, c'est à dire que l'on peut exhiber des phrases à portée métatextuelle explicitant le fonctionnement de la langue et obéissant aux mêmes lois de constructions grammaticales que les phrases de la langue. L'ensemble de ces phrases forme le **métalangage**. Par exemple, la phrase « Max est sujet de *Max mange une*

pomme » est une phrase du métalangage. Harris pose qu'en règle générale ce métalangage est réduit mais que cette réduction laisse des traces dans la langue.

Ce principe a été appliqué aux propriétés morpho-dispositionnelles du texte : ces propriétés peuvent être décrites à l'aide de phrases métatextuelles (par exemple : « je segmente mon texte en trois parties », « je commence par une introduction consacrée à la linguistique », etc.) et la réduction de ces phrases laisse des traces dans le texte qui sont des propriétés typographiques et dispositionnelles. L'ensemble des phrases décrivant les propriétés morpho-dispositionnelles et lexico-syntaxiques forment le **métalangage architectural** (Pascual, 1991).

Cette approche a pour conséquence l'existence d'un continuum entre des formes de MFM entièrement discursives et des formes visuelles : on passe d'une extrémité à l'autre de ce continuum en effaçant (au profit de marques visuelles) ou en reconstruisant (par interprétation) le métalangage architectural. L'intérêt d'utiliser cette approche dans l'étude de l'oralisation des structures textuelles est qu'elle fournit un cadre général d'interprétation et de reformulation des phénomènes typo-dispositionnelles qui ne sont pas directement transposables à l'oral. Nous allons chercher à reconstruire une partie du métalangage architectural qui est entièrement ou partiellement effacé de façon à expliciter de manière discursive les propriétés structurelles des textes.

La formalisation du modèle d'architecture textuelle nécessite de définir précisément les notions déjà évoquées ci dessus. Les définitions et notations suivantes sont extraites de Pascual (*ibid.*) et Luc (2000).

Définition 1 : Objet Textuel (OT)

Un objet textuel est un segment de texte caractéristique, accompagné de son entête s'il en a un (liste, énumération, chapitre, section, paragraphe, introduction, préface, théorème, avertissement, rubrique, etc.). Dans un texte, l'objet textuel le plus grand est le texte lui-même.

Définition 2 : Mise en Forme Matérielle (MFM)

La mise en forme matérielle d'un texte est l'ensemble des propriétés de réalisation appliquées aux objets textuels. Ces propriétés sont de nature syntaxique (nominalisation, numérialisation, formes interrogatives, etc.), typographiques (caractères, polices, corps, styles, couleurs, etc.) et dispositionnelles (justification, colonnages, marges horizontales et verticales, sauts de lignes, de pages, etc.).

Définition 3 : Architecture

L'architecture est une composante abstraite du texte, qui est rendue perceptible par un jeu de contrastes de la mise en forme matérielle : mise en relief, mise en parallèle, etc.

Définition 4 : Métaphore

Une métaphore est une phrase du sous-langage spécialisé relatif à l'archituration textuelle. Elle correspond à la forme discursive d'un phénomène de mise en forme matérielle, exprimant l'intention architecturante sous-jacente à ce phénomène. Une métaphore réalise un acte de discours à vocation textuelle.

Définition 5 : Unité Textuelle (UT)

Une unité textuelle est un segment de texte ne comportant aucun objet textuel, c'est-à-dire un segment entièrement discursif du texte.

Définition 6 : Métadiscours

Suite d'instances de métaphrases qui entretiennent entre elles des relations de cohérence et de cohésion. Un métadiscours associé à un texte représente l'architecture de ce texte.

Pour déterminer une liste de phrases du sous-langage spécialisé dans l'architecture textuelle, la méthode exposée par Virbel (1987) repose sur l'utilisation de schémas syntaxiques dont le but est d'inventorier toutes les expressions discursives associées aux phénomènes de mise en forme matérielle. Le schéma initial

N₀acteur V [Prep] N₁acteur que P

N₀acteur et **N₁acteur** représentent les acteurs textuels (Mojahid, 1990)

V représente une famille de verbes dénotant l'intentionnalité de **N₀acteur** (exprimer, traduire, manifester, souligner, indiquer, ...)

[prep] est une préposition qui marque le sens de l'intentionnalité de **N₁acteur** (à l'attention de, pour, auprès de, aux yeux de, ...)

P est une proposition explicitant l'acte performatif⁶, elle-même décrite par des schémas de phrases. Le schéma général pour P est de la forme : **N₂acteur V' d N_{AT}**

N₂acteur est instanciable par un acteur textuel, éventuellement identique à **N₀acteur** ou **N₁acteur**

V' est un verbe performatif⁷ (ou une expression) ayant une contrepartie en terme de mise en forme matérielle ;

⁶ Énoncé qui constitue simultanément l'acte auquel il se réfère (« la séance est ouverte » peut ouvrir la séance).

⁷ Verbe dont l'utilisation constitue un acte en soi, se confondant avec l'acte d'énonciation, tel que juger, promettre, baptiser, bénir... Ils s'opposent aux verbes constatifs.

- d** signale un marqueur de lien de dépendance (préposition, conjonction, complétive, etc.) et sa nature dépend de V'
- N_{AT}** représente le ou les arguments textuels sur lesquels est dirigé l'acte performatif représenté par V'

permet la construction de schémas dérivés qui sont alors instanciés par des verbes et des substantifs exprimant la performativité. Un certain nombre de notations supplémentaires sont utilisées pour exprimer les métaphrases :

- **M_i** représentent des entités instanciables par des substantifs, pris dans un ensemble **E_i** associé à M_i ;
- **id_i** représentent les identificateurs d'objets textuels ;
- **id_i.M_j** signifie que l'identificateur id_i doit être du type M_j, c'est-à-dire doit appartenir à l'ensemble associé à M_j ;
- **UT_i** représentent les unités textuelles ;
- **n** est un entier.

Les métaphrases relatives à l'architecture du texte peuvent être regroupées selon leur rôle vis à vis du texte : création, organisation globale, introduction d'objets textuels particuliers, liage des unités de texte aux objets textuels ou organisation des objets textuels entre eux.

Par exemple

L'auteur intitule $id_0.M_{ucn}$ par un titre identifié id_x .

*$M_{ucn} \in E_{ucn}$ où $E_{ucn} = \{texte\} \cup \{\text{livre, tome, volume, partie, chapitre, etc.}\}$
 $\cup \{\text{rubrique}\} \cup \{\text{item}\} \cup \{\text{théorème, hypothèse, exercice, problème, définition, etc.}\} \cup \{\text{commentaire, annotation, glose, précision, explication, interprétation, justification, etc.}\}$*

est la forme générale d'une métaphore introduisant l'objet textuel particulier « titre ».

L'architecture du texte peut être également représentée par son graphe architectural qui « traduit » visuellement le métadiscours. Ainsi, un texte composé d'une part d'un titre et d'autre part d'une énumération en deux points précédée d'une phrase introductrice pourra être représentée par la suite d'instances de métaphrases suivantes :

L'auteur crée un texte identifié $texte(1)$.

L'auteur intitule $texte(1)$ par un titre identifié $titre(1)$.

L'auteur attache $UT(1)$ à $titre(1)$.

- L'auteur compose titre(1) de UT(1).*
- L'auteur organise texte(1) en 2 parties identifiées partie(1) et partie(2).*
- L'auteur distingue une énumération identifiée énumération(1).*
- L'auteur agence énumération(1) en deux items identifiés item(1) et item(2).*
- L'auteur chapeaute énumération(1) par un chapeau identifié chapeau(1).*
- L'auteur attache UT(2) à chapeau(1).*
- L'auteur compose chapeau(1) de UT(2).*
- L'auteur attache UT(3) à item(1).*
- L'auteur compose item(1) de UT(3).*
- L'auteur attache UT(4) à item(2).*
- L'auteur compose item(2) de UT(4).*
- L'auteur compose texte(1) de titre(1) puis chapeau(1) puis item(1) puis item(2).*
- L'auteur utilise le système arabe pour énumérer item(1) et item(2).*

ou bien par un procédé dit d'Image De Page (IDP – Figure 5) sur le quel nous reviendrons plus en détail par la suite :

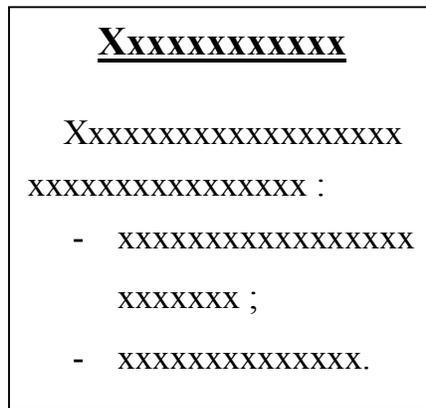


Figure 5 : Représentation d'un texte par une IDP

Un autre système de représentation a été proposé par Luc (2000), mais dans un objectif de combinaison du Modèle d'Architecture Textuelle avec une autre représentation liée à la théorie des structures rhétoriques (RST – Mann, 1988), en particulier pour rendre possible la prise en compte des structures énumératives dites non parallèles.

Enfin, précisons que le modèle est augmenté de propriétés :

- la relation de précédence P qui permet d'exprimer que X P Y signifie que dans un métadiscours la métaphore X requiert la présence de la métaphore Y ;
- la relation d'obligation O qui permet d'exprimer que X O Y signifie que dans un métadiscours la présence de X entraîne obligatoirement la présence de Y ;

- les lois de compositions qui concernent le fait de savoir de quels objets textuels et de combien d'objets textuels peut-être composé un objet textuel englobant.

Ces propriétés permettent de définir ensuite les notions de cohérence, de complétude et de contraintes de composition du métadiscours.

Nous concluons pour l'instant sur la description de ce modèle en résumant son triple intérêt :

- expressivité de la notation qui permet de créer facilement de nouvelles métaphrases ;
- prise en compte de structures non hiérarchiques ;
- souplesse des propriétés : ajout ou suppression de contraintes en fonction du niveau d'exigence sur le contrôle du document.

Nous attirons l'attention sur le fait que la terminologie développée dans cette partie sera très largement utilisée dans la suite de notre texte.

Cette dernière approche se distingue dans le domaine de la linguistique textuelle par le fait qu'elle aborde directement et de manière globale le problème de l'inscription graphologique du texte et propose un modèle de représentation potentiellement intéressant pour sa prise en compte lors du traitement automatique de la langue écrite.

Ce dernier travail est également un exemple de la nécessité de la constitution de typologies fines et rigoureuses pour chacun des Objets Textuels susceptibles de composer un texte. Nous concluons donc cette section dédiée aux aspects linguistiques de la morpho-disposition par la description de trois Objets Textuels relativement fréquents : le paragraphe, la structure énumérative et le titre.

III.7 Observation de 3 Objets Textuels

III.7.1 Le paragraphe

Bien que d'un point de vue historique le terme de paragraphe désigne simplement le signe d'écriture « § » dont l'usage était de marquer les différentes étapes d'un exposé (d'où sa liaison avec la composition, la disposition et donc le contenu du texte), il peut être désormais considéré comme dénommant les objets du texte écrit qui se visualisent comme des sections délimitées par des passages à la ligne.

Ainsi, son inscription graphologique se caractérise par un alinéa de début de paragraphe et un alinéa final. Le paragraphe se présente

comme une phrase ou une suite de phrases entre deux alinéas, c'est-à-dire comme un bloc de phrases tenant son unité d'un artifice graphique, l'alinéa. (Mitterrand, 1985).

Malgré ces deux propriétés « théoriques », le paragraphe reste relativement souple (longueur extrêmement variable) et ambigu (non univocité de l'alinéa initial qui peut apparaître, par exemple, devant des éléments de discours direct ou ne pas apparaître si il s'agit d'un paragraphe de début d'ouvrage ou de chapitre ; effacement de l'alinéa final dans le cas particulier où la ligne de fin du paragraphe est pleine, c'est-à-dire lorsqu'elle se confond avec la fin de la ligne graphologique).

Nous retrouvons également ces caractéristiques formelles de souplesse et d'ambiguïté au niveau fonctionnel : le découpage du texte en paragraphe peut être dépendant du rédacteur selon ses habitudes et les principes qui sous-tendent les changements de paragraphe semblent complexes et entremêlés. Cela dit Longrace (1979) en fait une unité grammaticale et Gardes-Tamine (1998) parvient à dégager trois raisons pouvant expliquer une transition vers un nouveau paragraphe. Elle définit cet objet textuel comme pouvant relever de trois intentions :

Les paragraphes thématiques qui peuvent être eux-mêmes de deux types :

- ceux pour lesquels chaque paragraphe correspond au regroupement d'une série de prédicats mais traitant d'un thème unique ;
- ceux pour lesquels il est possible d'associer à chacun un nouveau thème relié à celui du paragraphe qui le précède (les marques nettes de liaison sont facultatives).

Notons que nous avons pu observer à plusieurs reprises une combinaison de ces deux principes, réalisée par une possibilité de double découpage offerte par l'inscription graphologique démarcative (Cf. pour un exemple l'encadré dans la section suivante dédiée au point de vue des informaticiens) : un premier niveau de découpage en paragraphe utilise classiquement l'alinéa, tandis qu'un second niveau est mis en évidence par l'augmentation de l'interlignage. Chacun de ces deux niveaux correspond respectivement aux deux possibilités d'organisation des thèmes en paragraphes telle qu'elle a été exprimée précédemment.

Une deuxième interprétation de l'utilisation du paragraphe est de rigueur lorsque les transitions entre eux coïncident non pas avec la volonté d'une organisation thématique mais avec le passage à un autre type de discours : transition d'un fragment narratif vers un fragment descriptif ou vers des commentaires par exemple.

Enfin, les paragraphes dits énonciatifs dénotent un changement de modalité en passant par exemple de l’assertion à l’interrogation ou au discours indirect libre.

Finalement,

le paragraphe a une fonction organisatrice. Il constitue des moments distincts du texte. La distinction entre ces moments sera d’autant plus sensible que chacun manifesterà une unité d’autant plus nette (Garde-Tamine, 1998 :83)

III.7.2 Les structures énumératives

Bien que ces structures soient très souvent utilisées, il n'existe pas de définition attestée sur l'objet énumération ; on appréhende souvent une énumération comme un procédé consistant à mettre sur le même plan d'équivalence (fonctionnelle et visuelle) des items. Cette vision nous semble trop restrictive au vu de la grande diversité des types de structures énumératives observés.

En effet, des travaux sur l'énumération ont montré la diversité de ces structures : (Luc, 2001) propose une typologie des énumérations non-parallèles c'est à dire des énumérations ne respectant pas le parallélisme entre présentation et fonction des items. En opposition à ce terme, on parlera d'énumération parallèle pour les énumérations respectant ce parallélisme (Cf. figure 6)

De manière général, Luc définit cet objet comme une composition de différents Objets Textuels:

- un item : c'est une entité co-énumérée ;
- une amorce : c'est une phrase introductrice précédant et introduisant l'énumération. Une amorce peut être syntaxiquement complète ou syntaxiquement incomplète ; dans ce dernier cas, le ou les constituants manquants sont fournis par un ou des items ;
- une énumération : c'est l'ensemble des items ;
- une structure énumérative : c'est l'ensemble amorce + énumération.

Quand à C. Bush (Bush, 2000), elle a dressé un modèle (sémantique et syntaxique) de l'amorce pour les énumérations d'entités nommées.

Dans le cadre d'un projet soutenu par la région Midi-Pyrénées (Vigouroux, 2002), nous avons analysé des textes issus de brochures de présentation destinées aux personnes âgées. Parmi ces documents, on trouve : (1) des dépliants publicitaires (maisons de retraites, appareils auditifs, etc.), (2) des brochures d'associations (aide

pour la maintien à domicile, etc.), et (3) des informations sur les aides sociales (Centre Communale d'Action Sociale – C.C.A.S.– , Prestation Expérimentale de Dépendance – P.E.D. –, etc.). Le travail d'observation approfondie des textes a été effectué sur six documents.

Ce travail d'introspection avait pour objectifs de dresser l'inventaire des objets textuels présents, d'analyser les propriétés structurelles et linguistiques de ces objets ; ces propriétés pouvant être exploitées dans le cadre de l'oralisation et d'examiner l'influence de la variabilité (de MFM, rhétorique, etc.) sur le processus d'oralisation (Maurel, 2002a). Nous présentons ici quelques résultats issus de cette analyse pour les structures énumératives. Ces résultats sont représentatifs des types de données recueillies.

Nous avons noté une très grande hétérogénéité dans la MFM de ces objets pour les énumérations ; on trouve aussi bien des énumérations parallèles que non-parallèles, avec plusieurs formes de non-parallélisme. Par exemple, comme l'illustre la Figure 6, outre le cas du titre sur lequel nous reviendrons dans la sous-section suivante, nous remarquons une amorce syntaxiquement incomplète, et une énumération non-parallèle : les items ne sont pas sur le même plan d'équivalence structurelle. En effet, différents items (de « restaurants... » à « salon de jeux ») se rapportent au premier item (« au rez-de-chaussée ») alors que les deux derniers items se rapportent à l'item « au 1^{er} étage ». De la même façon, seule une partie des items commencent par un article.

-
- B) LES LOCAUX RESIDENTIELS**
Sont :
- au rez-de-chaussée :
 - un restaurant ouvert sur une terrasse,
 - un bureau de médecin,
 - une pharmacie - infirmerie,
 - une salle de kinésithérapeute,
 - salon télé avec cheminée,
 - salon de jeux.
 - au 1^{er} étage :
 - salon télé bibliothèque,
 - salon de coiffure.

Figure 6 : Énumération extraite d'une brochure de maison de retraite

Est également associé aux structures énumératives un ensemble de marques textuelles (de mise en forme matérielle) de nature ponctuationnelle, typodispositionnelle ou lexico-syntaxique. Comme nous l'avons remarqué lors de l'observation du corpus il existe une très grande variabilité dans la présentation des textes. De la même manière les structures énumératives sont signalées par des marques de MFM variées et variables (Luc, 2001), (Maurel, 2002b). Parmi ces marques nous identifions :

- dans l'amorce :
 - **classifieur** : sert à définir la nature des items de l'énumération. Par exemple les items peuvent être regroupés sous le terme de « critères », « étapes », « types », etc.,
 - **numéral** : adverbes de quantités. Par exemple « un », « deux », « trois », « quelques », etc.,
 - **groupe nominal spécifique** : « suivants », « ci-après », etc. Ces termes forment, selon Bush (2000) une catachrèse spatio-temporelle,
 - **marques ponctuationnelles de fin d'amorce** : on trouve [:], [], [], [:] ;

- dans les items :
 - **organiseurs textuels** : « premièrement », « tout d'abord », « puis », « ensuite », « enfin », etc.,
 - **marque de début d'item** : « - », « 1., 2., 3., ... », etc.,
 - **marque dispositionnelle** inter-items : espacements verticaux et horizontaux entre les items,
 - **connecteur lexicaux** des items : « et », « donc », « puis », etc. présents en général en fin d'item,
 - **ponctuation** de fin d'item.

L'ensemble de ces éléments est repris dans le schéma de la Erreur ! Source du renvoi introuvable. et le Tableau 4 ci-après (qui complète cet ensemble avec des paramètres formels non encore évoqués et leurs variations possibles).

Dans l'amorce	
Paramètres formels	Possibilités de variation
Numéral	Indéfini (« quelques », « plusieurs », etc.) Défini (« un », « deux », etc.)
Classifieur	De générique (« type », « point », etc.) à spécifique ⇔ hyperonyme
Groupe nominal spécifique	Ex : « suivant », « ci-après », « ci-dessous », etc.
Complétude	Amorce syntaxiquement complète Amorce syntaxiquement incomplète
Marquage de fin d'amorce	« : », « , », « . », « ; »
Dans l'énumération	
Paramètres formels	Possibilités de variation
Marquage de début d'item	« - », « 1., 2., 3., ... », grasse, tabulation, etc.
Organisateurs textuels	Homogène (« premièrement ..., deuxièmement... ») Hétérogène (« premièrement..., puis, ...le troisième... »)
Classifieur	Repris de l'amorce
Longueur des items	De 1 mot à plusieurs phrases
Marquage de fin d'item	« : », « , », « . », « ; », etc.
Nombre d'items	De 2 à ∞
Type de l'énumération (Bouraoui, 2000) (Luc, 2000)	Classique (« hyperonyme ») Description d'une procédure/action Résumé Alternance Parallèle vs. non parallèle
Exhaustivité	Liste exhaustive d'entités énumérées Liste non exhaustive d'entités énumérées
Ordonnement	Liste ordonnée d'entités énumérées Liste non ordonnée d'entités énumérées

Tableau 4 : Variations des paramètres formels des structures énumératives

III.7.3 Le titre

Parmi les quelques linguistes qui se sont intéressés particulièrement aux titres comme Hoek (1981), Genette (1987), Sullet-Nylander (2001) nous retiendrons deux contributions récentes qui abordent la problématique de cet OT en lui conférant un rôle discursif primordial.

Une approche fonctionnelle de l'unité titre est fournie par Ho-Dac (2004). Ses travaux appliquent au titre le point de vue de la Systémique Fonctionnelle (Mattiessen, 1997) qui décrit les métafonctions remplies par les éléments textuels : interpersonnelle, idéationnelle et textuelle. La première, appliquée au titre, révèle sa propension à attirer l'attention du lecteur, la seconde se réfère à la capacité du titre à présenter un élément du monde extérieur qui doit être intégré au monde du texte, enfin, la troisième métafonction consiste à segmenter, regrouper, hiérarchiser les informations.

Ces deux dernières métafonctions du titre, mêlant organisation matérielle et contenu informationnel, sont mises en relation par les auteurs à travers différents degrés d'implication notionnelles. Sont ainsi distingués les titres uniquement référentiels, référentiels focalisateurs ou référentiels préparatoires, les titres thématiques et les titres à implication zéro. Cependant l'implication notionnelle des titres est le plus souvent réalisée par une combinaison de ces trois grands modes de participation à l'organisation du texte : en marge de ces combinaisons se trouvent les titres combinant implication zéro avec une implication soit référentielle soit thématiques ; plus fréquemment les titres semblent mêler implication thématique et référentielle en donnant une part plus ou moins grande à l'importance de chacun de ces deux aspects. Parmi ceux-là, les auteurs proposent un classement qui distingue, en plus des titres à implication essentiellement référentielle ou essentiellement thématique, les titres un peu plus référentiels que thématiques, les titres mi-référentiels / mi-thématiques, les titres un peu plus thématiques que référentiels.

Enfin, ce travail ayant pour objectif premier une catégorisation automatique des titres, les auteurs proposent une réflexion sur les indices d'implication notionnelle du titre selon sa forme, sa situation au regard des autres titres du document, les conditions de reprise de ses éléments.

Sur ces derniers aspects, Virbel (2002) nous présente quelques éléments d'analyse pour mettre en évidence les manques et les limites de la conception « classique » du titre qui stipule que celui-ci a pour rôle d'identifier un autre OT tout en proposant son résumé significatif. L'auteur précise qu'il est nécessaire de relativiser la portée générale de cette définition et de prendre en compte d'autres aspects liés aux relations entre les

différents titres, entre les titres et les OT qu'ils intitulent ou d'autres mécanismes tels que l'énumération, la mise en saillance, le résumé, la numérotation ...

Sa méthode d'analyse permet de faire émerger une première typologie de cet OT basée sur le type des relations mises en évidence : syntaxiques, spatiales ou logico-sémantiques.

En ce qui concerne les relations syntaxiques entre un OT et son titre elles peuvent être, selon l'auteur, de type *parallèle* lorsqu'il n'existe pas de relation syntagmatique entre l'OT et le titre (accompagnement en marge du texte principal, tables analytiques, ...), ou de type *extrait* lorsque le titre se révèle par un aménagement syntaxique particulier et l'utilisation de la typo-disposition (l'auteur recense 16 combinaisons impliquant 4 types d'aménagements syntaxiques et 4 stratégies typo-dispositionnelles). Le positionnement du début du titre par rapport à l'OT qu'il intitule peut également largement différer du traditionnel « en haut, à gauche de l'OT » dès qu'il est syntaxiquement autonome. Enfin, l'auteur distingue trois types de relations logico-sémantiques selon que le titre précise ce dont parle l'OT qu'il intitule (titres thématiques), exprime la fonction du titre au sein du texte (titres fonctionnels) ou encore compose ces deux relations en les enchaînant ou en les fusionnant (titres tels que *Introduction*, *Annexe*, ... sont des exemples de cette « fusion sémantique »).

Parfois, les titres sont mis en relation les uns avec les autres à travers des procédés syntaxiques (leur concaténation forme un segment textuel autonome, présence d'une reprise anaphorique d'un titre antérieur, ...) ou logico-sémantiques (chaque titre d'une section est composé d'un jour de la semaine, utilisation d'un *pattern* identifiable comme dans un ouvrage consacré aux recettes de cuisine, ...)

Nous percevons un intérêt évident de cette typologie pour mettre en évidence les problématiques que peut potentiellement soulever l'oralisation des titres.

Conclusion

Cette partie, consacrée aux relations entre synthèse de parole, typographie, psycholinguistique et linguistique nécessite d'éclaircir les points importants abordés qui sous-tendent, influencent, ou simplement inspirent, la suite du tapuscrit.

La première constatation qui met en évidence les verrous technologiques et scientifiques soulevés par notre problématique provient du rôle accordé à la typographie par les concepteurs de systèmes de synthèse de la parole à partir de texte et par les typographes eux-mêmes. Les premiers ne prennent pas en compte cette dimension des textes écrits à un haut niveau d'analyse linguistique alors que les seconds se sont aperçus au cours du temps que le rôle de la typographie ne peut être précisé qu'en mettant en commun les compétences d'autres champs disciplinaires tels que ceux de la psychologie et de la linguistique.

La section consacrée au point de vue des psycholinguistes nous fournit deux éléments essentiels :

- un certain nombre de travaux permettent de présumer un impact important du marquage typographique sur certains processus cognitifs alors que la morpho-disposition des textes ne trouve toujours pas sa place au sein d'un modèle de compréhension du texte : notre travail devra apporter des éléments pour participer à un tel objectif ;
- une propriété tout à fait spécifique des textes écrits pour être lus silencieusement dont il faudra tenir compte est l'impact du premier regard (*first glance*) sur la suite du processus de lecture aussi bien en terme de stratégies que de mémorisation/compréhension des informations.

L'observation de la littérature extraite du domaine linguistique nous permet de relever également un certain nombre d'idées importantes et productives du point de vue de notre problématique de transposition orale.

Tout d'abord Vachek nous apprend qu'il ne faut pas s'attendre à trouver dans la norme écrite un moyen de marquage des émotions (bien que la typographie expressive peut détourner cette norme écrite dans un tel objectif). Les recherches concernant ce dernier point sont donc *a priori* à écarter pour notre étude. Notons tout de même que l'apparition des nouvelles technologies de la communication semblent nuancer cette observation. En effet, l'utilisation du mël, par exemple, repose sur un contexte de communication beaucoup plus dialogique qui met à mal les principes de « réaction statique à un stimulus non urgent ». D'ailleurs, ceci participe probablement à expliquer

l'apparition d'un nouveau matériel graphique (Nouvelles Formes de Communication Écrite ou NFCE) dont un des objectifs peut être « d'inscrire » ses émotions, à l'instar de l'oral, de manière prompte et immédiate (Guimier de Neef, 2004), (Dausse, 2004).

Bernhart nous affirme ensuite qu'un angle de vue incontournable pour aborder la problématique de la présentation typographique des informations est celui des marqueurs de cohésion (Halliday, 1976). Si nous n'avons trouvé que peu de résultats de cet auteur, un rapprochement peut être opéré avec les notions de cadre de discours (Charolles, 1978) ; notions plus récentes et dans la continuité de ces travaux. Pour cet auteur, la cohésion repose sur des marqueurs tels que les connecteurs, les anaphores, les chaînes de référence, les expressions introductrices de cadres discursifs et **les marques configurationnelles** (telles que les alinéas). Les travaux sur ce dernier point (Charolles, 1994:128, 1997) nous permettent de présumer un rôle particulier pour l'inscription graphologique : l'explicitation des cadres du discours. Par exemple, si l'on considère le texte (extrait de stimuli expérimentaux proposés par Michel Charolles)

Paul faisait une croisière en paquebot. Il monta sur le pont et s'approcha du bord. A la surface de l'eau, quelques dauphins faisaient la course. Des poissons sautaient dans tous les sens. Sur les îles, des habitations apparaissaient de loin en loin. Plusieurs moulins dressaient leurs ailes au vent du large.

des cadres sont ouverts puisque certaines phrases sont implicitement dépendantes des précédentes. L'utilisation d'une inscription graphologique appropriée serait à même d'expliciter ce phénomène :

Paul faisait une croisière en paquebot. Il monta sur le pont et s'approcha du bord :

- *à la surface de l'eau*
 - *quelques dauphins faisaient la course,*
 - *des poissons sautaient dans tous les sens ;*

- *sur les îles*
 - *des habitations apparaissaient de loin en loin,*
 - *plusieurs moulins dressaient leurs ailes au vent du large.*

Waller nous précise que la non linéarité du texte est à nuancer par la linéarité de la lecture qui en est faite. Il nous semble pourtant risqué de prétendre avec lui que les processus cognitifs mis en jeu s'appuient alors sur une entrée linéaire, en atteste l'impact du *first glance* sur lequel nous reviendrons ultérieurement. Cela dit nous conservons l'idée d'une typographie dont le rôle est d'organiser la structure thématique du texte en fonction des attentes pressenties du lecteur. Comme nous l'avons évoqué dans la partie préliminaire de cette section pour la mise en gras, un point crucial est de

considérer une part de l'inscription graphologique comme un moyen d'explicitier la structure informationnelle du texte.

Les travaux sur la place de la ponctuation en linguistique nous ont révélé des notions intéressantes telle que le ponctème chez Catach, notion qu'elle compare avec celle d'intonème chez Delattre ; l'idée d'une grammaire de texte chez Nunberg afin de prendre en compte la hiérarchie imposée par la ponctuation à certains segments du texte ; et la gestion de l'indentation des segments de texte de Power et Scott pour généraliser les règles établies par Nunberg. Cela dit, bien que ces travaux s'appuient sur l'idée qu'il faut considérer la ponctuation comme la construction de configurations à visées sémantico-pragmatiques, les auteurs restent trop laconiques en termes de typologies de ces configurations et de leur interprétation pour une implication directe dans notre étude.

Nous rajouterons ici une réflexion qu'il nous a été nécessaire de faire pour affiner la notion de mise en saillance et la rendre exploitable par la suite dans le cadre de l'oralisation. N'ayant trouvé que peu de littérature scientifique sur cet aspect de l'inscription graphologique, nous décrivons rapidement ci-dessous ce que recouvre ce terme pour nous et en quoi cette discussion, ajoutée aux précédentes, nous oblige à redéfinir quelque peu notre taxonomie.

Le rédacteur peut recourir à des procédés qui jouent sur les caractères imprimés au niveau du corps, de la graisse ou du style. La mise en valeur peut être réalisée, selon Richaudeau (1989), en fonction de la puissance du contraste qu'ils produisent sur du caractère romain, bas de casse, maigre ; ainsi, du plus perceptible au moins perceptible nous trouvons : le soulignement (celui-ci est souvent dénoncé mais sans véritable raison – si ce n'est lorsqu'il est mal réalisé – puisque son absence dans la composition au plomb n'est que l'expression de contraintes techniques aujourd'hui disparues), le « **graisage** » (sur du maigre il s'agit du normal, gras ou extra-gras), LA CAPITALISATION, *la mise en italique*. Nous rajouterons des procédés dont la force contrastive peut être variable tels que l'**habillage du texte** (encadrés, trames, surlignage) et le changement de corps, ou de style.

Le problème de la non univocité de ces marques doit être soulevé pour défendre le fait que lorsque ces procédés sont utilisés en se substituant, voire en s'additionnant, à d'autres signes, et ceci de manière homogène dans l'ensemble du texte (par exemple le changement de style et la graisse dans ce manuscrit se substituent aux classiques guillemets pour démarquer les citations référencées), ils ne font pas partie de l'objet de cette discussion puisque participant *a priori* à l'inscription graphologique démarcatrice ou stylistique. Dans le même ordre d'idées, lorsqu'une inscription graphologique

spécifique utilisant ces marques est utilisée à des fins esthétiques, pour des raisons de lisibilité ou sur la base de recommandations ou de conventions, elle relève alors purement de l'inscription graphologique stylistique (par exemple la présentation visuelle de la bibliographie ou des titres). Le rôle de la mise en saillance ainsi précisé est en fait de hiérarchiser l'information d'un point de vue sémantico-pragmatique, et cette possibilité lui est donnée depuis le niveau graphologique du mot (nous avons déjà abordé ce sujet dans nos « justifications préliminaires » en début de partie pour la mise en saillance par la graisse) jusqu'à l'Objet Textuel de niveau juste inférieur au texte lui-même.

Par exemple un paragraphe pourra être agrémenté d'une inscription graphologique distinctive renforçant son statut important au sein d'un texte. Certains auteurs d'ouvrages scientifiques utilisent également la diminution de la taille des caractères pour la fonction métalinguistique qui est de signifier que le passage ainsi mis en saillance est réservé aux lecteurs aguerris, et que le texte a été construit pour que leur « non lecture » ne nuise pas à des lecteurs moins spécialistes du domaine. Il s'agit donc, dans tous ces cas, d'attirer l'attention sur le fait que le contenu sémantique du texte sur lequel l'inscription graphologique s'applique est jugé par le rédacteur plus ou moins important, plus ou moins accessible, plus ou moins pertinent, métalinguistique ou non, etc.

Ces exemples montrent toutefois une différence que ne prend pas en compte notre taxonomie : la mise en saillance peut non seulement s'appliquer à un objet textuel déjà démarqué, mais également à un regroupement d'objets textuels qu'elle démarque donc en même temps qu'elle les distingue. Une autre utilisation fréquente de ce second procédé est par exemple **la mise en saillance d'un groupe de mot**.

Finalement nous reviendrons sur notre taxonomie en retenant, en plus des inscriptions graphologiques imposées et stylistiques, trois autres types d'inscription graphologique :

- **inscription graphologique démarcative** : ensemble des marques qui expriment la transition visuelle d'un objet textuel à un autre, explicitant ainsi de nouveaux cadres du discours ;
- **inscription graphologique distinctive thématique** : ensemble des marques qui permettent d'explicitement visuellement tout ou partie de la structure informationnelle d'un objet textuel, en terme de thème/rhème, topique/commentaire, posé/présumé, donné/nouveau, ...

- **inscription graphologique distinctive scalaire** : ensemble des marques qui permettent de lier des objets textuels visuellement sur une échelle de valeur ou à travers une relation d'opposition. Elle regroupe également les marques qui attribuent visuellement un modus au *dictum* porté par un objet textuel ou un regroupement d'objets textuels.

Dans l'exemple de la structure énumérative ci-dessus nous observons ces trois types d'inscription graphologique à l'œuvre : 1 objet textuel « amorce » et 3 objets textuels « items » sont démarqués et dans chacun de ces derniers sont également démarqués deux objets textuels correspondant à un terme et à sa définition ; à l'intérieur de chaque définition est distingué un objet textuel « mot(s) soulignés » dont le rôle semble être d'extraire ce que nous pourrions qualifier de « noyau rhématique » de la définition ; le dernier item comprend deux objets textuels « mots en italique » pour distinguer leur contenu par une forme de modalisation visant à produire la dichotomie métalinguistique « mot étranger / mot français ».

Enfin, l'approche globale proposée par Virbel, à travers son Modèle d'Architecture Textuelle (MAT), stipule l'existence d'une équivalence fonctionnelle entre les marques typographiques, dispositionnelles et des marqueurs lexico-syntaxiques tournées non pas vers le monde mais vers le texte lui même. Si cette idée nous laisse entrevoir l'utilisation du concept de métaphore pour recomposer un texte exploitant les trois types d'inscription graphologique précédemment décrits (que ce modèle regroupe sous le terme de Mise en Forme Matérielle de nature typo-dipositionnelle) sous une forme entièrement discursive (grâce en particulier à la Mise en Forme Matérielle de nature lexico-syntaxique), et donc oralisable, elle nous inspire également de *tester l'hypothèse d'une équivalence fonctionnelle entre ces nouvelles marques de nature lexico-syntaxique et des marques que nous qualifierons de « typophoniques » ou prosodiques*.

Sur ces bases, nous concluons cette discussion par l'affirmation de la nécessité de concevoir une architecture informatique plus adaptée au traitement en amont du texte à oraliser dans la perspective d'une transposition automatique à l'oral des structures visuelles des textes. L'objet de la partie suivante est la description synoptique d'une telle architecture et de ses principes, l'élaboration d'un modèle qui permette d'envisager l'opérationnalisation d'un module central de cette architecture et l'opérationnalisation partielle de ce modèle en focalisant notre étude sur l'oralisation des structures énumératives.

« Ce qui n'a jamais été articulé en mots finit par devenir trop nébuleux pour s'inscrire dans la mémoire. »

Jane Smiley, *L'exploitation*

Partie B Architecture, modèle et opérationnalisation

Cette partie propose, à partir des conclusions précisées dans la précédente, un système synoptique dont l'unité de traitement du texte en vue de sa transposition à l'oral est l'Objet Textuel (OT).

Nous montrons la faisabilité de la mise en œuvre du module de reformulation, de manière à ce que les systèmes de synthèses actuels soient capables, à peu de frais, de transposer à l'oral cette nouvelle unité.

Dans cet objectif, nous proposons un modèle d'oralisation par reformulation des OT décrivant les stratégies envisageables et les relations qu'elles entretiennent entre elles. Afin de mettre en évidence l'intérêt de ce modèle pour l'ingénierie informatique nous proposerons son opérationnalisation pour la transposition à l'oral des structures énumératives.

A travers l'étude de cet OT spécifique et complexe, nous proposons, pour parvenir à l'automatisation de nos stratégies, une méthode générale basée sur notre taxonomie, des représentations informatiques canoniques de chaque version des OT et des règles de transformation d'une représentation dans une autre *a priori* directement interprétable par un TTS.

Les spécifications proposées pour générer automatiquement, à partir d'une représentation canonique, la version prosodique des structures énumératives ont été implémentées dans une version du moteur de synthèse d'un partenaire industriel (Proverbe d'Elan Speech).

Chapitre IV. Une architecture adaptée à l'oralisation des textes par « OT »

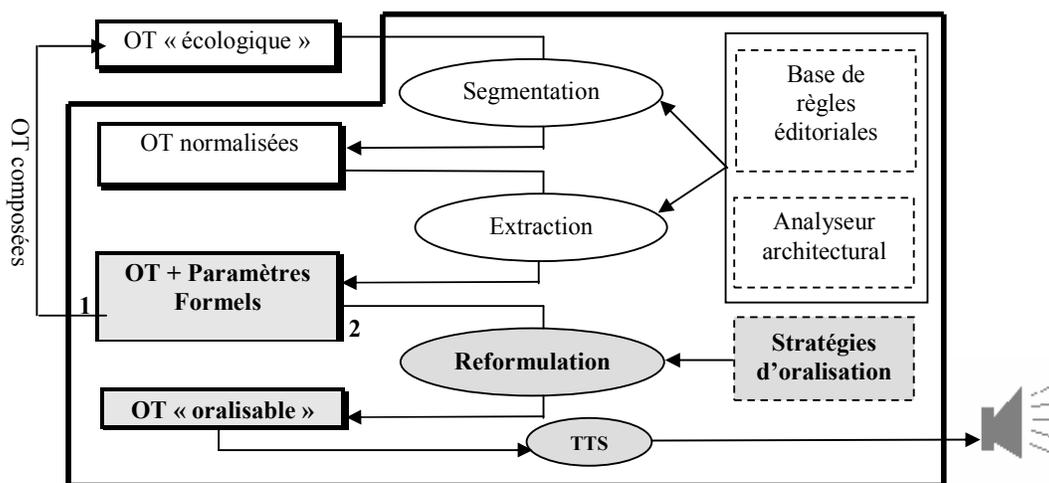


Figure 8 : Système synoptique d'oralisation des structures visuelles des textes

L'originalité du système proposé (Figure 8) réside dans l'unité d'analyse considérée pour le traitement des documents écrits (Maurel, 2002c). En effet, elle ne correspond plus à la traditionnelle phrase ponctuée et n'a pas non plus l'ambition, dans cette thèse, d'intégrer le texte globalement avec son co-texte et ses contextes. En fait, le processus d'oralisation décrit par cette architecture permet de faire « traverser » à un Objet Textuel (OT) plusieurs phases de transformation qui s'appuient sur un ensemble de connaissances associées à chacune d'elles. Ces trois phases permettent, à partir d'un OT « écologique », de fournir une représentation, associée à une stratégie d'oralisation particulière, interprétable par un TTS.

Les différentes étapes de ce processus sont décrites ci-après.

- phase de *segmentation* : une première étape qui s'appuiera sur des indications d'ordre syntaxique, pragmatique, stylistique et structurel, regroupées sous le terme de *base de règles éditoriales*. Cette première étape permet de contrôler en partie les phénomènes de variabilité du texte en reconnaissant et normalisant (du point de vue de son inscription graphologique) l'OT en entrée du processus. La segmentation de l'OT produira sa décomposition en un ensemble d'OT de niveau juste inférieur. Par exemple, la segmentation de l'exemple encadré ci-dessous (Figure 9) devra repérer un ensemble d'OT comme

s'enchaînant de la manière suivante : « titre », « titre », « paragraphe », « paragraphe », « paragraphe », « paragraphe » ;

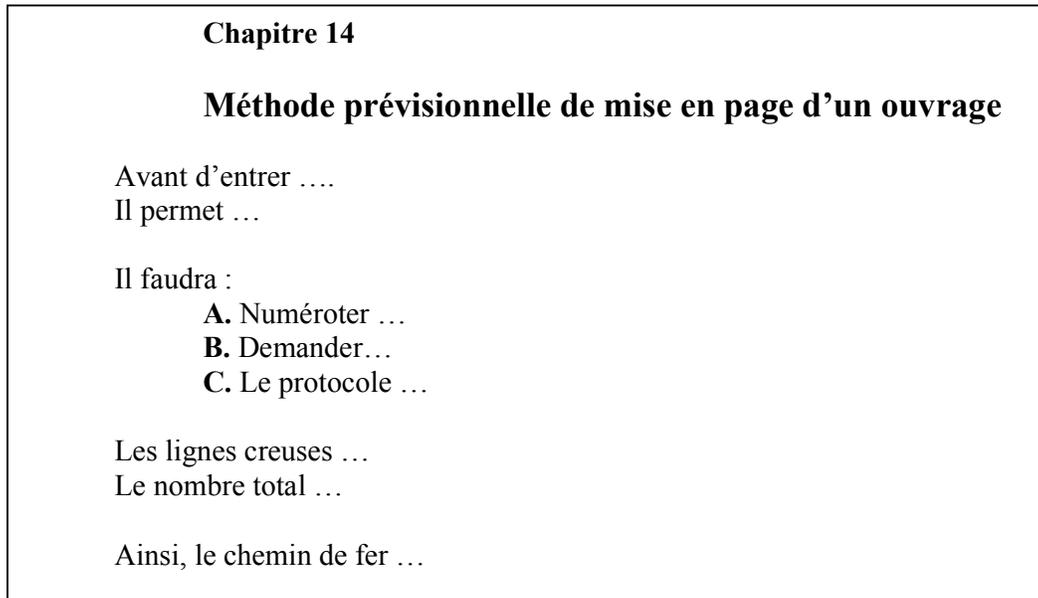


Figure 9 : Exemple de texte mis en page, inspiré de Duplan (1986 :127)

- phase d'*extraction* : la deuxième étape consistera à extraire des OT obtenus les différents *paramètres formels*, qui sont des *connaissances structurelles et linguistiques*, dont la variation semble *a priori* pertinente à étudier en terme d'influence sur l'élaboration des stratégies d'oralisation. Ce processus permet d'obtenir une représentation de l'OT original qui prenne en compte l'ensemble des paramètres formels, repérés s'ils sont explicites ou déduits s'ils sont implicites. En poursuivant l'exemple précédent nous obtiendrions un raffinement tel que : « titre **exploitant un système de numérotation arabe pour le chapitre 14** », « titre **thématique du chapitre 14** », « paragraphe **introductif** », « paragraphe **structure énumérative** », « paragraphe », « paragraphe **conclusif** ». (les apports de chaque passage par une étape de segmentation ou d'extraction sont mis en saillance par la graisse dans nos exemples) ;
- nous remarquons que dans notre exemple trois OT restent décomposables en OT de niveau inférieur. Les deux processus précédents s'appliqueront alors récursivement (chemin 1 du schéma) sur ces objets jusqu'à ce que chaque OT ne soit composé que d'OT de type « phrases ponctuées sans inscription

graphologique distinctive » (ces derniers, pouvant être qualifiés d'Unité Textuelle ou UT si l'on se réfère au MAT, étant traités convenablement par les systèmes de synthèse). Un passage supplémentaire de ces trois OT à travers les deux processus donnera pour notre exemple

- « paragraphe introductif »
 - « **sous paragraphe** »
 - « **sous paragraphe** »
- « paragraphe structure énumérative »
 - « **amorce** »
 - « **énumération** »
- « paragraphe »
 - « **sous paragraphe** »
 - « **sous paragraphe** »

pour la phase de segmentation. Puis la phase d'extraction produira (nous ne donnons que le cas de la structure énumérative car elle est la seule à subir des transformations) :

- « structure énumérative »
 - « amorce **incomplète** »
 - « énumération **parallèle, ordonnée et exhaustive classifiant des actions avec une numérotation alphabétique** »
- enfin une troisième itération sur ce dernier objet produira d'abord :
 - « structure énumérative »
 - « amorce incomplète »
 - « énumération parallèle, ordonnée et exhaustive classifiant des actions avec une numérotation alphabétique »
 - « **item** »
 - « **item** »
 - « **item** »

et finalement :

- « structure énumérative »
 - « amorce incomplète »
 - « énumération parallèle, ordonnée et exhaustive classifiant des actions avec une numérotation alphabétique »
 - « **premier item** »
 - « **deuxième item** »
 - « **troisième et dernier item** »

Le type des OT finaux repérés dans l'exemple ainsi que leurs paramètres formels sont récapitulés dans l'encadré suivant. Il s'agira de fournir au module suivant une représentation du texte à oraliser qui explicite l'ensemble de ces informations.

« titre exprimant le système de numérotation arabe du chapitre 14 »
« titre thématique du chapitre 14 »
« paragraphe introductif »
 « sous paragraphe »
 « sous paragraphe »
« paragraphe structure énumérative »
 « amorce incomplète »
 « énum. parallèle, ordo. et exhau. classifiant des actions avec une num. alphabé. »
 « premier item »
 « deuxième item »
 « troisième et dernier item »
« paragraphe »
 « sous paragraphe »
 « sous paragraphe »
« paragraphe conclusif »

Figure 10 : Types finaux des OT issus des phases de segmentation/extraction

- phase de **reformulation** : il s'agit, lors de cette troisième étape (chemin 2 du schéma), centrale à notre problématique, d'opérer un certain nombre de transformations qui visent à obtenir un OT oralisable en fonction (1) des paramètres formels (déduits et repérés), (2) des relations de dépendance qu'ils entretiennent, et (3) du type de stratégie envisagée.

Moyennant une représentation finale adéquate, ces trois étapes permettraient, à partir d'un objet textuel écologique, de fournir en entrée d'un système de synthèse de parole, un objet « oralisable » et, le cas échéant, les schémas intonatifs associés à une stratégie particulière, sans modifications de l'architecture classique du TTS (Cf. p33). Celui-ci pourra d'ailleurs s'affranchir de gérer ce que son système de prétraitement contrôle déjà, et prendre en compte, directement à travers les modules de phonétisation et de modélisation prosodiques, les nouvelles indications selon des spécifications qu'il nous impute de déterminer.

Dans l'état actuel des connaissances en linguistique textuelle, les deux premières étapes (segmentation et extraction) semblent difficilement abordables de manière computationnelle et en particulier en considérant des documents « tout venant ». En effet la première étape devrait tenir compte d'une variabilité quasi illimitée des textes

pour la reconnaissance automatique des OT et la seconde peut impliquer une analyse complexe pour extraire les paramètres formels associés à ces OT (le classifieur des items de l'énumération par exemple). Malgré tout, le respect de recommandations rédactionnelles pourrait faciliter l'automatisation de ces deux phases. Celles-ci seront donc effectuées « manuellement » de façon à obtenir la représentation canonique adéquate pour dériver les nouvelles versions fournies par le module de reformulation en entrée du système de synthèse classique.

Notre intérêt s'est ainsi porté sur le processus de reformulation qui implique les parties grisées de la Figure 7 : la mise en œuvre de ce processus implique (1) la conception d'un modèle des stratégies d'oralisation, (2) l'analyse des paramètres formels des OT étudiés, (3) l'élaboration de représentations informatiques pour les OT et chacune de leurs dérivations en fonction de la stratégie impliquée, (4) les spécifications pour la réalisation accoustico-prosodique par le TTS à partir de la représentation.

La partie suivante du manuscrit se consacre à ces différents objectifs.

Chapitre V. Modèle des stratégies d'oralisation

Notre premier objectif pour aborder la phase de reformulation a été de concevoir un modèle de stratégies d'oralisation prenant en compte l'inscription graphologique. L'axiome principal qui sous-tend le modèle prétend qu'il est nécessaire d'obtenir une forme discursive de l'OT commune aux deux modalités et que, de là, pourra être décrit un ensemble de stratégies utilisant des procédés spécifiques à la modalité orale. Sur ces bases, nous avons conçu le **Modèle d'Oralisation par Reformulation des Textes Écrits pour être Lus Silencieusement (MORTELS – Figure 11)**, (Maurel, 2004a).

A l'instar de l'hypothèse d'équivalence fonctionnelle, proposée par le Modèle d'Architecture Textuelle, entre la Mise en Forme Matérielle de nature typodispositionnelle et la Mise en Forme Matérielle de nature lexico-syntaxique (Harris, 1976), (Virbel, 1989), nous stipulons une équivalence fonctionnelle entre ces dernières et des configurations sonores. Nous pouvons nous convaincre de la plausibilité d'une telle hypothèse sur l'exemple simple de la phrase interrogative puisque celle-ci peut-être perçue comme telle :

- par un procédé discursif accompagné d'une conclusion intonative descendante : « est-ce que Max est venu ↓ » ; intonation conclusive identique à celle de l'affirmation « Max est venu ↓ » ;
- sans marqueurs discursifs particuliers mais avec une conclusion intonative montante spécifique à l'interrogation : « Max est venu ↑ »

Ainsi, la forme discursive peut être considérée comme le pivot central pour l'élaboration d'un continuum de nouvelles formes prosodiques. Celles-ci peuvent être générées par réduction de tout ou partie des marques de MFM de nature lexico-syntaxique, au profit d'un matériel oral non verbal équivalent du point de vue informationnel.

La forme discursive d'un Objet Textuel particulier peut être reconstruite selon trois stratégies :

- par une description objective de son inscription graphologique : forme discursive descriptive (S1) ;
- par un premier niveau d'interprétation, en attribuant un nom à la configuration graphologique : forme discursive dénominative (S2) ; ou

- par une interprétation plus profonde des intentions du rédacteur qui se traduisent par l'utilisation de cet OT particulier : forme discursive interprétative (S3).

Ces trois opérations pourront conduire ensuite à 4 continuums de nouvelles formes d'oralisation :

- la forme discursive descriptive pourra se réduire grâce à l'utilisation de conventions « typophoniques » par un procédé d'association bijective entre un ensemble de marques et un ensemble de sons (C1) ;
- la forme discursive dénomminative pourra se réduire par une convention « typophonique » qui associe de manière injective les marques définissant un type d'objet particulier à un son unique (C2a) ou par un schéma prosodique spécifique, repéré comme pouvant être associé au type d'OT étudié indépendamment de tout contexte d'énonciation (C2b) ;
- la forme discursive interprétative pourra se réduire par l'utilisation de schémas prosodiques dépendant de la situation d'énonciation et donc de la structure informationnelle de cette forme (en terme de thème, topique, focus, rhème - C3).

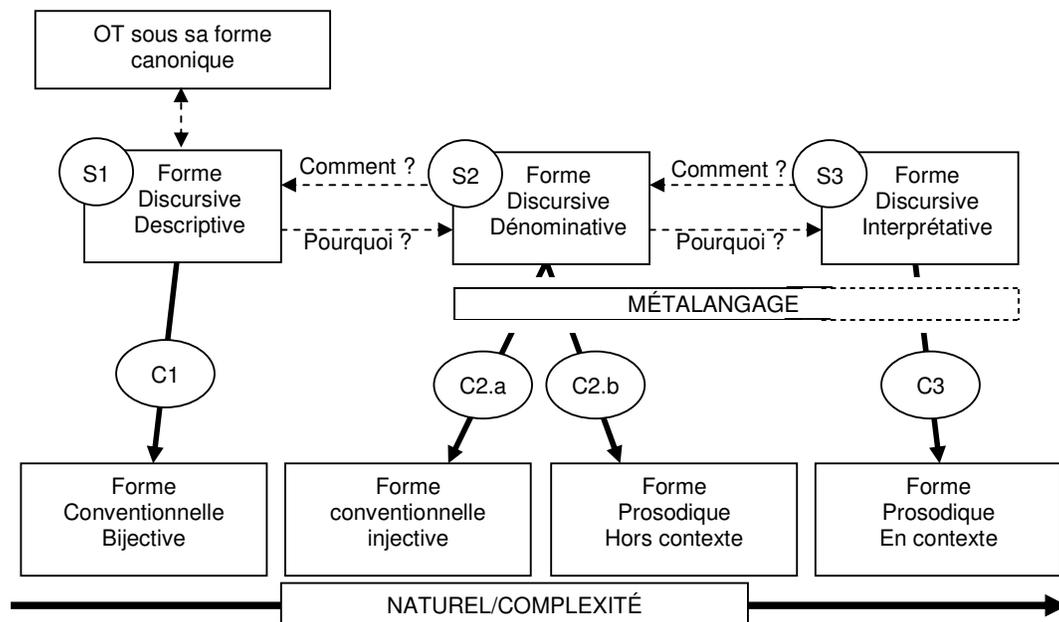


Figure 11 : MORTELS

Une illustration des trois stratégies dans leur forme discursive peut être donnée à travers l'exemple suivant. Soit le titre d'un document :

L'ORALISATION DES TEXTES

S1 : « **Le texte en gros, centré, en gras et en capitale est** l'oralisation des textes »

S2 : « **Le titre de niveau *n* est** l'oralisation des textes »

S3 : « **Je vais vous parler maintenant de** l'oralisation des textes »

Les parties en gras seront les marques lexico-syntaxiques restituées à partir de la morpho-disposition du texte. Ce sont aussi celles qui pourront potentiellement être effacées au profit de marques sonores telles que décrites plus haut.

Nous remarquerons, comme l'indique le schéma de gauche à droite, que les trois formes discursives sont en relation à travers une sémantique de « pourquoi » (relative au travail cognitif du **lecteur** d'un texte) ; il en découle de droite à gauche une sémantique de « comment » (relative au travail cognitif du **rédacteur** d'un texte) : dans l'exemple d'un titre *X*, la transition de (S3) à (S2) est effectuée à partir de la question *comment inscrire graphologiquement que l'on veut parler maintenant de X ?* par la réponse *en faisant de X un titre* ; la transition de (S2) à (S1) est effectuée à partir de la question *comment faire un titre ?* par la réponse *en centrant, soulignant, ...* qui permet au rédacteur de produire l'OT voulu.

Le métalangage nous permettra de proposer une version discursive dénomminative ou interprétative de l'objet. Le cadre du schéma se prolonge à droite par des traits pointillés car le choix d'un représentant d'une classe de métaphrases du métalangage nécessite des ressources qu'il ne sera pas toujours aisé d'obtenir pour la construction de la forme discursive interprétative : lorsque les indications nécessaires à l'opérationnalisation de cette stratégie sont exprimées d'une autre manière que visuellement ou lexico-syntaxiquement (contexte, co-texte, ...) le métalangage nous offrira sur cette base seulement des solutions de type dénomminatives.

Si nous prenons l'exemple de l'utilisation de la structure parenthésée de la dernière phrase du paragraphe précédent, une forme entièrement discursive descriptive se bornera à préciser « **l'auteur ouvre une parenthèse**, contexte, co-texte, point, point, point, **l'auteur ferme la parenthèse** » et la forme discursive dénomminative précisera plutôt « **l'auteur écrit** contexte, co-texte **et trois points, le tout entre parenthèses** ». Aucune marque visuelle ou lexico-syntaxique ne permet une interprétation plus approfondie de l'utilisation de la structure parenthésée (en l'occurrence, ici, une exemplification). Pourtant une telle interprétation, si elle présente moins d'intérêt dans le cadre de la génération de texte, est importante dans le cas de la transposition orale

puisqu'elle nous permettrait potentiellement d'accéder ensuite à des stratégies prosodiques plus adaptées. La présence des *trois points* en revanche est interprétable comme l'indication de la non-exhaustivité de la liste, indication qui trouvera une contrepartie métaphrastique. Ainsi la forme discursive interprétative devrait opérer la transformation

(contexte, co-texte, ...) → « , **par exemple le contexte, le co-texte, *et cetera***, ».

Finalement, la relation entre inscription graphologique et métalangage est non-ambiguë dans le cas de la forme discursive dénominative, alors que dans le cas de la forme discursive interprétative, l'inscription graphologique nous fournira un ensemble de métaphrases possibles (correspondant dans notre exemple de la structure parenthésée à un besoin passager de l'auteur d'expliquer, de préciser, d'ajouter une information, de faire un commentaire, ...); choisir parmi cet ensemble de métaphrases celle qui correspond à la bonne interprétation nécessite des ressources et une analyse autres que celles en relation avec l'inscription graphologique.

Ainsi nous voyons que si la qualité du naturel est d'autant plus augmentée que la stratégie utilisée se positionne à droite dans le schéma de la Figure 11, il en est de même de la complexité impliquée par l'opérationnalisation.

Une dernière remarque a trait au fait que nous ne traitons dans ce modèle que de l'oralisation de l'inscription graphologique, alors que le Modèle d'Architecture Textuelle met en évidence l'existence de marques de mise en forme de nature lexico-syntaxique qui, elles mêmes, poseraient, lors de leur transposition à l'oral, non pas des problèmes de naturel, mais des problèmes de cohérence spatio-temporelle : que faire d'un « ci-contre », d'un « dans la marge » ou d'un « l'antépénultième ligne du second paragraphe du quatrième de couverture » ! Dans les chapitres suivants, consacrés à l'opérationnalisation proprement dite de certaines stratégies décrites par le MORTELS, nous considérerons les abréviations suivantes :

FDDC : **F**orme **D**iscursive **D**es**C**riptive

FDDN : **F**orme **D**iscursive **D**é**N**ominative

FDI : **F**orme **D**iscursive **I**nterprétative

FCB : **F**orme **C**onventionnelle **B**ijective

FCI : **F**orme **C**onventionnelle **I**njective

FPHC : **F**orme **P**rosodique **H**ors **C**ontexte

FPC : **F**orme **P**rosodique **e**n **C**ontexte

Chapitre VI. Opérationnalisation des stratégies

Ce modèle ne peut être mis en œuvre de manière globale dans le cadre de ce travail puisqu'il est nécessaire pour chaque OT envisagé :

- d'effectuer une analyse détaillée en terme de paramètres formels. Nous nous concentrerons donc sur les objets décrits dans la partie précédente (Cf. Partie AIII.7) ;
- de préciser de quel type d'inscription graphologique est constitué l'OT. Le cas échéant, il s'agira également de mettre en évidence la structure informationnelle sous-jacente pour parcourir l'axe interprétatif du modèle. Ce point implique de choisir une situation d'énonciation particulière puisqu'un tel découpage en est fortement dépendant ;
- idéalement, de lui faire traverser tous les continuums de stratégies du modèle à partir de son point d'entrée : $S1 \rightarrow C1 - S1 \rightarrow S2 - S2 \rightarrow C2a - S2 \rightarrow C2b - S2 \rightarrow S3 - S3 \rightarrow C3$. D'un point de vue pratique, afin de limiter les études, nous opérerons de manière différente en considérant, comme nous l'avons évoqué, que l'axe interprétatif est prioritaire car *a priori* plus naturel (ce fait reste cependant discutable si nous nous référons aux nombreuses locutions métalinguistiques qui utilisent les autres stratégies de manière tout à fait naturelle dans le langage courant, qu'elles soient descriptives – « point final », « j'ouvre une parenthèse », « un premier point », ... – ou dénominatrices – « entre parenthèses », « entre guillemets », ... –). De plus nous ne considérerons que les stratégies situées aux pôles des continuums. Ainsi, notre point d'entrée sera mobile et variera en fonction de notre difficulté à réaliser les stratégies les plus interprétatives :
 - si FDI \rightarrow FPC non satisfaisant alors
 - si FDDN \rightarrow FPHC non satisfaisant alors
 - si FDDN \rightarrow FCI non satisfaisant alors FDDC \rightarrow FCB.

Les raisons de non satisfaction dans l'utilisation d'un axe stratégique particulier peuvent relever de différentes causes : manque de naturel ou d'intelligibilité de la solution, difficultés majeures pour l'opérationnalisation en terme de ressources ou de cadres théoriques, lourdeur du résultat, stratégie difficilement combinable avec celles déjà utilisées...

Nous avons choisi de détailler l'opérationnalisation de la phase de reformulation pour l'OT que nous avons étudié particulièrement en profondeur : les structures

énumératives. L'étude de l'oralisation de ces structures, qui est présentée ci-après, sera l'occasion d'exposer une modélisation de cet objet adaptée à notre problématique et un certain nombre de réflexions méthodologiques qui s'appliquent de manière générale quel que soit l'OT considéré.

VI.1 Les structures énumératives (SE)

La première partie, ainsi que ce qui précède cette sous-section, nous ont enseigné qu'il est nécessaire

- de choisir une situation d'énonciation particulière ;
- de fournir en entrée de notre modèle les OT dans une représentation normalisée, qui mette en évidence l'ensemble des paramètres formels nécessaires à l'application des stratégies d'oralisation ; ces paramètres sont *a priori* décrits dans le tableau 2 ;
- d'essayer de mettre en œuvre le cheminement suivant : **OT → FDI → FPC**.

VI.1.1 Type d'inscription graphologique et situation d'énonciation

Une particularité des structures énumératives est qu'elles relèvent, de notre point de vue, pour une part de l'inscription graphologique démarcative et pour une autre part de l'inscription graphologique distinctive thématique. En effet, l'amorce de la structure énumérative n'est en général pas démarquée du reste du texte qui la précède. Si tel est le cas, c'est que son introduction dans le texte coïncide avec l'explicitation d'un nouveau paragraphe, un nouveau cadre de discours. Dans ce cas, c'est à l'ouverture de ce cadre que réfère l'inscription graphologique démarcative et non à la mise en place d'une amorce de structure énumérative. En revanche, un certain nombre de marques typodispositionnelles permettent de distinguer, à travers l'amorce et potentiellement l'empreinte spécifique de l'énumération qui suit, globale de l'OT en terme de structure informationnelle. La valeur de cette explicitation sera attribuée par le lecteur en fonction du contexte situationnel de la prise d'information. Il s'agit donc de préciser notre position sur ce dernier point.

Une des origines des réflexions autour de la problématique que nous abordons est attachée à un projet soutenu par la Région Midi-Pyrénées dont l'objectif final est l'installation, dans des centres sociaux, de bornes interactives destinées à fournir des informations accessibles aux personnes âgées. Les questions potentielles qui peuvent être soumises à ce type de borne sont ouvertes, mais pas spécifiquement vers une réponse sous la forme de structure énumérative. Par exemple, la réponse « les 4 lieux où

l'on peut retirer un dossier de demande pour l'APA sont ...» proviendra plutôt d'une question du type « où puis-je retirer un dossier de demande pour l'APA ? » que du type « quels sont les 4 lieux où l'on peut retirer un dossier de demande de l'APA ? ». Nous verrons que ce fait a des implications sur le choix de la structure informationnelle que nous avons imposée à nos structures énumératives puisqu'elle nous permettra de considérer la construction d'une structure énumérative comme la composition d'un topique suivi d'une série de rhèmes.

Si l'on observe maintenant la structure énumérative dans son architecture interne, nous remarquons qu'elle exploite un ensemble de marques qui permettent de visualiser les transitions entre les divers OT qui la composent, et qui relèvent donc de l'inscription graphologique démarcative. Nous l'avons vu à travers l'exemple de la discussion générale de la partie précédente : ces marques permettent de matérialiser différents cadres du discours.

Par la suite, ces différentes remarques justifieront l'intérêt méthodologique de notre taxonomie pour l'opérationnalisation de transformations de la FDI vers la FPC.

VI.1.2 Le langage de représentation

Nous avons évoqué, lors de la présentation synoptique de l'architecture d'un système d'oralisation dont l'unité est l'OT que, dans l'état actuel des connaissances en linguistique textuelle, les deux premières étapes semblent difficilement abordables de manière computationnelle en considérant des documents « tout venant ». En effet l'étape de segmentation devrait tenir compte d'une variabilité quasi illimitée des textes et celle d'extraction implique une analyse à des niveaux à la fois syntaxiques, sémantiques et pragmatiques pour la reconnaissance automatique des paramètres formels associés aux OT. Ces deux phases seront donc effectuées « manuellement » de façon à obtenir une représentation adéquate de l'OT qui sera fournie en entrée du système à développer. En raison de son pouvoir expressif et des outils de traitement qui y sont associés, le langage XML⁸ nous semble adéquat comme descripteur du langage de représentation des OT fourni en entrée de la phase de reformulation.

Nous devons ici répondre à une objection naturelle qui pourrait être formulée vis à vis des propos précédents sachant qu'un des objectifs applicatifs que nous voulons faciliter dans notre démarche est l'oralisation de documents *Web* : ne pouvons-nous pas observer les possibilités déjà offertes par le langage de représentation utilisé pour ce type de document : XHTML (eXtended HyperText Markup Language) ?

⁸ eXtended Markup Language

XHTML a été conçu pour présenter l'information visuellement ; les recommandations du WAI⁹ sur la manière d'utiliser les marqueurs de représentation d'objets du texte particulièrement sensibles à leur accès non visuel (tableaux, titres, images, liens ...) ne sont pas, ou sont peu, respectées par la plupart des concepteurs de pages *Web*. Ainsi, dans une perspective d'oralisation, apparaissent un certain nombre de difficultés pour la prise en compte de ces marqueurs :

- la mauvaise utilisation de leur fonction usuelle : l'exemple des tableaux nous paraît convaincant de ce point de vue puisqu'ils sont de plus en plus utilisés pour contrôler la disposition des différents éléments de la page ; pourtant, l'oralisation devrait être bien sûr différente selon que l'objet conserve sa valeur sémantique classique (mettre en lien des données à travers différentes entrées) ou non (organiser des données sans véritables liens de nature tabulaire) ;
- le manque d'utilisation des marqueurs de la structure logique du texte : les concepteurs de page *Web* sont plus attentifs à l'aspect visuel de la page qu'à la représentation interne de sa structure logique. Par exemple, peu de concepteurs utilisent les marques réservées pour spécifier les titres (<H1>, <H2>, ..., <H6>) ; la plupart préfèrent utiliser un marquage qui influence directement l'aspect du texte (comme « centré », « gras », « italique », ...) ;
- la précision insuffisante de certains objets : les énumérations sont par exemple représentées par les deux marqueurs et qui permettent respectivement de déterminer le début et la fin de l'énumération et des items qui la compose ; mais, nous l'avons vu, les structures énumératives sont construites sur la base d'une segmentation plus complexe (avec l'amorce) et de paramètres formels plus nombreux qu'il resterait toujours à extraire.

Nous construirons donc notre propre DTD (Data Type Document) XML sur la base de la segmentation et des paramètres formels décrits dans la Partie AIII.7.2. Puis nous la ferons évoluer au fur et à mesure des étapes décrites plus loin.

Pour la troisième étape (la reformulation), le cheminement **OT → FDI → FPC** sera celui que nous considérerons pour obtenir, dans la suite de cette section, les formes finales que l'application de nos stratégies de reformulation doit fournir. Cela dit, une fois que les considérations théoriques auront produit les éléments nécessaires à l'établissement des deux formes interprétatives tournées vers la discoursivité ou vers la prosodie, nous serons capables de construire une seule représentation de l'OT qui

⁹Web Accessibility Initiative (www.w3.org/WAI)

pourra, selon la stratégie appliquée, conduire à l'une ou l'autre des deux formes. Ainsi le cheminement théorique diffère du cheminement informatique que traversera notre OT, puisque sa représentation devra finalement permettre au choix une transformation selon un schéma **OT** → **FDI** ou **OT** → **FPC**.

D'un point de vue informatique, il s'agira de transformer notre représentation source de l'OT en deux représentations cibles exploitables par un logiciel de TTS. De plus, afin que les principes établis par les feuilles de transformation (XSLT¹⁰) soient utilisables, les langages cibles de représentation devront également être de type XML.

VI.1.3 De l'OT normalisé et ses paramètres formels à la FDI

La normalisation impose de s'être fixé un *pattern* prédéfini de métaphrases interprétatives susceptibles de recevoir la reformulation de n'importe quelle structure énumérative écologique.

Un seul *pattern* ne pouvant regrouper à la fois les énumérations parallèles et non parallèles nous nous concentrerons sur les premières, intuitivement plus fréquentes ; c'est à dire dont les items sont en relation paradigmatique, homogène et isolée.

L'inscription graphologique distinctive d'une structure énumérative parallèle de N items résulte de l'effacement d'une mise en forme matérielle de nature lexicosyntaxique dans une série de métaphrases formulables selon le métadiscours (Cf. Partie AIII.6) suivant :

L'auteur crée un texte identifié texte (1).

L'auteur distingue une énumération identifiée énumération(1).

L'auteur agence énumération(1) en N items identifiés [item(i) et]_{0<i<N} et item(N).

L'auteur chapeaute énumération(1) par un chapeau identifié chapeau(1).

L'auteur attache UT(1) à chapeau(1).

L'auteur compose chapeau(1) de UT(1).

[L'auteur attache UT(i+1) à item(i)]_{0<i≤N}.

[L'auteur compose item(i) de UT(i+1)]_{0<i≤N}.

L'auteur compose texte(1) de chapeau(1) [puis item(i)]_{0<i≤N}

¹⁰ Extensible Stylesheet Language Transformation

Dans l'objectif de produire un métadiscours le plus naturel possible pour constituer le texte oralisable nous ne retiendrons du métadiscours, dans le cadre de notre problématique particulière, que les éléments qui ne peuvent être rendus implicites :

- (1) *L'auteur distingue une énumération identifiée énumération(1).*
- (2) *L'auteur agence énumération(1) en N items identifiés [item(i) et] $_{0 < i < N}$ et item(N).*
- (3) *L'auteur chapeaute énumération(1) par un chapeau identifié chapeau(1).*
- (4) *L'auteur attache UT(1) à chapeau(1).*
- (5) *[L'auteur attache UT(i+1) à item(i)] $_{0 < i \leq N}$.*

Pour une raison d'intelligibilité plus grande du texte final nous avons choisi d'éliminer les différents identifiants et de les remplacer par une fusion des métaphrases qui rendent encore une fois implicite un certain nombre d'éléments du métadiscours original : (1) et (4) deviennent inutiles à notre nouveau texte, (2) et (3) fusionnent ; nous avons également paraphrasé pour cette même raison trois termes : *chapeauter* se traduira par *annoncer ci-après*, *items* par *points* et *attacher* par *préciser que* ; de plus, une information visuelle n'est pas prise en compte par ce métadiscours : le fait que le dernier item, lors d'une lecture de l'énumération, est en train d'être abordé entraîne une formulation particulière pour la N^{ème} métaphrase représentée par (5). Ainsi, nous obtenons :

L'auteur annonce ci-après une énumération de N points.

[L'auteur précise que le I^{ème} point est UT(i)] $_{0 < i < N}$

L'auteur précise que le N^{ème} et dernier point est UT(N).

Nous considérerons également que l'explicitation de l'acteur textuel est inutile car dans notre cas il s'agira toujours de l'auteur du texte. Ceci implique un certain nombre de refontes syntaxiques pour effacer « auteur » et l'acte performatif qui lui est rattaché. De plus, l'obligation d'utiliser un classifieur générique (« point ») est trop contraignante car l'oralisation des structures énumératives lorsque celui-ci est précisé de manière moins générique entraînerait une perte d'information. Nous nous imposons donc de repérer ou expliciter le segment de texte classifieur (que nous noterons **UTC**) des structures énumératives à oraliser (ceci implique de vérifier certains accords en genre). Notre *pattern* sera modifié de la manière suivante :

Les N UTC sont énuméré(e)s ci-après. (Le premier, La première) UTC est UT(1). ...(Le, La) ième (1 < i < n) UTC est UT(i)... (Le, La) Nème et (dernier, dernière) UTC est UT(N).

Deux remarques viennent encore étoffer ce *pattern* :

- de manière générale l'exhaustivité d'une énumération n'est pas précisée et ne produira donc aucune modification dans le *pattern* précédent. Cela dit, elle peut parfois être explicitée par un matériel lexical particulier. La non exhaustivité peut également être parfois inscrite par un item supplémentaire de rang quelconque dont le contenu informationnel est réduit à « ... ». Jouer sur la présence de « Les » devant le numéral de l'amorce permet de modifier le sentiment d'exhaustivité ; « parmi d'autres » sera utilisé lorsque l'explicitation verbale ou l'inscription graphologique d'un item est tournée vers la non exhaustivité ;
- de la même manière, l'absence d'ordre ne peut être explicitée de manière non univoque que verbalement mais l'existence d'une logique d'ordonnement peut être déduite par la présence combinée d'un système de numérotation particulier (romain, arabe, grec, alphabétique, ...) et d'un type de classifieur précis (étapes, phases, ...). Ainsi, lorsqu'il est possible de préciser le critère d'ordonnement, nous augmenterons le *pattern* du terme « dans l'ordre » ou « dans le désordre ». Le cas échéant, ces différentes possibilités de dérivation du *pattern* conduiront à jouer sur l'aspect défini ou indéfini des articles placés devant les termes précisant le rang des items : la neutralité ou l'absence explicite d'ordre impliquera un article indéfini et l'explicitation de l'ordre entraînera un article défini.

Ainsi, nous pouvons proposer 9 dérivations différentes en fonction de la combinaison de la valeur d'ordre et d'exhaustivité des items. Si nous notons :

- **NE** : Neutralité de la MFM de la structure énumérative du point de vue de l'exhaustivité des items ;
- **NO** : Neutralité de la MFM de la structure énumérative du point de vue de l'ordre des items ;
- **E** : MFM autorisant une prise de position en faveur de l'exhaustivité des items de la structure énumérative ;
- **O** : MFM autorisant une prise de position en faveur de l'ordonnement des items de la structure énumérative ;
- **∩E** : MFM autorisant une prise de position en défaveur de l'exhaustivité des items de la structure énumérative ;

- $\neg O$: MFM autorisant une prise de position en défaveur de l'ordonnement des items de la structure énumérative,

et exprimons les possibilités de variation du *pattern* par :

- **DO** : présence de l'expression « dans l'ordre » ;
- **DD** : présence de l'expression « dans le désordre » ;
- \neg **DODD** : absence des expressions « dans l'ordre » et « dans le désordre » ;
- **AI** : article indéfini devant les termes précisant le rang des items ;
- **AD** : article défini devant les termes précisant le rang des items ;
- **Les** : présence du déterminant « les » en tête d'amorce ;
- \neg **Les** : absence du déterminant « les » en tête d'amorce ;
- **PA** : présence de l'expression « parmi d'autres » ;
- \neg **PA** : absence de l'expression « parmi d'autres »,

nous pouvons alors exprimer les 6 contraintes suivantes évoquées dans la remarque précédente :

$$\begin{array}{l}
 (1) \quad O \rightarrow DO \wedge AD \\
 (2) \quad \neg O \rightarrow DD \wedge AI
 \end{array}
 \left| \begin{array}{l}
 (3) \quad E \rightarrow Les \wedge \neg PA \\
 (4) \quad \neg E \rightarrow \neg Les \wedge PA
 \end{array} \right|
 \begin{array}{l}
 (5) \quad NO \rightarrow \neg DODD \wedge AI \\
 (6) \quad NE \rightarrow \neg Les \wedge \neg PA
 \end{array}$$

Les 9 dérivations sont alors obtenues en mettant en relation les 9 combinaisons possibles entre les deux ensembles (**NE**, **E**, \neg **E**) et (**NO**, **O**, \neg **O**) et les contraintes associées (1) à (6) :

$$\begin{array}{l}
 (NE \wedge NO) \rightarrow (\neg Les \wedge \neg PA \wedge \neg DODD \wedge AI) \\
 (NE \wedge O) \rightarrow (\neg Les \wedge \neg PA \wedge DO \wedge AD) \\
 (NE \wedge \neg O) \rightarrow (\neg Les \wedge \neg PA \wedge DD \wedge AI) \\
 (E \wedge NO) \rightarrow (Les \wedge \neg PA \wedge \neg DODD \wedge AI) \\
 (E \wedge O) \rightarrow (Les \wedge \neg PA \wedge DO \wedge AD) \\
 (E \wedge \neg O) \rightarrow (Les \wedge \neg PA \wedge DD \wedge AI) \\
 (\neg E \wedge NO) \rightarrow (\neg Les \wedge PA \wedge \neg DODD \wedge AI)
 \end{array}$$

$(\exists E \wedge O) \rightarrow (\exists Les \wedge PA \wedge DO \wedge AD)$

$(\exists E \wedge \exists O) \rightarrow (\exists Les \wedge PA \wedge DD \wedge AI)$

Notons que nous avons pu passer de la formule logique $(A \rightarrow B) \wedge (C \rightarrow D)$ à la formule $(A \wedge C) \rightarrow (B \wedge D)$ car elles sont équivalentes lorsque A et C sont tous deux vrais, et la dérivation adéquate sera bien sûr choisie en fonction de ce critère.

Remarquons également que pour être utilisable dans notre *pattern*, la tête syntagmatique de l'UTC doit avoir la catégorie syntaxique de *nom* ; la production de l'UTC peut donc impliquer la nominalisation d'un syntagme verbal issu de l'amorce écologique. De plus, la reprise complète de cette UTC dans les items peut-être très lourde lorsque cette unité textuelle est de grande taille. Nous l'avons donc réduite dans les items au seul nom en tête de l'UTC, mis au singulier (Cf. exemple ci-après).

Une dernière remarque concerne l'utilisation systématique du verbe « être » pour introduire les UTI. Ce choix semble quelque peu rigide car, en raison du *pattern* syntaxique choisi, la tête de l'hyponyme du classifieur dans les UTI devra être, le cas échéant, reformulée sous la forme d'un nom ou d'un verbe à l'infinitif :

- dans le cas d'un nom au pluriel, le verbe être au présent fonctionne mal : « **une** deuxième UTC **est les ...** ». Le passage à la troisième personne du pluriel empire le phénomène : « **une** deuxième UTC **sont les ...** ». Nous proposons de permettre l'utilisation de verbes ayant une dynamique ensembliste : « **une** deuxième UTC **comprend les ...** » ;
- dans le cas du verbe à l'infinitif, « est de » sera plus approprié que « est ».

Ainsi, la version définitive du *pattern* des structures énumératives parallèles normalisées manuellement sera :

(Les) N UTC (, parmi d'autres,) sont énuméré(e)s ci-après (dans l'ordre, dans le désordre). (Le, Un, La, Une) (premier, première) UTC (est, comprend, est de) UT(1). {(Le, Un, La, Une) I^{ème} UTC (est, comprend, est de) UT(i)}_{1 < i < N} (Le, Un, La, Une) N^{ème} et (dernier, dernière) UTC (est, comprend, est de) UT(N).

Les parenthèses expriment la présence possible d'un des termes de la liste qu'elles bornent.

Comme nous l'avons précisé, le travail des modules de segmentation et d'extraction des paramètres formels doit à terme nous fournir une représentation respectant une DTD XML de manière à ce que sa transformation soit possible dans une quelconque des formes discursives dérivables de ce *pattern*.

Nous présentons ci-dessous, à travers un exemple, la DTD écrite (Cf. Annexe A) qui nous autorise une telle représentation (Maurel, 2003c).

Ainsi, la structure énumérative

Le Centre Social intervient entre autres dans :

- **la gestion et le suivi de 6 appartements d'urgence que propose la commune ;**
- **l'organisation de réunions d'information collective ;**
- **les étapes du montage de projets de formation.**

sera représentée par

```
<SE type="parallèle" exhaus="non" ordon="neutre">
<AMORCE genre_classifieur= fem nom_classifieur= intervention >
    interventions du Centre Social
</AMORCE>
<ENUMERATION>
<ITEM cs= nom_sing > la gestion et le suivi de 6 appartements d'urgence que propose la commune </ITEM>
<ITEM cs= nom_sing > l'organisation de réunions d'information collective </ITEM>
<ITEM cs= nom_plur > les étapes du montage des projets de formation </ITEM>
</ENUMERATION></SE>
```

qui est nécessaire et suffisante pour produire le texte suivant :

Trois interventions du centre social, parmi d'autres, sont énumérées ci-après. Une première intervention est **la gestion et le suivi de 6 appartements d'urgence que propose la commune**. Une deuxième intervention est **l'organisation de réunions d'information collective**. Une troisième et dernière intervention comprend **les étapes du montage de projets de formation**.

La signification de la mise en forme utilisée dans ce nouveau texte est la suivante :

- le texte en caractère maigre romain est la partie invariable du *pattern* ;
- le texte en gras est récupéré directement à partir du texte encadré par les balises AMORCE et ITEM ;
- le texte en italique est déduit de la valeur des attributs par recopie (reprise du classifieur) ou application de règles adéquates (gestion de l'exhaustivité, de l'ordonnancement ou de l'absence de prise de décision pour la valeur de ces paramètres) ;
- le texte souligné est obtenu par des opérations externes à la structure : comptage du nombre total de balises ITEM et du nombre de balises ITEM déjà prises en compte.

Nous considérons que ce type de texte peut être fourni à un logiciel de synthèse de la parole qui l'oralisera correctement en se basant sur une unité phrastique classique.

VI.1.4 De la FDI à la forme prosodique en contexte (FPC)

La première phase de ce travail était de mettre à jour les différentes méthodes possibles pour formaliser les marqueurs prosodiques des structures énumératives afin de les mettre en relation avec les marques de MFM lexico-syntaxiques de la forme discursive.

VI.1.4.1 De la difficulté d'utiliser une méthode par analyse de corpus

Une idée était de faire jouer par des sujets la situation d'oralisation de textes en faisant varier la forme des énumérations dans le texte. Un premier groupe oraliserait des énumérations discursivement développées telles que nous les avons définies dans la section précédente, un autre oraliserait ces mêmes énumérations mais réduites grâce à leur inscription graphologique. La comparaison de la prosodie utilisée par ces deux groupes devait permettre de faire émerger des marques prosodiques spécifiques à l'oralisation des structures énumératives en fonction de leur degré de discursivité et de potentiellement faire émerger une relation de suppléance entre lexico-syntaxe et prosodie.

Si cette approche avait l'avantage de permettre le contrôle par l'expérimentateur du matériel présenté, elle impliquait en même temps un aspect peu écologique de la situation et donc le risque d'observer un échantillon peu représentatif de stratégies prosodiques.

Une deuxième possibilité était de s'appuyer sur un recueil pré-existant présentant les caractéristiques suivantes :

- variabilité des locuteurs pour l'accès à de multiples stratégies ;
- situation de parole lue à partir de textes : en effet, la situation de lecture est celle qui se rapproche le plus de la situation de synthèse finale : il s'agit bien d'oraliser des textes écrits pour être lus et non de synthétiser « une parole spontanée » ;
- compétence des locuteurs dans le domaine de la lecture ;
- quantité importante d'énumérations ;
- bonne qualité acoustique des enregistrements ;
- taille suffisante du corpus.

Une première ébauche de corpus s'appuyait sur un recueil radiophonique de textes lus par des non professionnels de ce média, puisque cette situation avait l'avantage d'être la plus proche de celle désirée : le lecteur ne pouvait, à l'instar de la synthèse de parole, ni produire de signes extérieurs (mimiques, gestes), ni s'adapter en fonction d'une assistance perplexe, bruyante, etc., ni être influencé par une prosodie trop caractéristique des professionnels de la radio. Les seuls documents sonores que nous ayons pu obtenir, répondant *a priori* à ces caractéristiques, concernaient des discours d'hommes et de femmes politiques, proposés très régulièrement dans l'émission *France Europe Express* au cours de la campagne présidentielle de 2002. Ce corpus a rapidement été abandonné car si le professionnel de la radio propose une prosodie spécifique, elle reste relativement normalisée et donc assez neutre ou unilatérale du point de vue de son interprétation (c'est peut-être là l'effet escompté en terme de déontologie journalistique). En revanche, elle est souvent tout aussi spécifique chez nos responsables politiques mais pas au niveau du genre du discours ; il s'agit plutôt d'une variabilité interindividuelle poussée à son paroxysme (puisque le souhait de neutralité du discours va bien sûr dans un sens contraire des préoccupations du locuteur) qui use et abuse de tonalités dramatisantes, emphatiques, ironiques et presque publicitaires.

Finalement, c'est le recueil audio des conférences de l'« Université de Tous les Savoirs » (UTLS) qui a semblé le plus approprié pour la constitution d'un corpus d'énumérations. Celui-ci est composé d'enregistrements audio de conférences présentées dans le cadre d'une initiative ministérielle pour l'an 2000, « l'Université de tous les savoirs ». 262 conférences sont disponibles à l'adresse <http://www.telerama.fr>. Chaque conférence dure entre 50 min. et 1h. Sont disponibles sur ce site la liste des

thèmes et sous-thèmes, la liste des conférenciers par sous-thème, le fichier « son » de la conférence, le fichier « texte » (mais sa présence n'est pas systématique) et un résumé de la conférence. Aucune précision n'est donnée sur la situation de production du texte : qui l'a écrit et quand (avant ou après la conférence et dans ce dernier cas le conférencier lui-même ou bien un transcripneur ?). Nous avons considéré que ce type de conférence officielle avec un très large public (en terme de quantité et de qualité) limitait à la fois l'abus de techniques trop axés sur la persuasion de son auditoire, les influences réciproques des attitudes du public et du conférencier et l'impact de l'environnement sonore sur la prosodie du discours.

Afin de confirmer l'adéquation du recueil UTLS pour la constitution d'un corpus d'étude adapté à nos objectifs, nous avons procédé à une première écoute des conférences avec l'objectif de catégoriser grossièrement les énumérations rencontrées au niveau de leurs paramètres formels et d'évaluer ainsi leur variabilité. Nous avons pu également à cette étape sélectionner uniquement les conférences pour lesquelles nous étions perceptivement persuadés qu'il s'agissait d'un texte lu.

Or, nous avons ainsi remarqué que la très grande variabilité des structures énumératives rencontrées rendait le contrôle des paramètres formels impossible. Renoncer à ce contrôle et comparer la réalisation prosodique d'un ou deux paramètres formels dans des données qui varient sur l'ensemble des autres paramètres formels, engendrait une limite importante pour faire émerger des régularités significatives en raison d'une potentielle interaction au niveau de la réalisation prosodique des différents paramètres formels. Ainsi, des conclusions auxquelles nous sommes parvenu au terme de la pré-analyse se dégage principalement l'obligation de concentrer notre étude sur 2 types de structures énumératives ne se distinguant que sur un seul paramètre, toutes choses égales par ailleurs (ex. par la présence ou non du groupe nominal spécifique dans l'amorce ou bien par le marquage ou non des items à l'aide d'un organisateur textuel).

Sur un plan théorique, la restriction de l'analyse à un ou deux paramètres limitait beaucoup la portée de notre étude : se limiter à un corpus qui ne nous permet d'étudier qu'un à deux paramètres discursifs, entraîne une forte proximité des deux formes de la structure énumérative sur le continuum opposant forme discursive et forme prosodique ; affaiblissant ainsi la probabilité de faire émerger des résultats intéressants et réutilisables au regard du travail important de traitement des données à mettre en œuvre.

La démarche par corpus n'était donc, de notre point de vue, pas appropriée à nos objectifs d'étude. Si cette démarche est adéquate pour le test d'hypothèses précises, elle ne l'est pas compte tenu de la complexité des objectifs que nous poursuivons. Partant de ce constat nous avons choisi, pour dresser l'inventaire des marques prosodiques

fonctionnellement équivalentes aux marques lexico-syntaxiques de la forme discursive, d'adopter une approche séquentielle pluri-fondée :

- étude du recueil de l'UTLS, non dans le cadre de la constitution d'un corpus, mais dans le cadre d'une étude perceptive qualitative des structures énumératives observées : écoutes attentives et critiques ;
- introspection soumise à discussion avec des spécialistes ;
- déduction à partir des deux étapes précédentes de la manière de projeter des règles, dictées par les modèle prosodiques du français, sur la relation entre des marques lexico-syntaxiques d'un OT et sa version prosodique.

Cette dernière étape a rendu nécessaire la recherche d'un modèle prosodique adapté pour supporter sa projection sur notre problématique.

VI.1.4.2 Recherche d'un modèle prosodique adapté

Dans cet objectif, la distinction entre inscription graphologique démarcative et distinctive semble intéressante pour orienter ce choix. Le terme de graphologie nous renvoie dans un premier temps à celui de phonologie et donc nous impose de nous situer parmi les différents concepts que recouvre ce domaine : en particulier, y a t il un seul ou deux mécanismes distincts qui gouvernent le fait prosodique, l'intonation et l'accentuation ?

Étant donné le domaine d'application de l'inscription graphologique nous nous tournerons vers la seconde position (largement la plus admise aujourd'hui) puisqu'elle oppose l'accentuation qui agit au niveau du mot, et l'intonation qui exerce son influence au niveau de la phrase et des constituants syntactico-sémantiques. Ce deuxième aspect nous fait préférer les modèles tournés spécifiquement vers l'intonation.

Si nous nous basons sur une définition telle qu'elle est proposée (à titre provisoire et simplifié dans son introduction) par Mario Rossi :

l'intonation [...] est un système linguistique destiné à organiser et hiérarchiser l'information que le locuteur entend communiquer à l'allocataire ou aux allocataires dans son message, et à linéariser la hiérarchie des structures syntaxiques. (Rossi, 1999 :7)

nous remarquons alors le rapport d'imitation entre

- l'inscription graphologique démarcative et l'intonation qui permet de linéariser la hiérarchie des structures visuelles, en d'autres termes à faire « entendre » les transitions entre des configurations de nature phonologique ;

- l'inscription graphologique distinctive et l'intonation relative à la hiérarchisation de l'information.

De ce fait, selon le statut de l'objet textuel étudié, nous nous dirigerons vers les modèles centrés sur l'intonation syntaxique, censés régir le mode d'association de l'intonation avec la syntaxe, ou sur les modèles centrés sur l'intonation pragmatique dont une des fonctions est de hiérarchiser l'information.

Nous avons vu que la structure énumérative procède des deux systèmes d'inscription graphologique. Il s'agira donc dans un premier temps de diriger nos investigations vers les modèles prosodiques du français qui mettent en relation intonation et construction discursive.

VI.1.4.3 Intonation et constructions discursives

Nous reprenons ici un titre de Lacheret-Dujour (1999) qui consacre un chapitre, dont nous tirons les enseignements de cette sous-section, aux approches qui abordent les fonctions sémantiques et énonciatives de l'intonation.

L'approche phonologique de Piet Mertens (1987)

L'énoncé est segmenté en une succession de groupes intonatifs dont la configuration prototypique est **[[NA]AI] [NA] AF [NA]** où la mise entre crochets exprime la facultativité et les couples de majuscules symbolisent des syllabes **Non Accentuées** ou possédant un **Accent Initial** ou un **Accent Final**.

A chacune des syllabes est associé un ton (représenté par une lettre minuscule pour les syllabes non accentuées et majuscule sinon) sachant que le registre mélodique d'un locuteur se situe entre deux niveaux de hauteur primaire absolue : suraigu (**H+**) et infra-bas (**B-** ou **b-**). Deux niveaux primaires relatifs peuvent être distingués comme bas (**B** ou **b**) et haut (**H** ou **h**) dès qu'ils diffèrent d'au moins un intervalle tonal égal à une tierce majeure. Les tons peuvent subir un rehaussement (/) ou un abaissement (\) mais ils sont alors toujours précédés d'un ton de même hauteur.

Les morphèmes intonatifs (la combinatoire tonale) des accents initiaux et finaux peuvent être à base de tons statiques (H, B sur AI), statiques allongés (B-B-, H+H+, HH, BB sur AF) ou dynamiques (BB-, HB, BH sur AF). La construction intonative d'un énoncé est obtenue par l'application d'une comparaison récursive de gauche à droite d'une règle de dominance entre les différents groupes intonatifs (B- domine H+ qui domine H qui domine B, /X domine X qui domine \X, B-B- domine H+H+ qui domine HB, qui domine BH, qui domine HH, qui domine BB) : deux groupes intonatifs

fusionnent si le ton final associé au dernier domine celui du premier ; sinon on obtient deux entités juxtaposées.

Cela dit, certains morphèmes intonatifs sont marqués parce qu'ils portent en eux-mêmes, ou en se combinant, une fonction sémantique ou énonciative. C'est à ce niveau que le rapprochement avec les fonctions de l'inscription graphologique distinctive peut être faite :

- [B-B-] indique la fin d'une unité au contenu conceptuel homogène ;
- [H+H+] est un marqueur d'emphase ;
- [HB] est un marqueur de focalisation ;
- [HB-] a une fonction contrastive ;
- [BH] a une fonction énonciative centrée sur le récepteur ;
- [h + HH] ou [h + B-B-] apparaît souvent dans les citations ou les énumérations d'éléments présentés comme allant de soi

Enfin, une remarque intéressante du point de vue de l'objet que nous étudions précise que dans ce contexte

la hiérarchie syntaxique est [...] souvent violée par l'organisation intonative en français parlé au profit de regroupements sémantiques, et, au delà de la syntaxe, le principe de dominance est directement lié à l'interprétation sémantique des gestes prosodiques (Lacheret-Dujour, 1999 :175).

L'exemple qui suit cette remarque exprime de tels regroupements sémantiques possibles hors de toute syntaxe classique :

Lundi_{BB} lavage_{BH} mardi_{BB} repassage_{BH} mercredi_{HB} repos_{B-B-}

que, dans une transposition inverse à notre problématique, nous pourrions inscrire comme équivalente à :

- *lundi : lavage ;*
- *mardi : repassage ;*
- *mercredi : repos.*

L'approche phonologique énonciative de L. Danon-Boileau et M.-A. Morel (1995)

Le point de départ du modèle considère un énoncé oral comme composé d'un préambule (lui-même constitué au maximum d'un ligateur, un indice de modalité, un cadre et un sujet logique), d'un rhème (simple ou avec focus) et d'un postrhème (ou ajusteur).

Contrairement au modèle précédent où seuls les accents finaux étaient cruciaux, ce modèle en distingue 5 dans ce type d'énoncé : les hauteurs initiales et finales de l'énoncé, la hauteur finale du préambule, le point le plus haut du rhème et la hauteur finale du rhème. La hiérarchie tonale n'est basée que sur les deux tons « haut » et « bas ».

En terme de niveau intonatif, le point le plus haut de l'énoncé est de niveau 4 et le plus bas de niveau 1, le seuil des niveaux 2 et 3 étant obtenu par un découpage de cette « fenêtre tonale ». Ces niveaux s'interprètent en terme de fonction énonciative : les niveaux 3 et 4 sont utilisés pour se positionner par rapport à la co-énonciation selon une attitude consensuelle ou au contraire en discordance avec celle de l'allocutaire ; le niveau 2 est utilisé pour rompre la co-énonciation ou lorsque celle-ci n'a pas d'enjeu ; le niveau 1 pose une frontière absolue dans l'intervention du locuteur.

Les transitions entre ces différents niveaux dénotent également une valeur énonciative : l'intonation montante exprime la recherche du consensus lorsqu'il est mis en défaut, l'intonation descendante montre un repli sur soi et montre que l'on n'attend plus rien de son interlocuteur ; l'intonation plate marque une rupture dans le déroulement discursif de l'énoncé (cas de rattrapage ou de glose)

Si cette approche paraît difficile à relier directement à l'inscription graphologique distinctive, c'est qu'elle porte sur les particularités énonciatives de l'intonation spontanée qui impliquent, entre les interlocuteurs, une relation dialogique très forte que, par nature, l'écrit n'offre pas. Pourtant elle présente un intérêt pour la description du rhème avec focus (un élément peut être également focalisé par une inscription graphologique particulière) ou lorsqu'elle aborde la prise en charge de l'interprétation énonciative de modalités à l'origine de la distinction d'objets textuels par une inscription graphologique particulière (parenthèses, point d'interrogation ou d'exclamation).

- rhème avec focus : le rhème focalisé est souvent introduit par une forme clivée de type *c'est...que/qui* et veut, par une montée très rapide sur la syllabe finale, inciter l'allocutaire à un consensus de point de vue sur ce point précis du discours ;

- l'incise :
 - permet de corriger un manque dans la mise en place du préambule, marquer un changement de plan énonciatif, d'introduire une glose ou un commentaire explicatif,
 - elle est actualisée par une rupture intonative dans les niveaux bas suivie d'une remontée portée par la syllabe finale qui permet de rattraper le niveau intonatif en cours ;

- la question : s'ajoute à la traditionnelle montée intonative reconnue pour la question fermée, une absence de segment modulé à la suite et une syllabe finale brève. La question ascendante peut également présenter deux montées intonatives qui ont pour effet de séparer le thème et le modus interrogatif. Une question ouverte, donc généralement avec une intonation descendante, peut voir son segment initial particulièrement haut, ce qui exprime en principe un changement de sujet et appelle une réponse thématique qui complète le rhème établi dans la question. Enfin une question porteuse sur la syllabe finale d'un ton dynamique et allongé (dans le même sens que les tons dynamiques sur les accents finaux de Mertens) incite à la convergence vers un point de vue ou un avis personnel qui ne peut être refusé par l'allocutaire ;

- l'exclamation peut relever de trois catégories dans cette approche :
 - à finale haute et allongée elle marque un inachèvement et appelle à une convergence des points de vue,
 - à initiale haute qui fait croire à un point de vue consensuel et finale basse qui le rompt tout de suite après (sorte de formulation ironique),
 - à intonation haute et plate : sorte d'incise mais rendu autonome par rapport à son contexte pour exprimer un véritable étonnement.

L'approche morphologique de Mario Rossi (1999)

Trois modules séquentiels permettent de prédire la structure prosodique des phrases en français parlé aux niveaux sémantico-pragmatique, syntaxique et rythmique. Le premier module qui nous intéresse ici devrait traiter les trois fonctions de l'intonation pragmatique. La fonction appellative qui établit le lien avec l'allocutaire, la fonction expressive qui exprime un contenu de nature expressive et la fonction représentative qui hiérarchise l'information du message. C'est cette dernière qui semble la plus proche de l'inscription graphologique distinctive de l'écrit et dont nous décrivons les principes de fonctionnement selon l'auteur.

De manière générale, l'intonation est décrite à travers la notion de morphème intonatif qui est un signe intonatif dont le contenu a une valeur pragmatique ou syntaxique et dont l'expression, dans la perspective d'une théorie morphologique, ne relève pas d'une combinatoire de phonèmes tonals, mais de contours ou d'ensembles de traits prosodiques constituant des intonèmes. A ces intonèmes sont associés des fonctions linguistiques précises :

- l'intonème de continuatif majeur /CM/ ou /CT/ marque la hiérarchie syntaxique et énonciative ;
- l'intonème de continuatif mineur /cm/ ou /ct/ est un outil de démarcation ;
- l'intonème d'appel /CA/ marque l'opération de thématization ;
- l'intonème de parenthèse /PAR/ marque l'extraction énonciative ;
- l'intonème de conclusif majeur /CC/ est un marqueur syntaxique et rhématique ;
- l'intonème de conclusif mineur /cc/ signifie une disjonction syntaxique.

Ces différents intonèmes permettent de définir les quatre opérations énonciatives proposés par l'auteur : la topicalisation, la thématization, la rhématisation et la focalisation par l'accent énonciatif. Les unités pragmatiques marquées prosodiquement par le jeu de ces opérations sont respectivement les topiques, les parenthèses interne et externe, les rhèmes et les focus.

Il nous semble d'ores et déjà que ces définitions présentent un intérêt majeur dans la perspective de l'oralisation automatique de l'inscription graphologique distinctive : elles proposent un système global et complet, et donc directement utilisable, qui articule un ensemble d'unités pragmatiques du discours oral spontané qui semble pouvoir être confondu avec les unités pragmatiques que peut dégager l'analyse de l'inscription graphologique distinctive d'un texte.

Cela dit, le fait que les modèles décrits proposent la description prosodique de l'oral spontanée, semble les éloigner de notre objectif d'oralisation d'un écrit « réfléchi ». En effet, la forme finale de la FDI résulte de la linéarisation de l'inscription graphologique sous la forme d'un empilement syntagmatique de phrases. Son oralisation par les logiciels de synthèse de la parole sera donc calculée à partir d'une analyse syntaxique classiquement phrastique. Or, la plupart des spécialistes du domaine de l'oral spontané justifient leur modèle en arguant que

l'intonation est toujours étudiée dans le cadre restreint de [...] la phrase. Appliquées aux énoncés d'une conversation, ces descriptions s'avèrent insuffisantes, voire inadéquates (Mertens, 1987 :1)

Cela rendrait probablement difficilement intégrables leurs concepts dans ces technologies, sans opérer des modifications en profondeur des modules d'analyse syntaxique non fondée pour la conversation.

Cependant, le modèle de Rossi réduit cette incompatibilité entre modèles prosodiques de l'oral spontané et oralisation d'un écrit « réfléchi » puisqu'il considère au contraire que

les modèles construits pour décrire la structure syntaxique de la phrase et les outils conceptuels développés pour caractériser sa structure prosodique sont utilisables pour l'étude de la parole spontanée (Lacheret-Dujour, 1999 :193).

Ainsi, l'intonation pragmatique est pour l'auteur la première strate informative à isoler, afin de l'articuler ensuite à une intonation syntaxique basée sur des modèles classiques (le modèle Chomskien du gouvernement et du liage pour l'auteur) : les intonèmes décrits pour l'intonation pragmatique sont donc utilisables hors analyse syntaxique de la phrase et intégrables ensuite en terme de réalisation acoustique en fonction de l'analyse syntaxique choisie par les concepteurs de synthèses de parole.

Enfin, les notions de « hiérarchie syntaxique », de « disjonction syntaxique » et surtout « d'outil de démarcation » que nous trouvons dans la description de certains intonèmes nous permettent d'espérer des recoupements entre les principes régissant le modèle de l'intonation syntaxique de ce même auteur et ceux qui devraient gérer les effacements de mises en forme matérielle de type lexico-syntaxique issues de la reformulation de l'inscription graphologique démarcative.

C'est pour ces raisons que nous choisirons ce modèle afin de proposer les effacements des marques de mise en forme matérielle de nature lexico-syntaxique au profit des intonèmes préconisés. Plutôt qu'une description exhaustive du modèle de Rossi, nous ferons directement référence aux concepts qu'il véhicule en les détaillant, et en les discutant, au fur et à mesure qu'ils sont projetés sur l'oralisation des structures énumératives.

VI.1.4.4 Projection des intonèmes de Rossi aux structures énumératives

Le pattern d'une structure énumérative de N items, obtenu lors de l'élaboration de la forme discursive interprétative est le suivant :

(Les) N UTC (, parmi d'autres,) sont énuméré(e)s ci-après (dans l'ordre, dans le désordre). (Le, Un, La, Une) (premier, première) UTC (est, comprend, regroupe, est de) UTI_1 . {(Le, Un, La, Une) $I^{ème}$ UTC (est, comprend, regroupe, est de) UTI_i .} $1 < i < N$ (Le, Un, La, Une) $N^{ème}$ et (dernier, dernière) UTC (est, comprend, regroupe, est de) UTI_N .

Une part du matériel décrit par ce pattern est susceptible d'être effacé au profit d'une configuration prosodique impliquant les intonèmes de Rossi. Pour utiliser ces derniers et localiser leur lieu de réalisation, il est nécessaire de procéder au découpage de la structure informationnelle de l'OT en unités particulières : topiques, thèmes internes et externes, focus, rhèmes.

Définitions préliminaires

Devant la grande variété de définitions de ces termes, nous proposons ici leur définition tirée de Rossi (1999).

Topique : constituant de tête, dans l'énoncé, chargé d'une valeur contextuelle, qui représente ce dont on parle. Le contenu contextuel peut être référentiel (déictique) ou inférentiel (lorsque de la réponse on peut inférer qu'une question a été posée). Le topique est accompagné d'une marque d'extraposition syntaxique et/ou intonative.

Thème : constituant extraposé à la fin ou à l'intérieur de l'énoncé ; les marques d'extraction sont généralement syntaxiques et intonatives ; le thème externe peut n'être marqué que par l'intonation. Le thème réfère au donné, au connu, ou le présuppose.

Focus : item porté au premier plan par le morphème intonatif dit Accent de Focalisation (AF).

Rhème : noyau d'énoncé obtenu après que le ou les topiques, et le ou les thèmes ont été soustraits ; le rhème est porteur de l'information de premier plan susceptible de modifier les croyances de l'allocutaire.

En fait la différence avec l'acception classique de topique ou de thème est dans la volonté de ne pas amalgamer les notions de *ce dont on parle/ce qu'on en dit*

Le thème (T) est le sujet dit psychologique ou logique ou linguistique, c'est à dire ce dont on dit quelque chose, et le rhème (R) ce qu'on affirme, ce qui est dit sur le thème ; comme le disait Mathésius, l'ordre normal dans ce cas est l'ordre TR. C'est la perspective thématique. (Rossi, 1999 :60)

avec le donné/le nouveau.

T est ce qui est connu, donné par le contexte ou la situation et R contient alors l'information nouvelle. C'est la perspective contextuelle. [...] L'ordre TR n'est pas obligatoire. (Rossi, 1999 :60)

Ici, le thème et le topique sont tous deux *ce dont on parle* mais, alors que le thème réfère uniquement au *donné* le topique acquiert en plus une valeur contextuelle qui peut conférer à ce segment de discours le statut de *nouveau*.

Voyons maintenant comment les différentes substitutions nous guident dans l'application de ces définitions aux structures énumératives (Maurel, 2003d).

Substitution du numéral et du classifieur de l'amorce

Le numéral ne sera pas substitué en raison de l'information sémantique qu'il porte en sus de sa fonction architecturante et pour laquelle aucun moyen oral équivalent n'existe : on ne peut exprimer une valeur numérique précise par l'intonation. Le monde des possibles qu'offre chaque support lui est spécifique et ces mondes ne se recouvrent que partiellement entre supports (écrit, discursif et prosodique, voire gestuel, ...) ; ainsi, la prosodie ne peut exprimer une quantité que de façon relative tandis que l'écrit peut fournir l'information exacte. Le phénomène inverse peut être d'ailleurs constaté : la prosodie peut remplir une fonction « émotionnelle » que l'inscription graphologique peine à remplir, bien qu'il soit possible de noter le développement récent de symboles expressifs, tels que les *smileys* dans les courriers électroniques, dont le rôle semble être de pallier ce déficit pourtant intéressant dans des contextes où l'écrit veut conserver une dimension très dialogique.

Le fait de conserver le numéral implique la conservation du classifieur pour des raisons syntaxiques (de plus, le classifieur est l'objet de la même remarque que le numéral lorsqu'il n'est pas générique).

Ainsi, nous tirons deux enseignements de cette observation :

- la première conforte l'intérêt de transiter par une reformulation discursive pour obtenir la version prosodiquement marquée d'un OT : certains des éléments restitués par la première étape ne pourront plus être substitués lors de la seconde ;
- le deuxième enseignement concerne les répercussions sur la décomposition de la structure informationnelle de la forme prosodique de nos structures énumératives : le matériel utilisable comme support des intonèmes est augmenté du numéral et du classifieur.

Substitution de « sont énuméré(e)s »

Le groupe verbal « sont énumérées » annonce de manière anticipée dans le pattern choisi :

- par sa construction syntaxique passive, la partie de l'amorce qui est concernée directement par l'énumération : le classifieur ;
- par sa morphologie et sa classe sémantique, la pluralité de ces informations (le participe passé du verbe « énumérer » est très rare au singulier) et la co-énumérabilité des hyponymes du classifieur.

La conservation de « sont » permet de conserver le rôle indiqué par le premier point. L'amorce est ainsi rendue incomplète, comme elle pouvait l'être par l'utilisation des deux points pour mettre en facteur les items à l'écrit, mais les implications à l'oral sont, de notre point de vue, plus importantes puisqu'elles vont entraîner plusieurs refontes intonatives. En effet, il semble qu'étant donné

- la dislocation syntaxique provoquée par l'incomplétude de l'amorce ;
- le fait que l'amorce, dans ce contexte, référence ce dont on va parler ;
- que la situation d'énonciation choisie confère plutôt un statut de *nouveauté* aux informations de l'amorce,

la mise en amorce incomplète joue le rôle d'opération de topicalisation, c'est-à-dire, selon Rossi, d'extraction en tête : dans notre pattern, le syntagme nominal sujet qui précède « sont » sera terminé par le morphème intonatif de topique **CT0**.

De plus, pour compenser l'effacement de « énuméré(e)s » il s'agira de focaliser l'importance des termes de l'amorce liés à la pluralité et la co-énumérabilité : le numéral et le classifieur. Chez Rossi, deux stratégies sont envisageables pour focaliser un élément : (1) la rhématisation et (2) la focalisation marquée par un **Accent de Focalisation (AF)**. Étant donné la remarque précédente, le statut de topique affecté à l'amorce, et donc en particulier au numéral et au classifieur, nous interdit d'utiliser la rhématisation. Nous étudierons donc la possibilité d'utiliser **AF** sur le numéral.

Précisons tout d'abord que la structure accentuelle de la phrase pose des accents virtuels (accents lexicaux sous-jacents – **as**) qui peuvent participer à la structure intonative selon qu'ils sont réalisés ou non par des morphèmes intonatifs. Ainsi, d'après Rossi :

1. Si **AF** est amalgamé avec **as**, sa portée est circonscrite au mot domaine de l'accent lexical et a valeur de proéminence sémantique ; l'intonème continuatif qui clôt le domaine est inversé ;

2. Si **AF** est initial (donc non amalgamé avec **as** en français) :

- **2.1.** si l'intonème continuatif qui clôt le domaine est inversé alors **AF** a une valeur contrastive,
- **2.2.** si l'intonème continuatif qui clôt le domaine n'est pas inversé alors **AF** a une valeur focalisatrice pleine ; dans ce cas la portée focalisatrice de **AF** est le domaine circonscrit par **AF** et le continuatif. La portée de **AF** se confond alors avec son domaine et détermine un **focus étendu**. Ce domaine est inclus dans un arc prosodique. Le contenu de l'arc prosodique est intégralement focalisé,
- **2.3.** si l'intonème qui clôt le domaine de **AF** est le conclusif **CC** (marqueur de la rhématisation) alors la portée de **AF** est circonscrite au mot porteur de **AF**.

Dans le cas de **AF** sur le numéral de notre pattern, nous pouvons dire que, celui-ci étant monosyllabique (en tous les cas jusqu'à 13 inclut), **AF** est à la fois à l'initiale et amalgamé avec l'accent lexical sous-jacent **as** ; de plus le premier continuatif à droite est sur l'**as** porté par le classifieur.

Nous éliminons donc le cas **2.3.** qui ne correspond pas à notre pattern ; ainsi que le cas **1.** qui n'est pas approprié car il sous-entendrait que le numéral, de par sa valeur particulière, est étonnant ou remarquable. Bien qu'envisageable, le cas **2.1.** mettrait en contraste le numéral mais avec l'effet secondaire de reléguer à l'arrière plan le nom classifieur : ceci engendrerait potentiellement un conflit avec l'objectif d'anticipation de la co-énumérabilité des hyponymes du classifieur.

Le cas **2.2.** sera retenu comme le plus adapté puisqu'il permettrait d'insister à la fois sur la pluralité des informations liées au classifieur et sur le classifieur lui-même (et donc globalement à la co-énumérabilité de ses hyponymes).

Remarquons que le choix du morphème continuatif de topique exprimé par **CT0** n'est pas suffisant, car ce morphème peut se décliner comme inférentiel (il devient alors **CTi**) ou référentiel (il devient alors **CTr**). La substitution suivante nous permet de prendre position pour un de ces deux morphèmes intonatifs contextuels.

Substitution de « ci-après »

CTi permet d'inférer qu'une question a été posée : ce morphème permet de séparer dans la réponse ce qui est de l'ordre de la reprise partielle de la question et de la réponse proprement dite. Quant à **CTr**, il montre en quelque sorte de quoi le locuteur va parler

de la même manière qu'il pourrait le montrer du doigt s'il voulait diriger la discussion autour d'un objet se situant à portée de vue.

C'est ce dernier morphème qui nous semble le plus approprié dans le cas de l'amorce des structures énumératives : l'effacement de « ci-après » peut être intonativement contrebalancé par l'utilisation d'un morphème intonatif de topique contextuel référentiel à la fin de l'amorce afin de la charger d'une valeur contextuelle référentielle qui permettra à l'allocutaire de savoir qu'il prend connaissance d'un référent **juste** avant de recevoir une information le concernant.

Les traits et propriétés acoustiques de l'intonème du morphème intonatif contextuel référentiel sont en annexe.

Substitution des termes jouant sur la place et l'ordre

La première remarque concerne la précision dans l'amorce de l'ordonnancement ou non des items. En fait, généralement, aucune information ne sera donnée dans la forme originale de la structure énumérative quant à l'ordre des items et l'option choisie sera de ne pas s'avancer sur ce paramètre très difficile à évaluer lorsqu'il n'est pas explicite. La question ne se pose donc que pour le peu de fois où cette précision est donnée explicitement et donc récupérée dans le *pattern* sous la forme « dans l'ordre » ou « dans le désordre ». Nous n'effacerons pas dans ce cas ces termes.

Les adjectifs ordinaux semblent quant à eux indiquer de manière absolue la position de l'item dans l'énumération et à ce titre, ils paraissent *a priori* non effaçables au profit d'une marque prosodique fonctionnellement équivalente. Cela dit, le nombre exact d'items étant explicité par la conservation du numéral dans l'amorce, il sera peut-être suffisant pour un repérage minimum dans la structure (est-on plutôt en début, au milieu ou proche de la fin de l'énumération ?) d'avoir recours à une combinaison de marques prosodiques indiquant le positionnement relatif des items. Nous remarquerons ici que ce type de marque lexico-syntaxique ne provient pas de la reformulation de l'inscription graphologique distinctive mais de l'inscription graphologique démarcative des items : en effet si la mise en structure énumérative relève globalement de la première, l'organisation interne de cet objet, composé d'autres objets, est gérée par la seconde. Nous ne trouverons donc pas dans le système préconisé pour l'intonation pragmatique du français d'indices pour une telle substitution. Nous nous sommes plutôt basés, pour cela, sur nos écoutes empiriques des recueils déjà évoqués afin de repérer une stratégie prosodique qui semble être utilisée dans l'intention de faire anticiper la quantité totale d'items et ressentir, pour chacun d'eux, le nombre d'items restant à énumérer.

Nous avons perceptivement remarqué plusieurs fois deux phénomènes qui nous semblent en lien avec notre objectif. Une marque prosodique pourrait être une **déclinaison intonative régulière** lors de chaque transition vers un nouvel item ; le dernier item étant prononcé sur la ligne mélodique de base du locuteur ; nous exploitons ici le changement de registre du locuteur au départ du « *downdrift* » (ou ligne de déclinaison : phénomène physiologique d'abaissement mélodique progressif observé sur l'énoncé) de chaque item (Figure 12 : 3 premiers items). Cette stratégie pourrait être accompagnée d'une « **prise d'inspiration** » en début d'énumération dont la force dépendrait de la longueur de l'énumération.

De la même manière la suppression de « dernier » est de nature organisationnelle et nos observations nous ont conduit à choisir comme la plus remarquable perceptivement, une intonation particulière pour l'ensemble du dernier item : celui-ci sera naturellement terminé par un conclusif mais sa pente sera étirée de manière à être plus forte que celle des autres items.

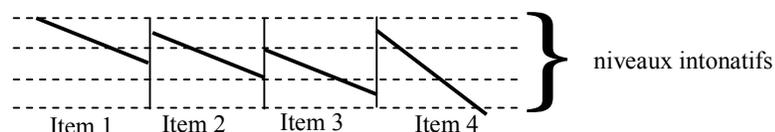


Figure 12 : Changements de registre à chaque déclinaison intonative

Enfin, une stratégie pausale adaptée devra soutenir l'organisation interne de la structure énumérative, puisque la pause est un moyen fort de regroupement des unités : nous considérerons qu'il existe une relation entre la hiérarchie des objets du texte et la stratégie pausale utilisée pour en rendre compte. Autrement dit, en déclinant l'objet du texte *Structure Enumérative* comme composé de deux objets (*Amorce* et *Enumération*), l'énumération étant elle-même composée d'objets appelés *Items*, nous utiliserons une durée de pause de fin d'objet d'autant plus courte que l'objet considéré est bas dans cette « architecture ». Ainsi, la pause après « sont » de l'amorce sera inférieure à la pause entre paragraphes mais supérieure à la pause de fin d'item ; cette dernière devra rester malgré tout supérieure à la pause entre deux phrases.

Substitution du groupe impliqué par la reprise du classifieur

Il s'agit en premier lieu de donner un statut, en terme de structure informationnelle, à la partie de l'item introduite par le groupe impliqué dans la reprise du classifieur. L'effacement de ce dernier conduira à une série de syntagmes mis en facteur par « sont » et la stratégie pausale. Cela dit, ce groupe mettrait également l'accent sur les

syntagmes contenant les hyponymes du classifieur qui sont, dans la situation d'énonciation choisie, les noyaux informationnels de l'item. Les items pourront donc être considérés comme une succession de rhèmes eux mêmes composés d'un « noyau rhématique » (l'unité minimale contenant l'hyponyme du classifieur) et, le cas échéant, d'autres éléments qui seront considérés selon leur place par rapport au noyau rhématique comme des topiques ou des thèmes.

Dans une langue comme le français, l'identification de l'information nouvelle est fondamentalement opérée par le morphème intonatif conclusif CC, marqueur de la rhématisation. Selon Rossi, les procédés d'organisation énonciative autorisent trois modes de présentation de l'information nouvelle :

- l'indication : CC a pour rôle d'indiquer le nouveau dans le matériau linguistique qui le précède. Ce mode est donné par construction, c'est à dire par le prédicat verbal présent devant CC. Par l'indication le rhème présente l'acte de prédication ;
- l'identification : lorsque le prédicat verbal est extraposé dans les structures Rhème/Thème. Les arguments ou les adjoints devant CC sont présentés non plus par simple indication, ils sont identifiés hors de l'acte de prédication. L'identification est un procédé de reconnaissance à la fois par indication et par contraste, mais où domine l'indication. L'identification est un mode d'indication marqué ;
- la sélection exclusive : la sélection exclusive, associée généralement à la notion d'accent contrastif, évoque immédiatement les constructions clivées.

Le mode de présentation dans notre *pattern* semble être celui de l'identification : CC $\Theta\epsilon$ en fin du noyau informationnel de l'item (syntagme contenant l'hyponyme du classifieur) aurait ainsi pour rôle d'identifier les arguments du prédicat extraposé (« sont ») non pas syntaxiquement par une construction Rhème/Thème, mais intonativement par sa mise en facteur dans l'amorce. L'opérateur d'abaissement $\Theta\epsilon$ entraîne dans ce contexte une mélodie plate à la hauteur de la fin de CC.

Nous ne détaillerons pas ici les intonèmes qui peuvent être utiles à ce niveau pour marquer les rhèmes, thèmes internes et thèmes externes mais renvoyons aux annexes et à Rossi (1999).

Implications pour la gestion de la fin des items

Deux cas peuvent survenir. Le premier se réalise lorsqu'il y a congruence entre la fin du noyau rhématique et la fin de l'item. L'item est donc terminé par les morphèmes CC

et $\Theta\varepsilon$. Dans le cas contraire, il s'agit de donner un statut informationnel à la suite de l'item. Nous considérerons ce segment informationnel comme le fait d'une thématization interne. En effet, dans ce cas, le constituant extrait est malgré tout informatif (contrairement au cas de l'extraction externe) et les éléments qu'il contient conservent une certaine saillance. Le rôle de cette extraction interne est donc de conserver ce côté informatif tout en faisant de cet élément, *a priori* rhématique par sa construction syntaxique, un thème, en repoussant vers l'arrière plan le poids de cette unité au sein de l'item. Selon Rossi, l'intonation du thème interne peut soit, dans le cas général, copier, en l'abaissant, les caractéristiques de l'intonème continuatif précédent, soit provoquer une parenthèse basse dans le cas où ce qui précède est un topique. Nous ne sommes ici dans aucun de ces deux cas. Nous nous rapprochons plutôt du second dans le sens que le segment précédent de notre configuration a un statut particulier, mais celui de rhème au lieu de celui de topique. La différence caractéristique de ces deux types de structure réside en particulier dans la terminaison conclusive pour l'une et continuative pour l'autre. Aussi, nous inverserons le sens de la parenthèse proposée par Rossi en donnant à la partie de l'item qui suit le noyau rhématique une intonation de parenthétique haute au lieu de basse.

VI.1.4.5 Le langage de représentation

Notre représentation des structures énumératives doit permettre une transformation vers un nouveau langage de marquage de type XML, interprétable par les logiciels de synthèses selon les spécifications acoustico-prosodiques des intonèmes et des stratégies retenues dans la sous-section précédente.

La spécification par le W3C du langage de balisage de synthèse vocale SSML offre une norme de balisage fondée sur XML afin de fournir aux auteurs de contenus synthétisables le contrôle de certains aspects de la sortie vocale (tels que la prononciation, le volume, la tonalité, le débit, etc.) entre différentes plateformes capables de synthèse vocale.

Nous avons donc procédé à l'examen systématique des différentes balises édictées par le consortium et trois cas de figure ont pu être observés pour représenter les structures énumératives selon nos besoins :

- utilisation de SSML sans modification des spécifications prescrites ;
- utilisation de SSML avec modification des spécifications prescrites ;
- spécification de nouvelles balises.

Quatre balises sont présentées ci-après. Une première respectant la norme SSML (BREAK) mais pour laquelle une recommandation est posée relativement à son interprétation, une deuxième dont tous les éléments sont conformes aux spécifications si ce n'est que ces dernières ont conduit à faire des choix au niveau de l'interprétation pour la réalisation acoustico-prosodique d'un des attributs (PROSODY), une troisième pour laquelle aucune modification ou précision supplémentaire n'a été nécessaire (AUDIO), et une quatrième que nous ne pouvons utiliser (EMPHASIS), nous conduisant à la suggestion d'ajouter aux spécifications du langage une nouvelle balise (FOCUS).

L'élément BREAK

L'élément break contrôle les pauses. Les attributs de cet élément sont les suivants :

- *strength* : attribut optionnel ayant l'une des valeurs "none", "x-weak", "weak", "medium" (valeur par défaut), "strong" ou "x-strong". Il sert à indiquer la force de la rupture prosodique dans la sortie vocale. "none" indique qu'aucune pause ne doit être produite. Les autres valeurs indiquent une force spécifique de rupture. Les intitulés "x-weak" et "x-strong" correspondent respectivement à *extra weak* et *extra strong* ;
- *time* : attribut optionnel indiquant la durée d'une pause à insérer dans la sortie, exprimée en seconde ou milliseconde.

Les valeurs de l'attribut *strength* ne sont pas censées être calculées relativement au débit de la parole de synthèse ; elles réfèrent plutôt à une notation pratique qui est en fait traduite ensuite par le système en une valeur prédéfinie en milliseconde. Cependant, les pauses interphrastiques sont calculées par le synthétiseur en fonction du débit de parole. Ce dernier point implique que nous imposons que le traitement des valeurs de pauses des structures énumératives soit également spécifié en fonction du débit de parole de la synthèse.

L'élément PROSODY

L'élément prosody permet la gestion de la tonalité, du débit et du volume de la sortie vocale. Les attributs sont les suivants :

- *pitch* : définit la tonalité de référence pour le texte contenu. Les valeurs légales sont un nombre suivi immédiatement de la chaîne Hz, un changement relatif ou "x-low", "low", "medium", "high", "x-high" ou "default" ;

- *contour* : fixe le contour de tonalité réel du texte contenu ;
- *range* : définit l'amplitude tonale (variabilité) du texte contenu. Les valeurs légales sont un nombre suivi immédiatement par la chaîne Hz, un changement relatif, ou "x-low", "low", "medium", "high", "x-high" ou "default" ;
- *rate* : provoque un changement dans le débit de la parole du texte contenu. Les valeurs légales sont un changement relatif, ou "x-slow", "slow", "medium", "fast", "x-fast" ou "default" ;
- *duration* : précise une valeur en secondes ou millisecondes indiquant la durée souhaitée pour lire le contenu de l'élément ;
- *volume* : modifie le volume sonore du texte. Les valeurs légales sont un nombre, un changement relatif, ou "silent", "x-soft", "soft", "medium", "loud", "x-loud" ou "default".

Nous utiliserons seulement les attributs *pitch* et *range* pour respectivement le calcul des changements de registre des (n-1) premiers items en fonction du nombre total d'item et du rang de l'item courant et « l'étirement » du dernier item.

L'élément AUDIO

L'élément audio gère l'insertion de fichiers sons enregistrés et l'insertion d'autres formats de son en conjonction avec une sortie de synthèse vocale. L'élément audio peut être vide. S'il n'est pas vide, alors il devrait contenir un texte à dire au cas où le document sonore ne serait pas disponible. L'attribut *src*, qui représente l'adresse URI d'un document avec un type MIME approprié, est obligatoire.

L'élément EMPHASIS

L'élément emphasis demande à ce que le texte contenu soit dit avec emphase. Le synthétiseur détermine la manière de restituer cette emphase, dans la mesure où la nature de l'emphase diffère entre les langues, les dialectes ou même les voix. Son seul attribut est décrit ci-après.

level : attribut optionnel indiquant la force de l'emphase à appliquer. Les valeurs définies sont : "strong", "moderate", "none" et "reduced". L'interprétation des niveaux "strong" et "moderate" dépend de la langue utilisée (les langues indiquent une emphase en combinant, éventuellement, un changement de tonalité, des changements de rythme,

une augmentation du volume et d'autres différences acoustiques). Le niveau "reduced" représente effectivement l'opposé d'une emphase. On utilise le niveau "none" pour empêcher que le synthétiseur mette en exergue les mots que, normalement, il aurait accentués.

Nous voyons que les spécifications de cette balise, qui se rapproche le plus de la notion de focus, est extrêmement vague (« la balise EMPHASIS met de l'emphase !!! ») et son format ne correspond pas à ce que nous attendons pour la réalisation d'un focus étendu ou rhématique dans notre forme prosodique. Nous proposons donc de rajouter la possibilité de marquer un focus et de préciser son type en attribut par un nouvel élément : **FOCUS doté d'un attribut *type* pouvant prendre les valeurs "étendu" ou "rhématique"**.

Le Tableau 5 met en regard le segment de texte à substituer, la stratégie prosodique de substitution appliquée et la représentation SSML traduisant le résultat de cette substitution.

Segment supprimé	Stratégie prosodique	Marque SSML résultante
« <i>sont énuméré(e)s</i> »	Focus étendu sur le numéral et le classifieur	<focus <i>type</i> = "etendu">
« <i>ci-après</i> »	Intonème du CTr en fin de focus étendu	</focus>
<i>Adjectifs ordinaux</i>	Prise d'inspiration Changement de registre à chaque item	<audio src=URI/> <prosody <i>pitch</i> = f(nbre item, rang item)> </prosody>
<i>Reprise classifieur</i>	Focus sur le noyau rhématique de l'item	<focus <i>type</i> = "rhématique"> </focus>
« <i>et dernier</i> »	« Étirement » de la F0 du dernier item	<prosody <i>range</i> = +80Hz>
<i>Structure interne</i>	Parenthétique haute Stratégie pausale	</focus> fin d'item </prosody> <break <i>strength</i> = "strong" "medium" />

Tableau 5 : de la FDI à la FPC

Si nous reprenons l'exemple précédent, la forme finale de notre version prosodique serait :

```
<SE> <FOCUS type="étendu"> trois interventions du centre social </FOCUS>, parmi d'autres,
sont <BREAK strength="strong"> <AUDIO src="uri"/> <PROSODY pitch="calculé pour
item 1/3"> <FOCUS type="rhématique"> la gestion et le suivi de six appartements d'urgence
</FOCUS> que propose la commune <BREAK strength="medium"> </PROSODY>
<PROSODY pitch="calculé pour item 2/3"> <FOCUS type="rhématique"> l'organisation de
réunions d'information collectives </FOCUS> <BREAK strength="medium"> </PROSODY>
<PROSODY range=Xst> <FOCUS type="rhématique"> les étape du montage des projets de
formation </FOCUS> </PROSODY> </SE>
```

Pour que cette version puisse être dérivée de notre représentation source, il convient de compléter cette dernière afin de conserver les noyaux rhématiques récupérés lors de la phase d'extraction.

```
<SE type="parallèle" exhaus="non" ordon="neutre">
<AMORCE genre_classifieur= fem nom_classifieur= intervention >
    interventions du Centre Social
</AMORCE>
<ENUMERATION>
<ITEM cs= nom_sing > <NOYAU> la gestion et le suivi de 6 appartements d'urgence
</NOYAU> que propose la commune</ITEM>
<ITEM cs= nom_sing > <NOYAU> l'organisation de réunions d'information collective
</NOYAU> </ITEM>
<ITEM cs= nom_plur > <NOYAU> les étapes du montage des projets de formation
</NOYAU> </ITEM>
</ENUMERATION></SE>
```

Ainsi, à partir d'une telle représentation canonique des structures énumératives, il sera possible d'obtenir sa FDI et sa FPC. Le texte de la première forme pourra être fourni directement au logiciel de synthèse pour son oralisation, tandis que la seconde forme nécessite l'implémentation des balises de la représentation cible afin que le logiciel de synthèse puisse les interpréter. Les DTD et les feuilles de transformation qui permettent de réaliser nos deux stratégies sont fournis en annexe de ce document.

VI.1.4.6 Intégration de nos spécifications prosodiques dans un système de TTS

Nous avons, en collaboration avec un partenaire industriel (Elan Speech), participé à l'implémentation des règles d'interprétation de nos marques en terme de génération acoustico-prosodique. Les spécifications ci-dessous ont été élaborées selon un cycle de vie de type itératif.

Principes de bases

5 niveaux intonatifs sont proposés par Rossi (1999), fondés sur les bases d'expériences psycho-acoustiques. Appliqués au registre de la voix de synthèse utilisée, nous définirons :

- -100 Hz = niveau **grave** ;
- 100 Hz -170 Hz = niveau **médium** ;
- 170 Hz -220 Hz = niveau **infra-aigu** ;
- 220 Hz - 250 Hz = **niveau aigu** ;
- 250+ Hz = **niveau suraigu**.

Les paramètres acoustico-prosodiques caractérisant les morphèmes intonatifs sont fournies en Unités de Perception (UP). Cette unité se définit comme suit :

1 UP = une fois la valeur du seuil de perceptibilité (ou seuil différentiel, ou Sd) du paramètre considéré. C'est à dire la plus petite différence perceptible entre deux signaux (Les Sd pour la parole sont plus grands que ceux qui ont été obtenus pour les sons purs). La valeur de Sd est fonction du paramètre prosodique considéré :

- **Sd de fréquence fondamentale (F)** entre deux voyelles = 5% (*i.e* : en gros 1 demi-ton) ;
- **Sd d'intensité (I)** = environ 3 dB (*i.e* : augmentation de l'amplitude de 40% et doublement de la puissance) ;
- **Sd de durée (D)** = 20% en moyenne pour les durées courantes situées entre 100 et 200 ms.

Ainsi, pour transformer la valeur initiale d'un paramètre prosodique X_i en une nouvelle valeur X_f par l'augmentation (respectivement par la diminution) de n UP, le calcul suivant est effectué :

- $D_f = D_i * (\text{respectivement } /) 1,2^n$;
- $F_f = F_i * (\text{respectivement } /) 1,05^n$;
- $I_f = I_i + (\text{respectivement } -) 3*n$ (Decibels) →
 - Amplitude_i * (respectivement /) $1,4^n$
 - Puissance_i * (respectivement /) 2^n

L'interprétation de nos balises de représentation des structures énumératives nécessite la réalisation acoustico-prosodique des morphèmes intonatifs suivants :

Le morphème de continuatif de topique référentiel **CTr** dont les propriétés acoustiques sont les suivantes :

- F0 : induit une mélodie toujours dominante dans l'énoncé ; présence d'un glissando sur la syllabe accentuée d'au moins 3 UP, dont le sommet est situé entre les niveaux Infra-aigu et Aigu de la tessiture du locuteur ;
- Durée : allongement de la syllabe accentuée de 5 UP en moyenne qui crée une pause subjective ;
- Intensité : pic d'intensité sur tout ou partie de la voyelle qui induit un maximum de sonie.

Le morphème de focalisation **AF** dans un contexte de **focus étendu** présente les caractéristiques prosodiques suivantes :

- Propriétés acoustiques de l'intonème du morphème de focalisation
 - o La F0 doit se situer entre l'infra-aigu et le suraigu selon la force de focalisation (le suraigu correspond au domaine de l'expressivité),
 - o La durée entraîne
 - Une voyelle brève si AF à l'initiale, mi-longue ou longue lorsqu'il est amalgamé à l'accent lexical sous-jacent **as**,
 - Un allongement de la ou les consonnes initiales,
 - o L'intensité doit être supérieure ou égale à 5 dB ;
- Propriétés acoustiques de l'arc prosodique du focus étendu
 - o effacement de toute frontière vers le bas du médium (entre une tierce et une quinte) sur tout le domaine de AF,
 - o effacement des pauses démarcatives.

Le morphème conclusif majeur **CC** qui se réalisera, dans le contexte de la rhématisation par identification, par un intonème qui présente les caractéristiques suivantes :

- mélodie basse dans le niveau Grave ;

- chute d'intensité supérieure à 8 dB ;
- allongement de 4 UP.

L'opérateur d'abaissement $\Theta\epsilon$ sera interprété, pour produire une parenthèse haute, par une mélodie montante d'un ton au-dessus de celle du constituant précédent (CC), soit jusqu'au niveau medium.

Contraintes d'implémentations

L'implémentation a été réalisée en collaboration avec un partenaire industriel, *ElanSpeech*¹¹, pour le moteur de synthèse *Proverbe Speech Engine*. Le logiciel de recopie de la prosodie *Prosel*¹² et un cycle de vie incrémental ont permis l'implémentation des spécifications dans une version de leur moteur de synthèse, telle qu'elle est décrite plus loin. Quelques contraintes ont aussi du être prises en compte.

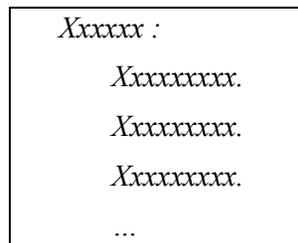


Figure 13 : Image de Page de la structure visuelle du texte source synthétisé

Tout d'abord, la structure prosodique de départ, sur laquelle s'est appuyée l'implémentation, est la synthèse du contenu articulable de la version prosodique en contexte, mais ponctué de la manière exprimée par l'Image De Page de la Figure 13.

Le fichier prosodique généré par la synthèse de parole à partir de ce texte est ensuite modifié par l'interprétation de notre marquage de type SSML puis réalisé acoustico-prosodiquement. Les fichiers prosodiques ainsi construits ne permettent pas de gérer le paramètre d'intensité. Ce dernier n'a aussi pas été implémenté.

¹¹ <http://www.elan.fr/index.html>

¹² http://www.elan.fr/products/fp_prosel.html

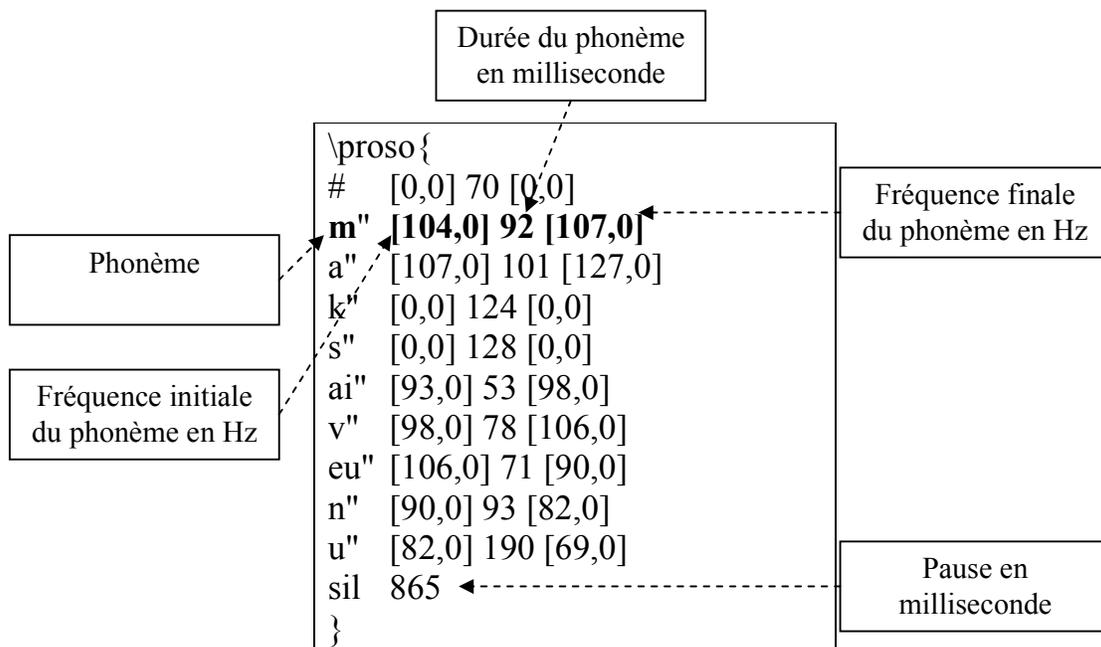


Figure 14 : Exemple de fichier prosodique généré à partir de « Max est venu. »

Un fichier prosodique se présente sous la forme d'une suite de phonèmes et d'indications de pauses silencieuses. Pour chaque phonème sont indiquées la fréquence fondamentale de départ en Hertz, la fréquence fondamentale d'arrivée en Hertz et la durée en millisecondes (Cf. Figure 14 ci-après pour un exemple de fichier prosodique et, dans l'encadré suivant, la mise en œuvre prosodique du « sont » de fin d'amorce de la FPC).

Interprétation finale de nos spécifications

Les spécifications et précisions précédentes conduisent à l'interprétation acoustico-prosodique des structures énumératives telle qu'elle est décrite ci-après.

Le « sont » de fin d'amorce incomplète est modifié de manière à respecter les caractéristiques suivantes :

<p>s" [0,0] 110 [0,0]</p> <p>on" [110] 200 [120,0] 100 [150,0]</p>
--

où le second phonème est décrit en utilisant la possibilité offerte de générer plusieurs phases intonatives à l'intérieur même d'un phonème.

<break strength = "strong" | "medium"/>

Pour le débit de la voix de synthèse utilisée :

<break strength = "strong"/> entraîne une pause de 600 ms

<break strength = "medium"/> entraîne une pause de 400 ms

<focus type="étendu"> numéral + classifieur </focus> :

Sur le numéral (mots toujours d'une syllabe) :

- durées consonantiques initiales * par 1.2 ;
- finale de la vocalique amenée à 200 Hz ;
- sommes des durées entre voyelles et semi-voyelles ≤ 100 m s (le cas échéant les voyelles et semi-voyelles seront réduites dans des proportions identiques) ;
- intensité * racine de 2.

Sur la dernière syllabe (celle qui contient la dernière vocalique différente du « shwa » ou « e muet ») :

- vocalique : $[F_{i1}] d_1 [F_{f1}]$ devient $[F_{i1}] d_2 [F_{f2}]$ avec
 - o $d_2 = d_1 * 1.2^5$,
 - o $F_{f2} = F_{i2} * 1.05^5$,
- conservation du contour pour les phonèmes suivants : $[F_{i1}] d [F_{f1}] \rightarrow [F_{i2}] d [F_{f2}]$ avec :
 - o $F_{i2} = F_{f0}$ (fréquence fondamentale finale du phonème précédent),
 - o $F_{f2} = (F_{f1} / F_{i1}) * F_{i2}$
- Intensité * racine de 2.

Sur les autres phonèmes :

- Si $F > 110$ alors F ramené à 110 ;
- Si $F < 100$ alors F amené à 100.

< focus type="rhematique"> noyau rhématique </focus>:

- élimination des pauses ;
- dernière syllabe :
 - o fréquences des sonores : 90 Hz,
 - o durée vocalique : 180 ms,
 - o intensité ramenée à 70% (* 0.7) ;
- durées des autres phonèmes vocaliques : tout ce qui est > 120 ms est ramené à 100 ms.

<PROSODY range="+80Hz">

L'interprétation entraîne un « étirement » de la pente de fréquence fondamentale pour que la fréquence de l'item soit augmentée de 80 Hertz (soit environ un départ à 180 Hertz avec la voix de synthèse utilisée) tout en conservant la valeur des durées et de la fréquence finale.

<PROSODY pitch = "+X Hz">

Le changement de registre sur les n-1 premiers items d'une structure énumérative composée de n items est interprété de manière à ce que les fréquences F_{init} de l'item soient recalculées en F_{fin} selon la formule : $F_{fin} = F_{init} * 1.05^X$. X est calculé par la feuille de transformation selon la formule $X = 4 * (n-i) / (n-1)$ qui permet le calcul du changement de registre perceptible le plus grand possible sans dépasser les capacités de la voix de synthèse utilisée. Cette dernière formule n'est pertinente que jusqu'à 5 items mais à l'avantage de proposer une solution non choquante pour plus d'item (les différences de niveau intonatif du registre entre deux items ne seront simplement pas perceptibles).

De plus,

en ce qui concerne la dernière syllabe item

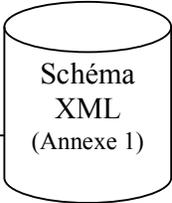
si il y a du texte entre <FOCUS type="rhematique"> et </PROSODY> et qu'il ne s'agit pas du dernier item nous procédons à une inversion du contour de la syllabe (par symétrie horizontale). Si la dernière vocalique est inférieure à 150 Hertz elle est ramenée à cette valeur.

Les spécifications pseudo-algorithmiques que nous venons de décrire pour l'interprétation des balises SSML finalisent deux formes de l'axe interprétatif du MORTELS : la forme discursive interprétative (FDI) et la forme prosodique en contexte (FPC). Nous résumons, dans le schéma suivant (Figure 15), à travers un exemple les différentes étapes et ressources qui ont été nécessaires pour transformer une structure énumérative « écologique » afin de la rendre oralisable par un système de TTS selon ces deux stratégies.

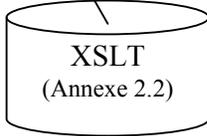
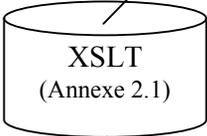
Le Centre Social intervient entre autres dans :

- la gestion et le suivi de 6 appartements d'urgence que propose la commune ;
- l'organisation de réunions d'information collective ;
- les étapes du montage de projets de formation.

Saisie manuelle des paramètres formels des SE dans une interface DELPHI pour simuler le rôle des phases de segmentation et d'extraction.



```
<SE type="parallèle" exhaus="non" ordon="neutre">
<AMORCE genre="classifieur" fem="nom" classifieur="intervention"> interventions du Centre Social</AMORCE>
<ENUMERATION>
<ITEM cs="nom_sing"><NOYAU> la gestion et le suivi de 6 appartements d'urgence </NOYAU> que propose la commune</ITEM>
<ITEM cs="nom_sing"><NOYAU> l'organisation de réunions d'information collective </NOYAU> </ITEM>
<ITEM cs="nom_plur"><NOYAU> les étapes du montage des projets de formation </NOYAU> </ITEM> </ENUMERATION></SE>
```



Trois interventions du centre social, parmi d'autres, sont énumérées ci-après. Une première intervention est la gestion et le suivi de 6 appartements d'urgence que propose la commune. Une deuxième intervention est l'organisation de réunions d'information collective. Une troisième et dernière intervention comprend les étapes du montage de projets de formation.

```
<SE> <FOCUS type="étendu"> trois interventions du centre social </FOCUS>, parmi d'autres, sont <BREAK strength="strong">
<AUDIO src="uri"/> <PROSODY pitch="calculé pour item 1/3"> <FOCUS type="rhématique"> la gestion et le suivi de six appartements d'urgence </FOCUS> que propose la commune <BREAK strength="medium">
</PROSODY> <PROSODY pitch="calculé pour item 2/3"> <FOCUS type="rhématique"> l'organisation de réunions d'information collectives </FOCUS> <BREAK strength="medium">
</PROSODY> <PROSODY range=Xst> <FOCUS type="rhématique"> les étapes du montage des projets de formation </FOCUS> </PROSODY> </SE>
```

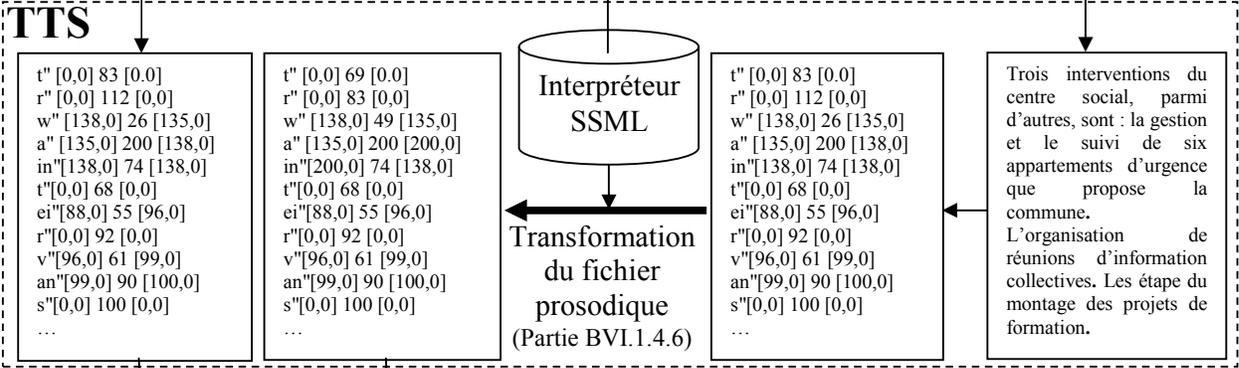


Figure 15 : D'une SE tout-venant à sa FDI et à sa FPC : processus et ressources

VI.2 Oralisation d'autres Objets Textuels

Cette section veut montrer l'intérêt des autres stratégies du modèle pour les autres OT plus particulièrement abordés dans la partie précédente.

Considérons en premier lieu le cas des titres. Cet OT relève de notre point de vue de l'inscription graphologique démarcative. Nous avons proposé ci-avant la métaphore suivante : *parlons maintenant de X*. Cette métaphore ne fonctionne que pour un titre dit thématique. D'autres métaphores moins problématiques peuvent être proposées, du type *abordons maintenant X* (*X* pouvant être remplacé par une UT ayant un contenu informationnel thématique ou non) ; mais cette formulation pourra impliquer des reformulations syntaxiques car la tête de *X* ne peut posséder n'importe quelle catégorie syntaxique.. Cela dit, l'écoute empirique de notre recueil ne semble pas montrer de différence intonative selon le type de titre oralisé alors que l'intention de titrer apparaît néanmoins. Nous en concluons que la prosodie que nous percevons dans ce cas ne provient pas d'effacements dans de telles métaphores, mais d'effacements dans des métaphores issues de la stratégie dénominative S2, par exemple *le titre est X*. Nous prônons donc l'utilisation d'une forme prosodique hors contexte d'énonciation dont les caractéristiques acoustico-prosodiques sont :

- un départ dans les niveaux aigus conjugué à un pic d'intensité sur la syllabe initiale ;
- une déclinaison intonative régulière qui termine dans les niveaux infra-bas du locuteur.

Pour traiter le cas des paragraphes, il convient de rappeler la grande difficulté des linguistes à statuer sur la raison graphique de cet OT. De notre point de vue, cela s'explique par le fait que l'intention architecturante n'est pas directement dépendante du type de contenu : la mise en paragraphe traduit seulement la volonté d'exprimer une fermeture et une ouverture de cadres du discours, les raisons de cette « cassure » sont expliquées ou analysables par l'étude du contenu et non par l'inscription graphologique. A ce titre la mise en paragraphe serait liée à une sorte de stratégie cognitive qui est proposée au lecteur pour lui donner le temps d'encoder les informations. Ainsi, ce caractère purement démarcatif et temporel ajouté à la lourdeur que la forte occurrence de cet OT imposerait à une stratégie dénominative, nous conduit à opter pour une stratégie descriptive dont la version discursive se formulerait par des *point à la ligne* et la version intonative reposerait sur l'utilisation d'une pause allongée.

Enfin, considérons le cas plus complexe des OT relatifs à la mise en saillance telle que nous l'avons définie, c'est-à-dire relevant de l'inscription graphologique distinctive :

- s'il s'agit d'un procédé distinctif thématique, c'est-à-dire dans ce cas d'une intention rédactionnelle visant à expliciter le lieu de la réponse à une question qu'est supposé se poser le lecteur, alors il semble complexe de concevoir un *pattern* générique susceptible de réceptionner les métaphrases impliquées par une telle intention architecturante ;
- s'il s'agit d'un procédé distinctif modalisant il s'agira dans un premier temps de mettre en évidence quel type de dichotomie est révélé par la mise en saillance. Dans le cas fréquent d'une intention de distinguer l'important du moins important, nous pourrions utiliser les variations d'une métaphore construite autour de l'expression *j'insiste sur*. Par exemple *Max est venu* serait transformé en *Max, j'insiste sur Max, est venu*. L'effacement dans ce type de métaphore pourra ensuite s'appuyer sur les nombreux travaux traitant de la focalisation prosodique. Remarquons qu'une forme hybride peut également être envisagée à travers une substitution au profit d'une construction syntactico-prosodique : les clivées ou les pseudo-clivées permettront en effet de transformer notre exemple en *C'est Max qui est venu*, voire *Celui qui est venu, c'est Max*.

De manière générale, il nous semble qu'il sera difficile lors des phases de segmentation et d'extraction de préciser si une configuration utilisant un procédé contrastif relève d'une mise en saillance ou d'une démarcation ; si, dans le premier cas, il s'agit d'une inscription graphologique distinctive de type thématique ou modalisante, et, ce cas échéant, quelle dichotomie est mise en évidence. Ainsi, nous opterons plutôt pour un axe stratégique différent.

Ce dernier choix nous permet de formuler une remarque concernant le MORTELS. La mise en saillance étant généralement provoquée par une seule marque (gras, italique, souligné, ...), les deux axes **S1-C1** et **S2-C2a** du modèle se confondent : d'une part nommer la configuration revient à décrire la marque utilisée et d'autre part la convention sonore injective est également bijective. Ainsi nous proposons une seule forme discursive qui donnerait, en continuant sur notre exemple, *l'auteur met en gras Max dans la phrase Max est venu*.

De là nous pouvons utiliser l'un des deux continuums **C2a** ou **C2b** : ce deuxième axe est intéressant puisqu'il permet de limiter la lourde insertion à *en gras*, à condition que

soit dégagé un moyen prosodique hors contexte d'énonciation, pour assurer que ne survienne aucune ambiguïté quant au segment de texte visé par ce terme métalinguistique : lors de l'oralisation de *en gras Max est venu* il convient de s'assurer d'un regroupement intonatif des termes *Max* et *en gras*. Notons enfin que la stratégie conventionnelle a été également prétestée lors d'une expérimentation psycholinguistique (Truillet, 1999) et qu'un des résultats semble indiquer que si la présence d'une indication sonore (un *bip* dans l'expérience) pour signifier la mise en gras d'un mot est lourdement décriée par les sujets, c'est la stratégie qui mène aux meilleurs résultats en terme de mémorisation !

Cette dernière remarque nous conduit à la partie suivante de ce tapuscrit puisqu'elle met en évidence qu'il est nécessaire de formuler des hypothèses cognitives sur l'impact de nos stratégies sur la mémorisation et la compréhension des informations véhiculées par le texte oralisé. Nous avons élaboré un protocole expérimental afin d'initier un tel objectif. Le *design* de ce protocole et les principes de la méthodologie expérimentale utilisée sont présentés dans la partie suivante, suivis de l'analyse des premiers résultats.

« J'ai lu l'Odyssée. - Vous l'avez lue dans le texte ? - Non, dans le train! - Ça va plus vite. - Oui, d'autant que c'était un rapide. Évidemment quelques passages m'ont échappé. Les passages à niveau, naturellement. »

Pierre Dac

Partie C Évaluation cognitive

La partie précédente avait entre autres pour ambition de répondre à la question : une petite prosodie vaut-elle un long discours ?

Nous essaierons ici de répondre à une nouvelle question : une petite prosodie vaut-elle **mieux** qu'un long discours ?

En d'autres termes, notre objectif est d'évaluer si une véritable équivalence intermodale est envisageable. Pour cela, l'obtention d'une équivalence en terme de contenu informationnel est, de notre point de vue, insuffisante.

Nous nous attacherons à présenter un protocole expérimental visant à évaluer l'équivalence en terme de traitement cognitif (mémorisation/compréhension) Ce protocole, à l'origine prévu pour évaluer le traitement cognitif selon deux populations d'utilisateurs (jeune contre âgée), n'a été partiellement mis en œuvre, et discuté ici, que pour la population jeune.

Chapitre VII. Hypothèses et variables

Le protocole vise à évaluer l'efficacité de deux formes interprétatives d'oralisation (discursive et prosodique) en termes d'impact sur le traitement cognitif (mémorisation et compréhension) (Maurel, 2004b).

Le protocole expérimental tel que nous le décrivons ici vise à tester deux hypothèses :

- H1 : la restitution de la signification des aspects visuels structuraux de l'écrit améliore les performances en termes de compréhension/mémorisation ;
- H2 : la stratégie la plus efficace peut dépendre du type de population envisagé.

Trois variables indépendantes sont envisagées pour mener à bien cette investigation.

VII.1 Variable indépendante

VII.1.1 F : forme d'oralisation du texte

Les deux formes d'oralisation que nous nous proposons d'étudier sont celles décrites dans la partie précédente : la forme discursive interprétative (FDI) et la forme prosodique en contexte (FPC). Deux hypothèses spécifiques relatives à cette variable sont proposées.

VII.1.1.1 Deux hypothèses spécifiques

Hypothèse 1 (choix d'une hypothèse de supériorité) :

Les formes d'oralisation restituant les aspects structuraux produisent de meilleures performances que la forme contrôle qui ne les restitue pas ; comparaison entre la forme contrôle et les 2 autres formes.

Hypothèse 2 (choix d'une hypothèse de différence simple) :

Les performances des sujets diffèrent selon la forme d'oralisation des textes qu'ils ont écoutée ; comparaison de toutes les formes deux à deux.

Nous n'avons pas proposé de prédiction quant à la supériorité d'une forme sur l'autre car des raisonnements différents permettraient de conduire aux deux hypothèses : en effet quel élément de décision prévaudra pour départager les deux formes entre :

- l'avantage de l'explicitation de l'architecture du texte dans la forme discursive par rapport à la version prosodique plus implicite de ce point de vue ;

- l'inconvénient de la longueur du message pour la version discursive par rapport à la version prosodique qui est sensiblement plus courte.

VII.1.1.2 Conception de la forme contrôle

Les hypothèses spécifiques posées impliquent de concevoir une troisième forme d'oralisation dite de contrôle. La forme contrôle (FC) est essentielle dans le protocole puisque c'est elle qui nous permettra de statuer sur l'hypothèse 1 selon laquelle la restitution des aspects visuels structuraux améliore les performances des sujets en termes de mémorisation/compréhension. Cette forme ne doit donc pas restituer les aspects visuels structuraux du texte écrit.

La contrainte à gérer est celle du vase communicant entre validité interne et validité externe : sur un plan écologique et applicatif, il conviendrait d'utiliser la forme d'oralisation obtenue directement par le même moteur de synthèse que pour les autres formes sans la rectifier ; le problème est que nous n'avons ni connaissance, ni contrôle des règles d'oralisation de la synthèse portant sur les aspects visuels structuraux. Nous savons que certaines marques sont traitées mais nous ne savons pas exactement comment. Le risque serait alors d'obtenir une forme contrôle qui soit quelque part entre une absence de traitement de l'inscription graphologique et notre forme prosodique. Nous ne contrôlerions pas dans ce cas réellement le strict effet de nos propres stratégies de reformulation.

Une possibilité est de fournir au moteur de synthèse, le texte de la forme prosodique dénué de toute inscription graphologique. Les tests autour d'une telle proposition montrent que le stimulus résultant présente des « bogues » dans l'oralisation, dus, non pas à l'absence de prise en compte de la structure visuelle, mais à un découpage syntaxique du texte défectueux.

La solution retenue finalement fut de se conformer à l'ordre proposé par le MORTELS : puisque la première forme de restitution des aspects visuels structuraux à l'oral est une FDI exploitant des marques lexico-syntaxiques, et que la seconde (FPC) est le produit d'effacements dans la FDI au profit de marques prosodiques, alors la forme contrôle devrait être obtenue par les mêmes effacements mais sans contreparties prosodiques. Ainsi, la forme contrôle adaptée est la version orale de la FDI privée des segments de signal correspondant aux marques lexico-syntaxiques architecturantes.

Du fait qu'elle soit formée à partir de la forme discursive orale, la forme contrôle hérite de ses caractéristiques prosodiques. Ces caractéristiques sont elles-mêmes dépendantes de la structure syntaxique de la forme discursive. La forme discursive exploite des structures de type phrastique (l'amorce constitue une phrase ainsi que

chaque item), alors que la forme contrôle, délestée de certaines marques lexico-syntaxiques, n'a plus la même structure syntaxique (c'est une seule phrase). Cela dit, une pré-évaluation perceptuelle des stimuli résultant montrent que l'intonation syntaxique de la forme contrôle ainsi constituée n'est pas choquante.

VII.1.2 A : âge des sujets

L'accès aux documents peut être compliqué par des facteurs nécessitant une adaptation des stratégies d'oralisation en fonction des caractéristiques spécifiques à une population d'utilisateurs ; c'est, nous semble-t-il, *a priori* le cas pour les personnes âgées en raison de potentielles difficultés de perception de la prosodie et/ou de déficiences mnésiques. Ainsi nous proposons de tester à terme les deux hypothèses spécifiques suivantes :

Hypothèse 1 (hypothèse de supériorité): les performances des sujets jeunes sont meilleures que les performances des sujets âgés ;

Hypothèse 2 (hypothèse d'interaction simple) : l'effet de la forme d'oralisation sur les performances des sujets dépend de leur âge.

VII.1.3 T : tâche demandée

La mémorisation/compréhension des informations sera évaluée à travers des tâches de rappel libre et indicé :

- La tâche de rappel libre consiste pour le sujet à rappeler le maximum d'informations du texte entendu ;
- La tâche de rappel indicé consistera à remplir un texte à trous reproduisant le texte entendu.

VII.1.4 Statuts des variables indépendantes

La décision arrêtée pour VII est celle d'un statut intergroupe. La population des sujets est divisée en sous-groupes ; chaque sous-groupe est soumis à une des 3 formes d'oralisation.

Du fait de sa nature, la variable VI2 aura un statut intersujet. Ses modalités sont croisées avec VII.

La variable VI3 a un statut intrasujet : chaque sujet effectuera pour chaque stimulus le rappel libre et indicé des informations.

Ainsi nous obtenons le plan décrit dans le tableau ci-dessous.

Personnes âgées			Personnes jeunes		
FD	FPC	FC	FD	FPC	FC
rappel libre et indicé			rappel libre et indicé		

On obtient donc le plan expérimental représenté par la formule de plan suivante :

$$S_{10} < A_2 * F_3 > * T_2$$

où S désigne la variable sujet, T la variable “type de tâche”, F la variable “forme d’oralisation” et A la variable “âge”. Les chiffres indicés désignent le nombre de modalités de la variable.

VII.2 Variables de contrôle

La principale difficulté liée à la constitution de six groupes indépendants repose sur les différences susceptibles de caractériser les sujets. En effet il est difficile de savoir si l’observation statistique des données recueillies sont réellement imputables aux facteurs étudiés ou aux spécificités des sujets. Pour notre évaluation, il est nécessaire de contrôler les facteurs intrinsèques au sujet susceptibles d’affecter la qualité du rappel.

- familiarité par rapport à la synthèse vocale : élimination des personnes familières à la synthèse vocale (étude, travail, loisir) afin d’éviter une influence de cette variable sur les résultats ;
- niveau d’étude : afin de faciliter la constitution de groupes équivalents, nous avons opté pour le choix d’un niveau bac +2/3. Il s’agira également de s’assurer, autant que faire se peut, de l’équivalence des groupes « jeunes » et des groupes « âgés » sur le niveau d’étude ; ceci, bien qu’il soit difficile d’évaluer le degré de correspondance d’un niveau d’étude donné actuel avec son équivalent il y a 40 ou 50 ans ;
- français langue maternelle : il est nécessaire de s’assurer d’un niveau égal de difficulté de la tâche pour tous les individus. Les personnes étrangères seront automatiquement éliminées, sauf celles vivant en France depuis leur enfance. Il s’agit ainsi d’éviter que des mauvais résultats dans les tâches de rappel ne soient dus à une mauvaise compréhension du vocabulaire ou de la syntaxe employés dans les stimuli test ;
- problème d’audition : l’ensemble de l’expérimentation sollicitant le canal auditif, il est nécessaire de contrôler ce facteur. On s’assurera ainsi lors du

recrutement que les personnes ne souffrent pas de pathologies avérées liées à l'audition. Par ailleurs, on donnera au sujet la possibilité de régler le volume de son casque en début d'expérience ;

- pathologie liée à la mémoire : il importe de s'assurer lors du recrutement que les sujets ne souffrent pas de pathologies avérées liées à la mémoire ;
- sexe : seules des expériences portant sur la mémoire émotionnelle ont montré des différences de performance selon le sexe des sujets, il n'est donc pas nécessaire pour notre expérience (qui sollicite seulement la mémoire à court terme) d'appareiller les sujets selon cette variable ;
- âge : cette variable est très importante pour notre expérience puisqu'elle permettra de statuer sur la validité de nos hypothèses. L'impact cognitif des stratégies d'oralisation sera testé auprès de deux populations jeune vs. âgée. Il s'agira de recruter des étudiants de niveau bac +2/3 dont la tranche d'âge sera comprise entre 19/30 ans. Concernant les personnes âgées, la population recrutée aura plus de 70 ans, puisque les changements cognitifs liés au vieillissement s'opèrent de manière plus nette après ce seuil.

Les informations concernant ces variables sont recueillies lors du recrutement pour exclure les sujets ne rentrant pas dans nos critères de sélection.

VII.3 Variables dépendantes

Nous ne détaillerons ici que les variables dépendantes liées à la procédure de rappel libre. En effet, la seconde tâche ne sera analysée que dans le cas d'une absence de résultat pour la première.

Étant donné que nous travaillons sur l'inscription graphologique des textes, et que nos stratégies interprétatives ont comme objectif de la restituer oralement, il nous a semblé intéressant de pouvoir observer si les sujets traitent, dans nos stimuli oraux, les marques prosodiques ou de mise en forme matérielle de nature lexico-syntaxique, au même titre que le contenu informationnel du message original. Ainsi, le choix d'un rappel écrit semble pertinent pour évaluer l'influence de l'oralisation et du traitement cognitif subis par le texte source sur le texte produit par le sujet : les formes d'oralisation discursive et prosodique favorisent-elles plus la restitution de ces marques lors du rappel écrit que la forme contrôle ? Et le cas échéant, quelle forme a été choisie pour leur rappel (inscription graphologique ou MFM de nature lexico-syntaxique) ?

De plus, les sujets effectuent leur rappel dans un cadre délimité sur une feuille A4. En effet, un enregistrement audio/vidéo de la scène de rappel est récupéré au moyen

d'une *Webcam* dont le champ est focalisé sur ce cadre. Les enregistrements seront susceptibles de fournir des informations sur la dynamique même du processus de restitution écrite. Cette remarque implique une précision supplémentaire dans la consigne : bien que le sujet ne puisse pas prendre de notes pendant le rappel, il a la possibilité de raturer et de réfléchir à voix haute. Il indiquera le moment où il considère son rappel terminé pour que lui soit proposé l'exercice suivant.

Les critères sur lesquels on peut coder les informations du rappel libre sont les suivant :

- la quantité d'informations rappelées (quelle que soit leur justesse) par rapport à la quantité d'informations présentées ;
- la quantité d'informations justes rappelées par rapport à la quantité d'informations présentées ;
- la quantité d'informations justes rappelées par rapport à la quantité d'informations rappelées ;
- le degré de littéralité du rappel des informations justes : il peut être codé en attribuant des pondérations. Se pose alors le problème du choix du système de pondération : il s'agit de décider de la valeur pondérée à attribuer à un rappel en fonction de sa proximité avec l'information présentée. Or, la notion de proximité pouvant être sujette à débat, le système de pondération comportera inévitablement une dimension arbitraire. Une solution possible à apporter aux limites du système de pondération est d'étudier dans chaque rappel la proportion de rappel littéral, non littéral, etc. Il s'agirait alors de comparer les pourcentages de chaque type de rappel selon la forme d'oralisation ;
- le respect de l'ordre de présentation dans le rappel : on peut mesurer grâce à l'enregistrement vidéo du rappel libre l'ordre de restitution des informations retenues lors du rappel et l'ordre des informations à la fin du rappel en ne tenant compte que de l'aménagement spatial du texte rappelé. La question est de savoir si l'on code le respect de l'ordre de façon dichotomique ou non. Si oui, il faut déterminer à partir de quand on considère qu'un protocole suit l'ordre de présentation ou non ;
- la tendance ou non à confondre des structures du texte et à les fusionner lors du rappel ; c'est à dire que le sujet associe deux informations présentes dans le texte mais pas à la bonne place.

VII.3.1 Quantité d'informations justes rappelées et du degré de littéralité du rappel

Pour mesurer la quantité d'informations justes rappelées, il faut se fixer un référent (la quantité d'informations justes rappelées pour un rappel parfait), définir la notion de justesse de l'information rappelée et préciser la notion d'unité d'informations.

VII.3.1.1 Choix du référent

Classiquement, deux types de référents peuvent être choisis:

- on compare le rappel au texte présenté ;
- le référent est l'ensemble des informations rappelées.

Étudier des indicateurs construits sur les deux référents peut être intéressant. En effet, en admettant qu'il y ait une variabilité interindividuelle très importante au niveau de la quantité d'informations mémorisées et que cette variabilité brouille d'éventuels effets des variables indépendantes, des indicateurs construits sur le référent de l'ensemble des informations rappelées permettraient de s'affranchir un peu de cette variabilité.

Il faut noter que pour la forme discursive, le texte référent contient davantage de mots du fait de l'explicitation sous forme de métaphrases. On peut alors envisager plusieurs solutions. Soit, on exprime les scores en pourcentage, soit, on ne tient pas compte des informations rappelées, correspondant aux segments supplémentaires.

VII.3.1.2 Niveaux d'analyse des données

Le modèle d'oralisation s'appuyant sur le Modèle d'Architecture Textuelle, le rappel de la SE sera analysée par objets textuels (amorces, items 1, 2, et 3).

On décomposera ensuite chaque objet textuel en « mots pleins » (MP). On élimine ainsi tous les mots outils (pronoms, conjonctions, articles etc.).

Un premier indicateur « grossier » pourrait être l'expression en pourcentage du nombre de mots pleins justes rappelés (voir ci-dessus les critères de justesse) par rapport au nombre de mots pleins présentés (et rappelés). Cet indicateur ne prend donc pas en compte « en quoi » le mot plein rappelé est juste ou faux.

Un second indicateur est de distinguer trois niveaux dans les données du rappel en fonction des mots pleins (MP) présentées : MP absents, MP faux et MP justes.

Un troisième indicateur consistera à distinguer dans la catégorie des « faux » : les confusions (le sujet rappelle un mot présent dans le texte mais pas à la bonne place dans la SE) et les reconstructions.

Un quatrième indicateur distinguera parmi les MP justes les mots pleins littéraux et les MP non littéraux. En effet, le critère général pour juger que le rappel est juste se base sur la compréhension des informations présentées aux sujets. Ainsi, on considérera comme justes les reformulations plausibles : par exemple, si le sujet rappelle que « le dossier est délivré sur 3 sites » pour « les trois sites possibles pour retirer un dossier sont », on considère que « délivré » est un rappel non littéral par paraphrase du mot plein « retirer ».

Les catégories « non littérales » que l'on peut analyser comme justes sont les synonymes, les hyperonymes, les hyponymes, les co-hyponymes et les paraphrases.

Pour le dernier indicateurs (littéral vs. non-littéral), l'analyse consistera à étudier la proportion de chaque catégorie dans le rappel global. Ces indicateurs nous permettraient de juger si une forme d'oralisation favorise un rappel littéral par rapport à une autre forme d'oralisation. Bien que nous n'ayons pas d'hypothèse précise à ce propos, les résultats peuvent être intéressants, surtout si l'analyse quantitative ne révèle pas de différence significative selon les formes d'oralisation.

Les différents niveaux d'analyse sont récapitulés dans le schéma suivant (Figure 16).

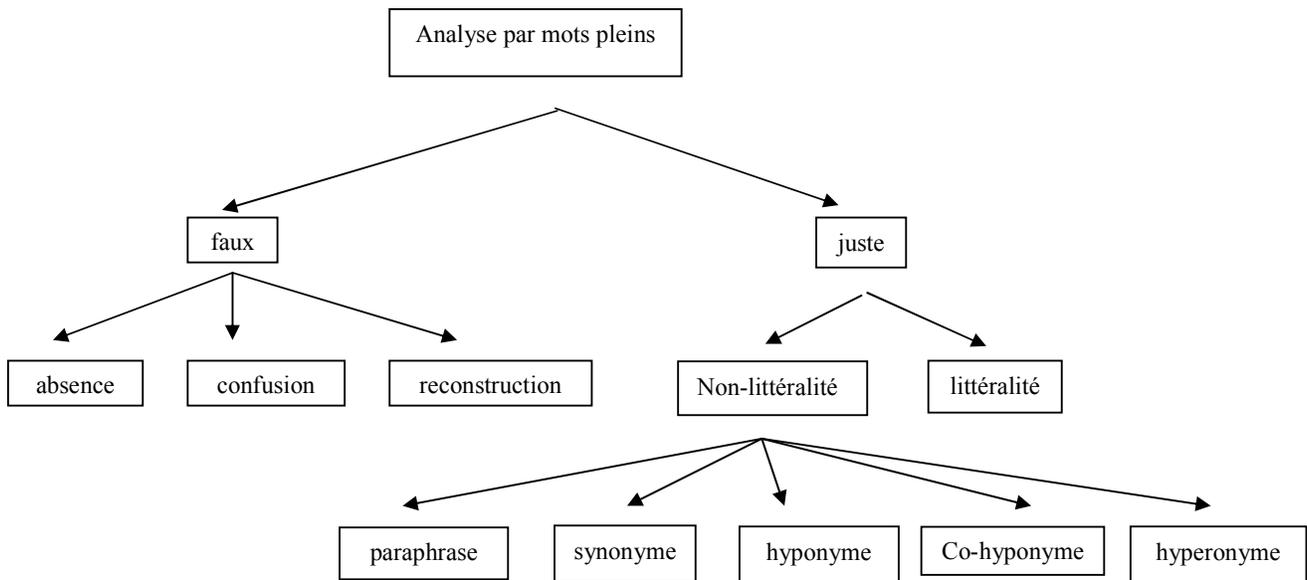


Figure 16 : Schéma récapitulatif des différents niveaux d'analyse

Il conviendra, une fois l'analyse par mots pleins réalisée, de faire une analyse par objets textuels selon les mêmes critères : nombre d'items rappelés, degré de littéralité de ce rappel.

VII.3.2 Restitution des informations architecturales

La modalité du rappel choisie (l'écrit) nous permet ainsi de comparer la structure du texte avant oralisation et sa restitution écrite après oralisation et traitement cognitif.

VII.3.3 Modalité de restitution du nombre d'items :

Le nombre d'items présent dans l'énumération est véhiculé par le numéral de l'amorce pour les trois formes d'oralisation, mais il est également présent dans les items sous forme discursive et prosodique pour les stratégies interprétatives. On peut ainsi analyser s'il existe des différences selon les formes d'oralisation. Par ailleurs, la restitution de cette information peut se faire à plusieurs niveaux : sous forme lexico-syntaxique mais cette information peut être suggérée par la mise en forme matérielle. Ainsi, un sujet peut très bien retranscrire trois tirets mais n'associer un contenu qu'à deux de ces tirets. Il conviendrait donc de coder :

- le nombre suggéré d'items ;
- le mode de restitution du nombre d'items : le sujet peut indiquer qu'il y avait 3 items en restituant le numéral de l'amorce, en restituant les marques lexico-syntaxique de troisième item de la forme discursive, en exploitant une configuration de MFM (tirets, indentations, ponctuation) suggérant la présence de trois items et en combinant ces trois types d'indices. A ce niveau, il peut être intéressant d'étudier quel mode de restitution de cette information les sujets utilisent et si ce mode est influencé par la forme d'oralisation ou non.

Il est important également de prendre en compte la spécificité de la forme discursive. En effet, que deviennent les segments lexico-syntaxiques rendant compte de la mise en forme matérielle de l'énumération ?

Il sera intéressant de voir si ces segments sont rappelés sous forme lexico-syntaxique ou typo-dispositionnelle.

VII.3.4 Mesure du respect de l'ordre

Comme mentionné plus haut, deux dimensions d'ordre peuvent être comparées à l'ordre des informations présentées :

- l'ordre de rappel lors du processus de rappel lui-même ;
- l'ordre suggéré par le rappel, celui-ci terminé.

VII.3.4.1 Codage de l'ordre lors du processus de rappel

Une fois le texte décomposé en objets textuels (enrobage de début, amorce, item 1, 2, 3, enrobage de fin) on peut coder pour chaque sous-structure, son ordre d'apparition lors du rappel grâce aux enregistrements vidéo. Si l'on souhaite dichotomiser l'indicateur en « a respecté l'ordre », « n'a pas respecté l'ordre », il faudra se fixer un critère distinguant clairement les deux catégories. La difficulté est de tenir compte de la quantité de rappel.

Il conviendrait peut être également d'étudier le nombre de retour sur un OT produit auparavant lors du rappel (le sujet rappelle une partie de l'amorce, le premier et le deuxième item puis revient sur l'amorce pour la compléter).

VII.3.4.2 Codage de l'ordre du rappel final

On peut envisager de numéroter les OT en fonction de leur ordre de présentation et de leur ordre de rappel (configuration finale et « processus » de rappel).

Pour chaque objet textuel présenté, on codera si la place est mémorisée et si le contenu est mémorisé, la place pouvant être mémorisée indépendamment du contenu et vice-versa (le sujet peut indiquer la présence d'un item en exploitant par exemple le tiret sans pour autant rappeler son contenu).

Chapitre VIII. *Design final*

VIII.1 Constitution des stimuli

Le matériel expérimental sera constitué de :

- un texte d'entraînement/distracteur : le sujet ne sera pas averti qu'il s'agit d'une phase d'entraînement ; le texte aura comme double emploi
 - o de familiariser le sujet avec la synthèse de parole à partir de texte, les deux tâches de rappel et certains éléments lexicaux des stimuli tests ;
 - o d'éviter que le sujet n'adapte sa stratégie de mémorisation en fonction de notre objet d'étude ;
- deux textes tests : pour augmenter la robustesse des données, nous optons pour un plan à mesures répétées, c'est-à-dire que pour chaque sujet, on recueille deux séries de mesures correspondant à chacun des deux textes tests ; chaque texte test contiendra une énumération à 3 items et un enrobage textuel neutre (pas d'objet textuel autre que du corps de texte) avant et après la structure énumérative. Cette précaution nous permet de contrôler d'éventuels effets de récence et de primauté dans la tâche de rappel libre et d'éviter une focalisation de l'attention du sujet sur l'objet étudié ;
- un texte distracteur court et facile à mémoriser, ayant pour double vocation de ne pas démotiver les sujets (surtout du groupe contrôle) et de ne pas focaliser le traitement de l'ensemble des sujets sur les structures énumératives ;
- 6 textes à trous pour la tâche de rappel indicé : chaque texte test aura une version correspondant à l'oralisation de la forme discursive et une autre correspondant à l'oralisation des formes contrôles et prosodiques ; les deux distracteurs n'impliqueront qu'un texte à trou chacun.

Le thème des textes sera celui de l'Allocation Personnalisée Autonomie car il est en relation avec le projet de mise en place d'une borne interactive à destination des personnes âgées.

Le texte oralisé pour produire le texte d'entraînement et sa version pour le rappel indicé sont encadrés ci-après.

L'allocation personnalisée d'autonomie, ou A.P.A, est une nouvelle prestation. Il s'agit d'un nouveau droit pour aider à mieux vivre les personnes de plus de 60 ans.

Texte oralisé pour constituer le stimulus « texte d'entraînement »

L'allocation _____ d' _____, ou A.P.A, est une nouvelle prestation. Il s'agit d'un nouveau _____ pour aider à mieux vivre les personnes de plus de _____.

Texte à trous écrit pour le rappel indicé du « texte d'entraînement »

En ce qui concerne les deux stimuli tests chaque groupe est soumis à une forme d'oralisation différente, les segments entre parenthèse des transcriptions encadrés ci-après ne sont à prendre en compte que pour la forme discursive. L'ordre de présentation des stimuli « test » fera l'objet d'un contre-balancement pour éviter un éventuel effet d'ordre. Avant la présentation du premier stimuli-test, le sujet est prévenu qu'il sera soumis à l'oralisation d'un texte plus long que le texte d'entraînement, afin d'éviter une posture cognitive adaptée à la mémorisation d'un texte court qui risquerait de trop dégrader les performances.

Une fois le rappel écrit effectué, nous présentons aux sujets les textes à trous correspondant aux textes oralisés. Les segments entre parenthèse ne seront, là aussi, présentés que pour la forme discursive.

Un certain nombre de précisions méritent d'apparaître ici pour justifier de la forme finale de ces deux stimuli. Afin de pouvoir cumuler les données extraites de chacun des stimuli, ceux-ci ont été calibrés de manière à ce que la fréquence d'utilisation des mots les composant soit sensiblement la même d'un stimuli à l'autre. (l'hypothèse est que notre fréquence de confrontation avec un mot est corrélée à notre capacité à le retenir). Ne pouvant construire de cette manière deux textes cohérents pour lesquels chaque mot d'un texte a une contrepartie dans l'autre en terme de fréquence, nous avons opté pour une équivalence des moyennes de fréquences d'utilisation des mots composant chaque OT des stimuli. L'obtention des fréquences d'utilisation de chaque mot a pu être réalisée par l'interrogation du corpus journalistique BRULEX¹³ (Content, 1990).

Une dernière difficulté posée par le test de rappel indicé était de positionner les « trous ». Il s'agissait de construire à partir de nos stimuli, un texte dont les segments effacés dans la SE ne soient pas reconstituables autrement que par leur mémorisation préalable ; au contraire, la reconstruction de l'enrobage devait être facilitée pour ne pas

¹³ BRULEX est un outil développé pour la recherche en psycholinguistique. C'est une base de données lexicales informatisée pour le français écrit et parlé qui donne, pour environ 36.000 mots de la langue française, l'orthographe, la prononciation, la classe grammaticale, le genre, le nombre et la fréquence d'usage. Il contient également d'autres informations utiles à la sélection de matériel expérimental (notamment, point d'unicité, comptage des voisins lexicaux, patrons phonologiques, fréquence moyenne des digrammes) – <http://www.lexique.org/moteur/>

focaliser le sujet sur cette partie du stimuli. La méthode pour parvenir à réaliser cet objectif a été de proposer le texte à trou à une dizaine de personnes qui n'avaient aucune connaissances du texte original. Les segments de texte candidats à l'effacement étaient retenus si les « prétesteurs » étaient capables de retrouver l'enrobage mais que toute tentative de combler un trou de la structure énumérative se soldait par un échec.

Pour bénéficier de l'A.P.A., l'intéressé doit en faire la demande. Il est indispensable de constituer un dossier et de l'adresser au conseil général. Les trois sites possibles pour retirer un dossier sont (énumérés ci-après. Le premier site est) le point d'accueil retraite du département. (Le deuxième site regroupe) les établissements d'hébergement de personnes âgées. (Le troisième et dernier site est) le centre médico-social. La mairie du domicile peut fournir un plan et une brochure explicative.

Pour prétendre à l'A.P.A., le demandeur doit constituer un dossier complet. Il est nécessaire de rassembler un certain nombre de documents. Les trois documents requis pour constituer un dossier sont (énumérés ci-après. Le premier document est) l'avis d'imposition du demandeur. (Le deuxième document est) la fiche individuelle d'état civil. (Le troisième et dernier document est) le certificat médical signé. Les rubriques du formulaire de demande doivent être dûment remplies.

Textes oralisés pour constituer les stimuli « test »

Pour _____ de l'A.P.A., l'intéressé doit en faire la _____. Il est indispensable de constituer un dossier et de l'adresser au conseil _____.

Les trois _____ possibles pour _____ un dossier sont (énumérés ci-après) :
(le premier site est) le _____ d' _____ retraite du département ;
(le deuxième document est) les _____ d' _____ de personnes âgées ;
(le troisième et dernier document est) le centre _____ - _____.

La _____ du domicile peut fournir un _____ et une brochure explicative.

Pour _____ à l'A.P.A., le demandeur doit constituer un _____ complet. Il est nécessaire de rassembler un certain nombre de _____.

Les trois _____ requis pour _____ un dossier sont (énumérés ci-après):
(le premier document est) l'avis d' _____ du demandeur ;
(le deuxième document est) la fiche _____ d' _____ ;
(le troisième et dernier document est) le _____ signé.

Les _____ du formulaire de demande doivent être dûment _____.

Textes à trous écrits pour le rappel indicé des « tests »

Entre les deux tâches de rappel libre et indicé de la phase test, est inséré le texte distracteur dont la transcription et la version à trou sont les suivantes :

La loi du 20 juillet 2001 pose comme principe général que l'allocation personnalisée d'autonomie est versée par le département où réside le demandeur. Pour tout cas particulier, s'adresser au service social le plus proche.

Texte oralisé pour constituer les stimuli « distracteur »

La loi du 20 juillet _____ pose comme principe général que l'allocation personnalisée d'_____ est versée par le _____ où réside le demandeur.
Pour tout cas _____, s'adresser au service _____ le plus proche.

Texte à trous écrit pour le rappel indicé du « distracteur »

Enfin un questionnaire est proposé pour l'évaluation subjective de la forme d'oralisation soumise et vérifier l'absence de familiarisation au thème des textes oralisés. Les aspects concernant l'intelligibilité et le naturel de la voix de synthèse sont évalués par le sujet sur une échelle allant de 1 à 5. afin de statuer sur l'existence ou non d'interactions entre la performance des sujets et la qualité de la voix écoutée. De plus un trop fort degré de connaissance de l'A.P.A. éliminera le sujet.

VIII.2 Procédure de présentation

Les trois groupes sont soumis chacun à une seule forme d'oralisation. Les trois formes d'oralisation seront présentées de manière successive : le sujet 1 sera soumis à la forme contrôle, le sujet 2 à la forme discursive, le sujet 3 à la forme prosodique, le sujet 4 à la forme contrôle, ainsi de suite... L'ordre de présentation des deux stimuli-test est contrebalancé. Les différentes phases du protocoles suivent le plan suivant :

- 1) Introduction à l'évaluation
 - a. Consigne sur l'écran

Vous allez écouter un texte court, lu à haute voix par un système de synthèse vocale.
Votre tâche sera de mémoriser ce que vous entendez. Une fois l'écoute terminée, vous devez rappeler par écrit le maximum d'information.
Soyez attentif car vous ne pourrez écouter le texte qu'une seule fois.
Cet exercice sera répété sur quatre textes différents.

b. consigne orale

Vous n'avez plus besoin d'être devant l'écran à partir de maintenant. Dès que l'écoute est terminée, vous écrirez dans le cadre ce que vous avez retenu. Vous pouvez raturer et réfléchir à voix haute pendant que vous écrivez. Vous ne pouvez pas prendre de notes pendant l'écoute. Lorsque vous aurez fini de rappeler les informations, vous me l'indiquerez et on passera à l'exercice suivant.

c. réglage du volume sonore pour une écoute au casque confortable

2) Entraînement

- a. écoute du stimulus audio
- b. rappel libre
- c. consigne orale rappel indicé

Vous devez remplir ce texte à trou qui correspond au texte que vous venez d'écouter. Chaque trou correspond à un seul mot .

d. rappel indicé

3) Test 1

- a. consigne orale

L'exercice suivant est du même type mais sur un texte plus long. Vous devez toujours mémoriser le maximum d'information pour les rappeler à l'écrit.

- b. écoute du stimulus audio
- c. rappel libre
- d. rappel indicé

4) Distracteur puis Test 2

- a. écoute du stimulus audio
- b. rappel libre
- c. rappel indicé

5) Questionnaire à remplir

Chapitre IX. Analyse des premiers résultats

Ce protocole a été utilisé sur un groupe de 30 sujets jeunes, et les données recueillies pour le rappel libre cumulé des deux stimuli audio ont été dépouillées pour la variable dépendante relative à la quantité d'informations justes rappelées et au degré de littéralité du rappel.

IX.1 Variables contrôles en fonction de la forme d'oralisation

En raison d'un effectif de sujets limité, nous avons choisi d'appliquer, pour les analyses statistiques, la famille des tests non paramétriques pour échantillons indépendants (test de Kruskal-Wallis pour l'effet global et test de Mann-Whitney pour les comparaisons 2 à 2).

IX.1.1 Effectif des sujets en fonction du sexe

Forme d'oralisation	Sexe		Total
	homme	femme	
forme contrôle	3	7	10
forme discursive	3	7	10
forme prosodique	2	8	10
Total	8	22	30

Sur 30 sujets, 22 sont de sexe féminin. La proportion d'hommes et de femmes pour chaque groupe expérimental est quasi-identique.

IX.1.2 Age des sujets selon les groupes expérimentaux

Forme d'oralisation	Moyenne
forme contrôle	23.200
forme discursive	22.400
forme prosodique	21.400

L'âge moyen de l'ensemble des sujets est de 22.3 ans.

IX.1.3 Niveau d'étude en fonction des groupes expérimentaux

Forme d'oralisation	Niveau d'étude		Total
	bac + 2	Bac + 3	
forme contrôle	5	5	10
forme discursive	4	6	10
forme prosodique	6	4	10
Total	15	15	30

L'ensemble des sujets est soit de niveau bac + 2 ou bac + 3. Les deux catégories de niveau d'étude sont correctement réparties dans chacun des 3 groupes.

IX.1.4 Activité impliquant l'utilisation de la mémoire

Forme d'oralisation	Utilisation de la mémoire		Total
	non	oui	
forme contrôle	7	3	10
forme discursive	4	6	10
forme prosodique	5	5	10
Total	16	14	30

Sur l'ensemble des 30 sujets, 14 déclarent exercer ou avoir exercé une activité impliquant l'utilisation de la mémoire (3 dans le groupe « forme contrôle », 6 dans le groupe « forme discursive » et 5 dans le groupe « forme prosodique »).

Le groupe « forme contrôle » comprend donc un déséquilibre des effectifs sur cette variable.

IX.1.5 Familiarité des sujets avec le thème abordé dans les SE : l'APA

Forme d'oralisation	Familiarité de l'APA		Total
	non	oui	
forme contrôle	10	0	10
forme discursive	6	4	10
forme prosodique	7	3	10
	23	7	30

Sur 30 sujets, 7 déclarent avoir déjà entendu parler de l'APA. Parmi ces sujets, 4 appartiennent au groupe « forme discursive » et 3 au groupe « forme prosodique ». Cette répartition est préjudiciable à l'analyse des effets principaux de la variable « forme d'oralisation ». En effet, selon nos hypothèses, les formes discursive et prosodique entraînent un meilleur rappel que la forme contrôle. Or, au vu de la configuration décrite ci-dessus, si de tels résultats sont observés, ils peuvent également être imputables à la surreprésentation, dans les groupes formes discursive et prosodique, de sujets connaissant le thème abordé par les textes à mémoriser.

Il convient donc d'étudier si ces sujets rappellent significativement plus d'informations que les sujets n'ayant pas de connaissances sur l'A.P.A.

Pour vérifier l'absence d'effet de cette variable, nous avons étudié le pourcentage de MP justes rappelés par rapport aux MP présentés.

Les résultats du test de Mann-Whitney indiquent que les médianes des pourcentages ne diffèrent pas significativement entre les deux groupes :

Familiarité de l'APA	N	Rang moyen	Somme des rangs
non	23	15.00	345.00
oui	7	17.14	120.00
Total	30		

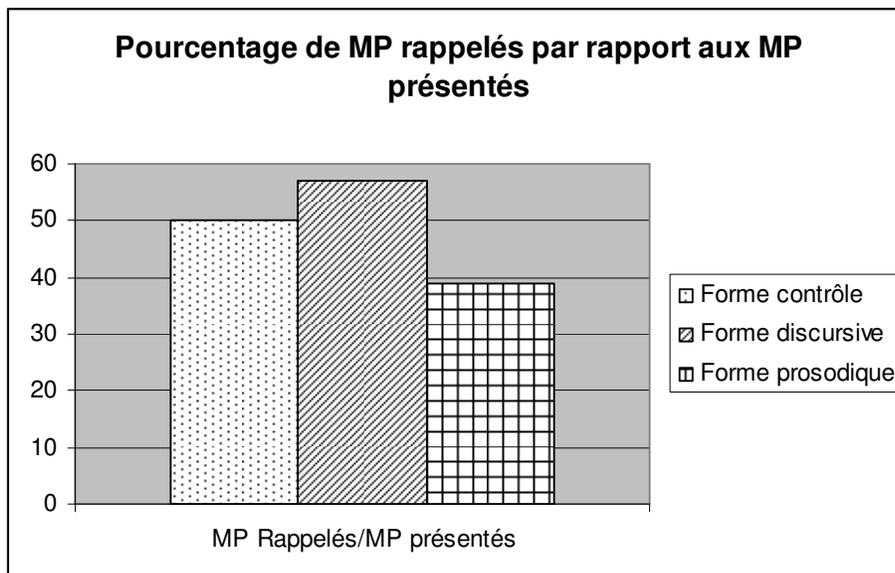
	pourcentage de MP justes rappelés
U de Mann-Whitney	69.00
W de Wilcoxon	345.00
Z	-.568
Signification asymptotique (bilatérale)	.570
Signification exacte [2*(signification unilatérale)]	.598

Nous avons donc choisi de conserver ces sujets dans l'analyse.

IX.2 Analyse d'une variable dépendante

Les résultats montrent qu'aucun résultat significatif sur le plan statistique ne semble émerger pour cette population d'usagers en terme de mémorisation/compréhension. Malgré tout, on observe, comme en atteste les paragraphes suivants, certaines observations purement descriptives intéressantes voire certaines tendances.

IX.2.1 Pourcentage de MP rappelés par rapport aux MP présentés



Sur un plan purement descriptif, nous constatons que le pourcentage de **mots pleins (MP) rappelés** par rapport aux mots pleins présentés est plus important avec la forme discursive (57%) qu'avec les deux autres formes (FC : 50% et surtout FPC : 39%):

Sur un plan statistique le test non paramétrique de Kruskal-Wallis ne révèle pas d'effet significatif de la forme d'oralisation sur la quantité d'informations rappelée (indépendamment de leur justesse) (khi-deux = 1.549).

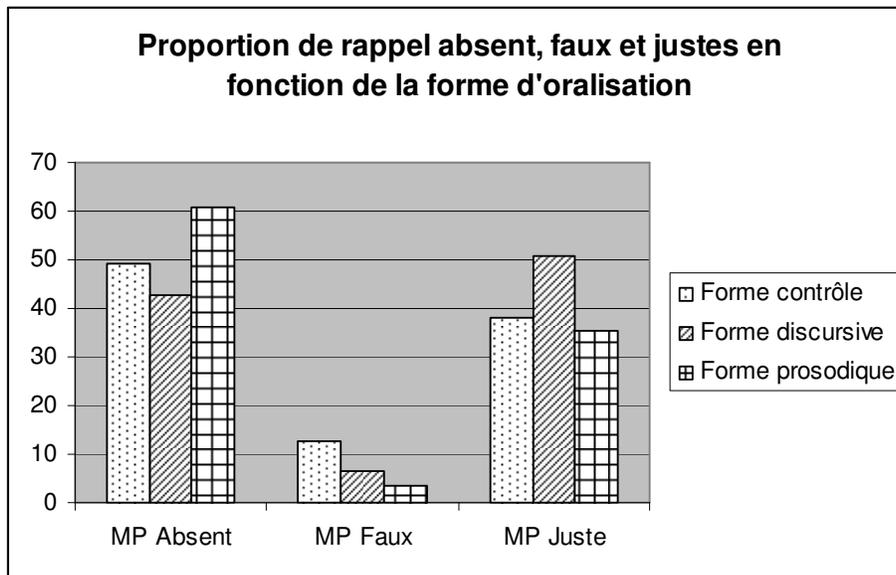
Test de Kruskal Wallis :

Forme d'oralisation	N	Rang moyen
forme contrôle	10	14.80
forme discursive	10	18.20
Forme prosodique	10	13.50
Total	30	

	% de MP rappelés
Khi-deux	1.549
Ddl	2
Signification asymptotique	.461

L'absence de résultat statistique peut être imputée à des valeurs d'écart-types élevées et hétérogènes (FPC=33.76 ; FDI=22.29 ; FC=16.15).

IX.2.2 Rappel faux, absent ou juste en fonction de la forme d'oralisation :



Les proportions de rappel de **MP justes** sont respectivement de 38.2 %, 50.9 % et 35.5 % pour les formes contrôle, discursive et prosodique. Sur un plan descriptif, on observe donc une supériorité de la forme discursive sur les deux autres formes. Par contre, sur le plan statistique, cette supériorité de la forme discursive n'est pas vérifiée [$\chi^2 (2, N = 30) = 1.954$].

Test de Kruskal Wallis :

Forme d'oralisation	N	Rang moyen
forme contrôle	10	14.10
forme discursive	10	18.75
Forme prosodique	10	13.75
Total	30	

	% de MP justes rappelés
Khi-deux	1.954
Ddl	2
Signification asymptotique	.376

Les proportions de rappel de **MP faux** sont respectivement de 12.7 %, 6.4 % et 3.6 % pour les formes contrôle, discursive et prosodique. Les formes discursive et prosodique semblent entraîner ainsi une proportion de rappel faux moins importante que la forme contrôle. Mais, sur le plan statistique, le test de Kruskal-Wallis ne fait pas apparaître d'effet global significatif [$\chi^2 (2, N = 30) = 3.431$].

Test de Kruskal Wallis :

Forme d'oralisation	N	Rang moyen
forme contrôle	10	18.95
forme discursive	10	15.30
Forme prosodique	10	12.25
Total	30	

	% de MP faux rappelés
Khi-deux	3.431
Ddl	2
Signification asymptotique	.18

Les proportions de rappel de **MP absents** sont respectivement de 49.1 %, 42.7 % et 60.9 % pour les formes contrôle, discursive et prosodique. Sur un plan descriptif, la forme prosodique entraîne une proportion d'absence de MP plus importante que les formes contrôle et discursive. Sur le plan statistique, le test de Kruskal-Wallis ne fait pas apparaître d'effet global significatif [χ^2 (2, N = 30) = 1.549].

Test de Kruskal Wallis :

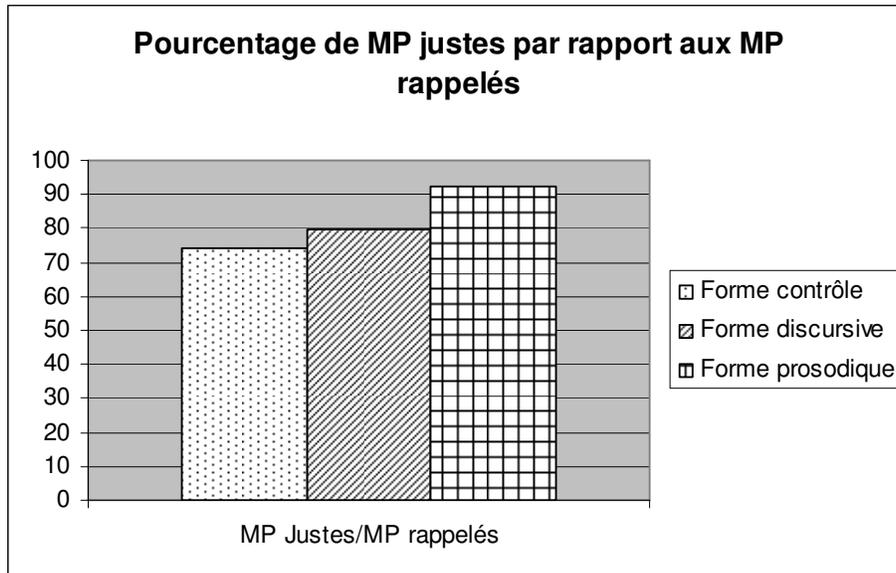
Forme d'oralisation	N	Rang moyen
forme contrôle	10	16.20
forme discursive	10	12.80
Forme prosodique	10	17.50
Total	30	

	% de MP absents rappelés
Khi-deux	1.549
Ddl	2
Signification asymptotique	.461

Ainsi, bien qu'il apparaisse que la forme contrôle est meilleure que la forme prosodique en terme de quantité d'informations rappelées, les résultats suivants nous poussent à constater que : **la quantité d'informations rappelées pour la forme contrôle n'est pas associée à leur justesse**. Si la forme contrôle favorise un rappel quantitativement plus important que la forme prosodique, elle favorise également un rappel qualitativement « plus faux ».

IX.2.3 Pourcentage de MP justes par rapport à l'ensemble des MP rappelés ::

Si nous avons observé la justesse relativement à l'information présentée, un autre angle de vue est celui de la justesse relativement à ce qui a été réellement rappelé. Afin de pouvoir procéder à un calcul statistique, nous ne pouvions considérer pour cet indicateur les sujets n'ayant rappelé aucune information. Cela nous a conduit à l'élimination de trois sujets dans un effectif déjà limité.



Test de Kruskal Wallis :

Forme d'oralisation	N	Rang moyen
Forme contrôlée	10	11.10
Forme discursive	10	13.45
Forme prosodique	7	18.93
Total	27	

	% de MP faux rappelés
Khi-deux	4.251
Ddl	2
Signification asymptotique	.119

Malgré une absence de significativité statistique [$\chi^2(2, N = 27) = .119$], nous voyons que la tendance est en faveur de la forme prosodique (92,4% contre respectivement 79,9% pour la FDI et 73,9% pour la FC) ; ainsi celle-ci entraînerait un rappel quantitativement frileux mais dont le rapport quantité/qualité est le meilleur : on se rappelle de peu de chose mais on est assez sûr des informations retenues. Une explication peut-être liée au fait que les stratégies d'extraction prosodique par la topicalisation, la focalisation étendue et la rhématisation améliorent la mémorisation des mots marqués mais cela au détriment des autres segments mis ainsi en retrait. Nous retrouvons ici la remarque formulée par Glynn (1979) et évoquée auparavant selon

laquelle ce même effet s'observe pour le rappel des segments de texte marqués typographiquement.

Finalement, ce premier jeu de test nous indique une tendance qu'il reste à démontrer statistiquement :

- d'un certain point de vue **la forme discursive semble être meilleure que les deux autres formes pour une tâche de mémorisation**. En effet, c'est celle où on observe la plus grande proportion de mots pleins justes rappelés, et le moins de mots absents. En notant tout de même que, pour ce qui est de la proportion de mots pleins faux, elle est deux fois plus élevée que pour la forme prosodique mais deux fois moins élevée qu'avec la forme contrôle ;
- D'un autre point de vue **la forme prosodique peut être considérée comme meilleure que les deux autres formes pour un tâche de mémorisation**, puisque les éléments importants sont retenus bien qu'un grand nombre d'éléments ne soient pas rappelés. Notons qu'une explication alternative est envisageable puisque les trois sujets n'ayant fait aucun rappel appartenaient tous au groupe soumis à la forme prosodique : leur élimination a donc contribué à une augmentation du score de la forme prosodique par rapport aux autres (d'un point de vue descriptif, les résultats restent tout de même encore en faveur de la forme prosodique lorsque nous procédons à un rééquilibrage consistant à éliminer également les trois moins bons sujets des deux autres groupes de sujet).

Nous retrouvons, pour **conclure** vis à vis de ces deux angles de vue, la résistance que nous évoquions pour **prédire** la supériorité d'une forme par rapport à l'autre (résistance qui avait conduit à opter pour une hypothèse de différence simple). En fait, cette difficulté de prédiction est peut-être corrélée à la nécessité de choisir ce qu'il faut entendre par « une forme est supérieure à une autre » :

- le caractère implicite de la forme prosodique pour évoquer l'inscription graphologique du texte est accompagné d'une plus grande capacité de marquage que les marques lexico-syntaxiques de la forme discursive ;
- la longueur du message impliqué par l'utilisation d'une forme discursive est accompagnée par une plus grande quantité d'éléments à rappeler que la forme prosodique.

Cela dit, si nous parvenions à valider statistiquement ces résultats, nous avaliserions d'une certaine manière l'hypothèse 1 spécifique à la variable F : **les formes d'oralisation restituant les aspects structuraux produisent de meilleures performances que la forme contrôle qui ne les restitue pas.**

« *La mémoire ne filme pas, la mémoire photographie.* »

Milan Kundera, L'immortalité

Partie D Images De Page et Multimodalité

Une hypothèse expliquant les limites du protocole expérimental, dont les résultats sont décrits dans la partie précédente, est l'impossibilité d'obtenir une équivalence intermodale prenant en compte le *first glance* impliqué par la lecture visuelle d'un texte.

Sur cette base, cette partie est consacrée à l'étude d'un concept original qui nous permet d'envisager l'élaboration d'interfaces multimodales impliquant des moyens d'interaction originaux : les Images De Page (IDP).

Après avoir exposé l'origine de ce concept et ses principes fondateurs, nous proposons son utilisation pour la conception d'interfaces plus adaptées à la communication homme-machine en situation dégradée.

La plausibilité cognitive des Images de Page est démontrée à travers deux protocoles expérimentaux qui rendent intéressant le développement de deux prototypes visant à pallier une surface de présentation de l'information réduite ou l'impossibilité d'utiliser la modalité visuelle.

Chapitre X. *First Glance* et Images De Page

Si le MORTELS permet de reformuler l'inscription graphologique démarcative et distinctive, la transposition du texte dans la modalité orale subit les contraintes inhérentes à cette modalité et il semble difficile, dès lors, de restituer les fonctions cognitives étroitement liées à l'exploitation de la spatialité du document et qui agirait globalement et au premier regard sur le texte. Bien que la caractéristique anticipatrice de certaines fonctions peut être approchée en restituant très tôt dans le signal de parole certaines marques, l'effet *first glance* du texte semble bien impossible à restituer à l'oral. Ainsi ce travail, à partir de l'étude de la transmodalité de l'écrit vers l'oral, met au jour une faiblesse à laquelle nous proposons d'essayer de répondre par une stratégie multimodale.

Nous formulons pour cela une hypothèse sur un mécanisme unique responsable de ces différentes fonctions sur le plan cognitif : l'allègement de la charge mentale en mémoire de travail. Les traces visuelles issues de la réduction de métaphrases, permettraient d'alléger la charge cognitive imposée au sous-système de la mémoire de travail responsable du traitement des informations verbales en sollicitant en parallèle un autre sous-système de l'architecture cognitive décrite dans (Baddeley, 1986).

Ainsi, puisque

**l'image nous donne le réel dans sa totalité et surtout son immédiateté :
et la parole vient encore, de surcroît, s'ajouter au visuel et le renforcer
(Dagonnet, 1998)**

alors, pour restituer à l'oral la fonction globale d'allègement de la charge cognitive portée par les marques de MFM, ainsi que ses fonctions d'anticipation du traitement du texte, nous proposons une présentation basée sur une « multimodalisation » non redondante du texte (et de l'hypertexte) afin de distribuer sur deux canaux sensoriels différents le contenu « articulable » et l'architecture globale du message : un système de synthèse vocale à partir de textes se chargera du contenu et les IDP de sa structure visuelle.

Nous avons déjà évoqué, dans le chapitre consacré à la présentation du Modèle d'Architecture Textuelle, ce moyen visuel alternatif aux métaphrases pour représenter l'architecture du texte. Cette représentation est, à l'origine, un outil de communication et de discussion interne aux concepteurs du MAT, né de la nécessité d'appréhender les architectures plus intuitivement que par les formalismes existants et de favoriser ainsi, à la fois la conception de nouvelles expérimentations et le dialogue (homme/homme !) lors de collaborations interdisciplinaires. Certains principes ont alors été posés et ont

permis d'ériger un système notationnel dédié à la représentation de l'architecture du texte (Luc, 2001).

Les principes qui définissent une image de page s'appuient sur quatre impératifs auxquels ils doivent répondre : permettre l'observation de phénomènes architecturaux visuels, syntaxiques et rhétoriques ; jouer sur une représentation économique du texte pour ne laisser filtrer que les informations jugées utiles dans un objectif donné ; faciliter l'étude des propriétés des modèles issus de différentes disciplines et la composition de ces modèles ; servir d'outil d'expérimentation. Nous proposons une cinquième direction pour l'utilisation de ce système : la réalisation d'interfaces interactives multimodales pour la communication homme/machine.

X.1 Principes fondateurs

Les propriétés que les concepteurs ont voulu donner à ce système se réfèrent à la théorie de la notation et de la notationnalité de Nelson Goodman (Goodman, 1990) et peuvent s'exprimer de la manière suivante :

- conservation et utilisation par des acteurs différents, et manipulation réglée des objets représentés à travers leur notation ;
- normalisation de la représentation de phénomènes fonctionnellement équivalents mais superficiellement divers ;
- identification de la sélection des traits pertinents de la notation à l'exemplification ;
- mise au clair des fonctions propres des éléments alphanumériques et graphiques.

Il existe 5 types d'image de pages selon la finesse que l'on veut donner à la granularité de la représentation :

- l'image minimale qui représente l'organisation du texte en unités textuelles (UT) en conservant la casse du début des UT (M/m pour respectivement majuscule/minuscule) et la ponctuation encadrante. La suite de caractères constituant les UT est signifiée par un trait horizontal ;
- l'image augmentée 1 qui ajoute à la précédente les blancs utilisés pour rendre perceptible l'architecture de l'objet textuel représenté (indentation, marge, interlignage, ...). Ces marques sont représentées par des flèches verticales et horizontales de hauteur et longueur relative ;

- l'image augmentée 2 qui ajoute à la précédente les segments de texte mis en saillance dans l'OT représenté (guillemets, gras, italique, ...) et les marques lexico-syntaxiques (organisateur textuels, connecteurs, ...);
- les images augmentées 3 et 4 qui s'attaquent respectivement à la représentation des relations de dépendance syntaxique et à la représentation des relations de dépendance rhétorique.

La granularité des deux dernières représentations semble trop fine dans le cadre de cette étude puisqu'elles relèvent d'un niveau de description de la structure informationnelle de l'OT qui va au delà de ce que le *first glance* est *a priori* susceptible de nous fournir. Aussi, nous pensons que l'image augmentée 2 permet une représentation de l'OT intéressante pour un système de présentation de l'information.

Si la question de l'utilité d'une telle représentation pour une présentation multimodale du texte nécessite d'être posée à travers la conception et l'évaluation de prototypes, la première nécessité est de s'assurer de l'utilisabilité de ce principe. La section suivante s'attache donc à montrer comment nous avons tenté de prouver la plausibilité cognitive des IDP.

X.2 Évaluation de la plausibilité cognitive des IDP

D'après l'hypothèse avancée ci-dessus, la présentation conjointe de l'oralisation du contenu du texte et de sa représentation sous la forme d'une IDP ayant uniquement vocation à représenter les phénomènes architecturaux, permettrait de restituer la fonction d'allègement de la charge cognitive en sollicitant deux sous-systèmes de traitement de la mémoire de travail au lieu d'un.

Si ce type de présentation devrait être efficace au regard de nos fondements théoriques, il n'en reste pas moins que ce mode de présentation bi-modal peut présenter des inconvénients :

- en raison de son manque de familiarité pour les sujets, ce mode de présentation ne nécessite-t-il pas de faire l'objet d'un apprentissage ?
- si la représentation de la structure du texte sous la forme d'une image de page a fait ses preuves en tant que système notationnel, la question de sa plausibilité cognitive reste posée.

Nous avons donc mené deux études expérimentales visant à évaluer si l'IDP constitue une représentation plausible sur le plan cognitif, c'est-à-dire manipulable mentalement par le sujet. Nous présentons dans la section suivante les protocoles

expérimentaux utilisés, une description et une interprétation des résultats au regard de la question posée.

X.2.1 Premier protocole

L'étude que nous avons menée visait à évaluer si le sujet humain comprend la représentation du texte sous la forme d'une IDP en testant la faisabilité d'une tâche impliquant sa manipulation. Nous avons choisi d'évaluer cette compréhension en demandant au sujet de choisir, parmi plusieurs images de page (IDP), celle qui représente le mieux un texte oralisé par un système de synthèse vocale à partir de textes (Maurel, 2003a, 2004b).

X.2.1.1 Hypothèses, méthode et prédictions

Notre hypothèse générale est que le sujet humain est capable d'associer à un texte oralisé sa représentation sous la forme d'une IDP. Pour tester cette hypothèse générale, la méthode que nous avons employée a consisté à faire écouter à chaque sujet deux formes orales d'une énumération, exploitant la stratégie discursive interprétative et la forme prosodique en contexte. Simultanément à l'oralisation du texte, trois IDP étaient présentées au sujet sur l'écran de l'ordinateur. La tâche du sujet était de classer les IDP par ordre décroissant en fonction de leur capacité à représenter le texte oralisé.

Plus spécifiquement, voici quelles sont nos hypothèses :

- H1 : les sujets sont capables d'identifier une énumération à l'oral, qu'elle soit réalisée sous une forme discursive ou prosodique et ils ont des connaissances sur la réalisation des énumérations à l'écrit. Ils sont alors capables d'associer une réalisation adéquate d'une énumération sous la forme d'une IDP à une énumération entendue à l'oral ;
- H2 : les sujets sont capables de juger une représentation sous la forme d'une IDP incompatible si celle-ci représente un autre objet textuel que l'énumération ;
- H3 : les sujets sont capables de repérer qu'une énumération oralisée sous une forme discursive peut être représentée par deux IDP:
 - o sous la forme d'une image de page représentant une suite de phrases. Il s'agit alors d'une réalisation non redondante puisque les marques structurales sont réalisées dans la structure énumérative sous une forme lexico-syntaxique ;
 - o sous la forme d'une image de page représentant une énumération marquée par un *pattern* morphe-dispositionnel "classique".

Les sujets écoutent tour à tour les deux formes, la moitié des sujets dans l'ordre FPC-FDI, l'autre moitié dans l'ordre FDI-FPC. Après ou pendant chaque écoute, le sujet effectue le classement en cliquant tour à tour sur chaque IDP. Pour chaque écoute, l'expérimentateur demande aux sujets les raisons de leur classement et recueille leurs verbalisations. La consigne comprenait également une courte définition des IDP spécifiant qu'il s'agissait d'une représentation mettant en évidence uniquement les aspects visuels du texte. Une fois les deux écoutes et les deux classements effectués, le sujet remplissait une fiche signalétique. L'expérience durait environ 20 minutes.

Les mesures retenues pour l'analyse sont le temps de réponse du sujet, son ordre de classement des IDP et les catégories de justifications avancées par le sujet. Le matériel expérimental que nous avons utilisé est constitué de deux formes oralisées provenant d'un même texte décliné soit sous une forme discursive, soit sous une forme réduite et marquée sur le plan prosodique. Dans les deux cas, la synthèse utilisée est Proverbe, d'Elan Speech. Le texte original contient deux phrases introductrices, l'amorce d'une énumération, une énumération à 4 items et une phrase finale. Le thème général porte sur les documents qu'il faut fournir pour constituer un dossier complet de demande d'Allocation Personnalisée d'Autonomie (APA). Voici la forme discursive écrite oralisée par le système de synthèse (les segments entre parenthèses sont effacés dans la forme prosodique au profit de marques prosodiques) :

Pour prétendre à l'A.P.A., le demandeur doit constituer un dossier complet. Il est nécessaire de rassembler un certain nombre de documents. Les quatre documents requis pour constituer un dossier sont (énumérés ci-après. Le premier document regroupe) les derniers avis d'imposition du demandeur(. Le deuxième document est) la fiche individuelle d'état civil(. Le troisième document est) le certificat médical détaillé(. Le quatrième et dernier document est) le double d'une facture EDF. Les rubriques du formulaire de demande doivent être dûment remplies.

Trois IDP sont présentées au sujet :

- une IDP dite " Structure Énumérative " (SE) exploitant un *pattern* morpho-dispositionnel " classique " (cf. a) Figure 17);
- une IDP dite " Bloc " n'exploitant aucune marque de mise en forme morpho-dispositionnelle spécifique à un objet textuel particulier. Cette IDP est possible pour représenter l'énumération, mais elle n'en constitue pas une représentation spécifique (cf. b) Figure 17) ;
- une IDP dite " Autre " présentant des marques morpho-dispositionnelles pouvant indiquer un titre ou une structure d'insistance dans le texte. Cette IDP

est supposée ne pas représenter l'énumération du fait qu'elle n'en constitue pas une représentation adéquate (cf. c) Figure 17).

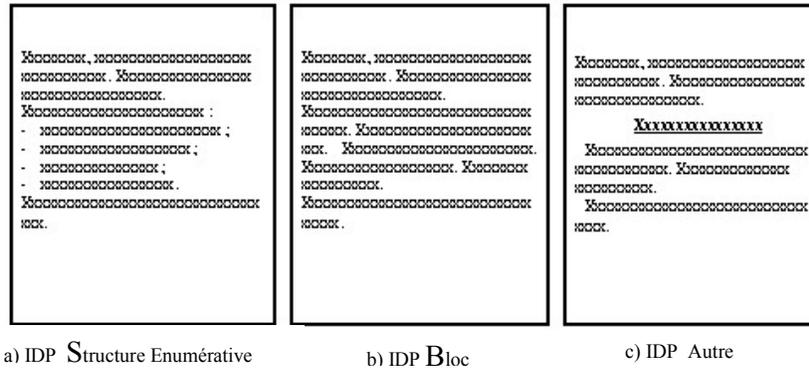


Figure 17 ; Images de page SE, Bloc, Autre

Les 3 images sont présentées côte à côte. Pour la première écoute, l'ordre de présentation des images sur l'écran est, de gauche à droite, bloc-autre-SE, alors que pour la deuxième écoute, il est autre-SE-bloc. Le choix de ces deux ordres s'est fait de façon à ce qu'aucun des deux ordres, dans le cas où le sujet réponde au hasard et dans l'ordre naturel gauche-droite, ne favorise des réponses allant dans le sens attendu par nos hypothèses. Le changement d'ordre pour les deux écoutes visait à s'assurer qu'il n'y ait pas une automatisation de la réponse entre les deux présentations.

Vingt étudiants de troisième cycle ont participé à l'expérience, deux femmes et 18 hommes. Tous étaient de langue maternelle française et seul un sujet a déclaré avoir des troubles auditifs mais qui ne l'ont pas gêné pour l'expérience. La moyenne d'âge des sujets était de 24 ans.

Trois prédictions découlent des trois hypothèses formulées plus haut :

- P1 : quelle que soit la forme entendue, l'ordre de classement favorise « l'IDP SE » en premier choix ;
- P2 : quelle que soit la forme entendue, « l'IDP Autre » est plus choisie en troisième choix, qu'en deuxième et qu'en premier choix ;
- P3 : la forme discursive favorise davantage le choix de « l'IDP Bloc » en premier que la forme prosodique.

X.2.1.2 Résultats

Nous présenterons ici les résultats quantitatifs obtenus sur les ordres de classement choisis par les sujets et les analyses statistiques effectuées ; dans la sous-section suivante, nous discuterons les résultats obtenus à la lumière de l'analyse qualitative des verbalisations de sujets.

Le Tableau 6 indique les effectifs de chaque position possible dans le classement pour chaque IDP.

	Premier choix	Deuxième choix	Troisième choix
SE	33 (A)	6	1
Bloc	6 (B)	21	13
Autre	1 (C)	13 (D)	26 (E)

Tableau 6 : Effectifs pour chaque couple (type d'image de page ; position)

Chacun des 20 sujets devant donner un ordre de classement pour chacune des deux écoutes, nous aboutissons au total à 40 ordres, dont 20 pour la FDI et 20 pour la FPC. Pour tester la première prédiction selon laquelle, le premier choix dans l'ordre de classement porte significativement plus sur « l'IDP SE » que sur les « IDP Bloc ou Autre », il faut vérifier que $(A) > (B)$ et $(A) > (C)$. Un test d'adéquation à la loi binomiale¹⁴ indique qu'il y a **significativement plus de premiers choix portant sur l'IDP SE que sur l'IDP Bloc et que sur l'IDP Autre** avec un seuil de significativité inférieur à 1 %.

Pour tester la seconde prédiction selon laquelle le dernier choix du classement favorise « l'IDP Autre » par rapport aux « IDP SE et Bloc », il faut vérifier que $(E) > (D)$ et $(E) > (C)$. Les analyses statistiques révèlent avec un seuil de significativité inférieur à 1% qu'il y a **plus de derniers choix portant sur « l'IDP Autre » que sur « l'IDP Bloc » et que sur « l'IDP SE »**.

Pour le test de la troisième prédiction, nous nous sommes intéressés à l'ordre de classement des sujets selon la forme d'oralisation écoutée. Le Tableau 7 présente les

¹⁴ le test d'adéquation à la loi binomiale est celui qui a été utilisé pour l'ensemble des analyses menées.

effectifs obtenus pour chaque ordre de classement possible selon la forme d'oralisation écoutée.

	FDI	FPC
SE-Bloc-Autre	8	12
SE-Autre-Bloc	6	7
Bloc-SE-Autre	6 (F)	0(H)
Bloc-Autre-SE	0 (G)	0(I)
Autre-SE-Bloc	0	0
Autre-Bloc-SE	0	1

Tableau 7 : Effectifs pour chaque ordre selon la forme d'oralisation

Pour tester la troisième prédiction selon laquelle la forme discursive donne lieu à plus de premiers choix portant sur l'IDP Bloc que la forme prosodique, il faut vérifier sur un plan statistique que **(F) + (G) > (H) + (I)**. Les analyses statistiques révèlent que cette différence est significative pour un seuil de 5 % (mais pas pour un seuil de 1 %). **Il y a significativement plus de premiers choix portant sur « l'IDP Bloc » lors de l'écoute de la forme discursive que lors de l'écoute de la forme prosodique.**

Toutefois, nous avons souhaité vérifier si cette différence portant sur le premier choix lorsque le sujet écoute la forme discursive n'était pas influencée par l'ordre de présentation des deux formes orales aux sujets. Rappelons que, pour la moitié des sujets, l'ordre de présentation était FPC-FDI tandis que l'autre moitié écoutait d'abord la forme discursive. L'étude des effectifs des ordres de classement en fonction de la forme d'oralisation et de son ordre de présentation semble révéler un effet d'interaction : le fait que la forme discursive favorise le choix de « l'IDP Bloc » en premier ne s'observe que lorsque la forme discursive est présentée en deuxième, après la forme prosodique, au sujet.

X.2.1.3 Discussion des résultats

Quelle que soit la forme entendue, les sujets favorisent un premier choix portant sur « l'IDP SE » dans leur classement. L'étude de leurs justifications montre que tous les sujets ayant fait ce choix le justifient par une correspondance entre l'énumération identifiée à l'oral et celle représentée à l'écrit et ceci de façon explicite. **Ainsi, les sujets sont capables d'identifier l'objet textuel « énumération » à l'oral, d'identifier ce même objet textuel représenté par des marques de MFM à l'écrit, et d'associer les deux.**

« L'IDP Autre » est davantage choisie en troisième qu'en deuxième ou en premier, quelle que soit la forme entendue. **Les sujets sont capables d'identifier l'énumération à l'oral et de constater l'incompatibilité avec l'image représentant un titre ou une insistance.** Les catégories de justifications avancées pour le troisième choix portant sur

« l'IDP Autre » renforcent cette explication ; en effet, sur 26 troisièmes choix portant sur « l'IDP Autre », 21 justifications évoquent une mauvaise correspondance entre le message entendu et ce qui est représenté sur l'image de page. Cinq sujets évoquent un choix par défaut.

L'écoute de la forme discursive de l'énumération favorise davantage en premier choix « l'IDP Bloc » que la forme prosodique. Ce résultat n'est vrai que lorsque la forme prosodique est écoutée en premier. Deux hypothèses explicatives peuvent rendre compte de cette interaction entre la forme écoutée et sa position dans l'ordre de présentation des deux formes. Une première explication est que lorsque l'ordre de présentation est FPC-FDI, lors de l'écoute de la FPC, le sujet a n'a eu aucune difficulté à y associer l'image de page SE ; lors de sa deuxième écoute portant sur la FDI, il aurait " envie " de donner le même ordre de classement avec « l'IDP SE » en premier choix mais il se sent obligé de changer d'ordre. Si ce changement d'ordre se fait au hasard, il devrait y avoir alors autant de premier choix portant sur « l'IDP Bloc » que sur « l'IDP Autre » associée à l'écoute de la forme discursive. Or, ce n'est pas le cas : sur 7 sujets ayant associé à la forme prosodique écoutée en premier, un premier choix de « l'IDP SE », 6 sujets ont choisi « l'IDP Bloc » pour la FD. Il semble donc que le changement d'ordre conditionné par l'ordre d'écoute des formes ne se fasse pas au hasard mais en faveur de « l'IDP bloc ». Une explication alternative est que l'effet d'interaction est dû à la différence de statut prise par la forme prosodique selon sa position dans l'ordre de présentation ; lorsque le sujet écoute en premier la forme discursive, ce qu'il perçoit en priorité, c'est la présence d'une énumération compatible avec la représentation de l'image de page SE ; il choisit donc en premier cette image. Lorsque le sujet écoute en premier la forme prosodique, il associe sans difficulté l'image de page SE à la forme prosodique ; lorsqu'il entend ensuite la forme discursive, il perçoit que l'énumération est différente de celle entendue sous une forme prosodique. Prenant celle-ci pour référent, il choisit une autre image que celle choisie en premier pour la forme prosodique et son choix porte sur celle qui est possible bien que moins probable, « l'IDP Bloc ». L'étude des justifications données par les sujets semble confirmer cette interprétation de l'interaction ; en effet, les 6 sujets ayant choisi en premier « l'IDP Bloc » considèrent que cette image correspond à la forme discursive car elle représente une suite de phrases. Certains sujets comparent la forme discursive à la forme prosodique qu'ils ont écoutée avant et évoquent le fait que « c'est des phrases plutôt que des points ».

En revanche, un résultat obtenu semble contradictoire avec nos hypothèses : la forme discursive, lorsqu'elle est écoutée en premier, donne uniquement lieu à des premiers choix portant sur « l'IDP SE », et donne autant de seconds choix portant sur « l'IDP

Bloc que Autre », alors que dans la logique de nos hypothèses, elle aurait dû davantage donner lieu à des deuxièmes choix portant sur « l'IDP Bloc ». L'étude des catégories de justifications évoquées pour un deuxième choix portant sur les « IDP Bloc ou Autre » pour la forme discursive écoutée en premier n'éclaire pas ce phénomène.

Malgré ce dernier résultat peu compatible avec nos hypothèses, l'hypothèse générale selon laquelle le sujet humain est capable d'associer au texte oralisé sa représentation sous la forme d'une IDP semble validée. Les résultats de ce premier protocole nous permettent donc d'envisager une présentation multimodale du texte combinant l'oralisation du texte et sa représentation visuelle sous la forme d'une image de page.

X.2.2 Second protocole

X.2.2.1 Hypothèses, méthode et prédictions

Le protocole vise à la fois à renforcer nos résultats sur la plausibilité cognitive des IDP et à valider la valeur « désambiguïsatrice » de nos stratégies : il s'agit pour le sujet de choisir parmi deux IDP celle qui correspond le mieux à l'oralisation du texte qu'il vient d'écouter selon 3 formes d'oralisation (en inter-sujets) : FPC, FDI et une forme dite de contrôle (FC) conçue de la même manière que pour les stimuli du protocole visant à évaluer l'impact des stratégies sur la mémorisation ; la fabrication de ces trois stimuli a été réalisée de la manière suivante :

- la FDI a été directement oralisée par le moteur de synthèse (texte encadré complet) ;
- la forme contrôle a été créée à partir de la FDI en supprimant les segments du signal acoustique correspondant aux marques lexico-syntaxiques « architecturantes » (texte entre parenthèse dans l'encadré) ; ces marques étaient donc effacées et non remplacées ;
- la FPC a repris la forme contrôle mais augmentée de marques prosodiques « équivalentes » fonctionnellement aux marques lexico-syntaxiques supprimées.

Une oralisation partielle de ces 3 stimuli est présentée aux sujets (l'amorce et le premier item d'une énumération). La question était de savoir si le sujet est capable de déclencher lui-même à la fin de l'écoute le raisonnement selon lequel il s'agit de l'oralisation d'un début de texte (un entraînement est effectué sur un texte complet et il est simplement précisé qu'il s'agira ensuite du même type d'exercice).

La prédiction est que, si l'on considère comme meilleure, la forme d'oralisation qui permet pour le plus grand nombre de sujets de désambiguïser le stimuli dès la fin du

premier item, alors, pour cette tâche : FDI supérieure ou égale à FPC elle même très supérieure à la forme contrôle.

Afin de renforcer l'ambiguïté de notre stimulus et susciter ainsi un effet négatif sur le traitement syntactico-sémantique de la structure du message (effet dit de « *garden path* ») nous avons conservé le classifieur de l'amorce et repris le même adjectif numéral en début de premier item :

Les **trois** points forts de la région Midi-pyrénées sont (énumérés ci-après.) (Le premier point fort comprend) les **trois** cultures agricoles locales.

La tâche d'identification d'image de page débute par une phase d'entraînement. On propose ainsi aux sujets un texte d'entraînement présentant des marques de MFM autres que celles caractéristiques des énumérations, qui permettront au sujet de se familiariser d'une part avec la synthèse vocale et d'autre part avec la notion d'IDP :

**CAISSE D'ALLOCATION FAMILIALE
DE LA HAUTE-GARONNE**

Comment nous contacter ?

La CAF reçoit le public du lundi au vendredi, en journée continue de 8 h 15 à 16 h 15.

Attention ! Cette nouvelle antenne se différencie des accueils administratifs traditionnels, et, notamment, de la permanence tenue jusqu'à ce jour 2 rue Emile Zola.

Téléphone : 08.21.01.99.31

Nos prédictions sont alors que :

- malgré la reprise du numéral dans le premier item qui semble appuyer l'interprétation d'une phrase autonome, la forme prosodique permettra de conférer au message, dans les mêmes proportions que la forme discursive, le statut de début d'énumération ;
- malgré un texte d'entraînement sur un texte complet et sans précision supplémentaire, la plausibilité cognitive des images de page est assez forte

pour permettre d'inférer le statut d'un objet textuel partiellement représenté par ce moyen.

La consigne donnée aux sujets est la suivante :

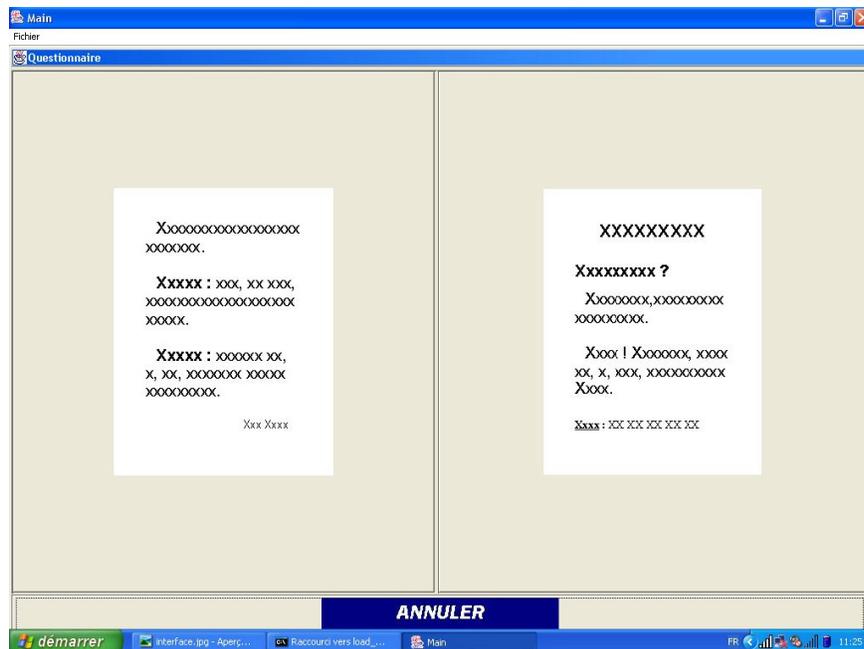
Vous allez entendre à l'aide d'un casque un texte lu à haute voix par un système de synthèse vocale. Après l'écoute, deux images, dites images de page, vont apparaître à l'écran.

Une image de page est une représentation schématique qui montre le texte comme une sorte de photographie vue de loin. Les mots sont en fait cachés en remplaçant les lettres du texte par des X mais la structure visuelle est ainsi conservée.

L'exercice consiste à choisir une image de page qui correspond le mieux à ce que vous venez d'entendre. Signalez à l'expérimentateur que vous êtes prêt, celui-ci déclenchera alors l'écoute. Restez concentré car le texte ne pourra être écouté qu'une seule fois. L'apparition des deux images de page sera déclenchée automatiquement dès la fin de l'écoute.

Il vous suffira alors d'indiquer à l'expérimentateur l'image de votre choix.

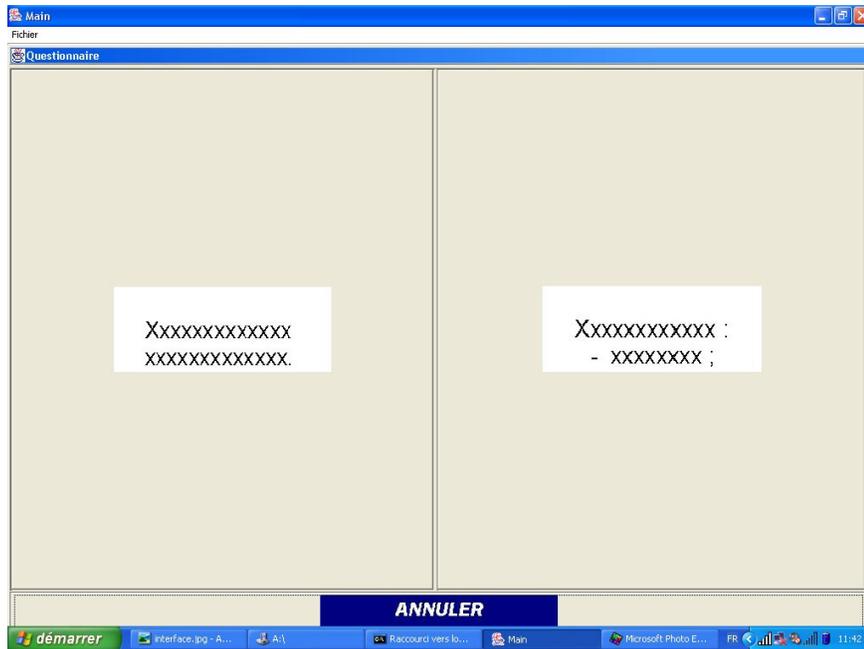
Pour le texte d'entraînement, chaque groupe expérimental sera soumis à la même forme d'oralisation du texte. Les deux IDP proposés sont :



Pour le stimulus test, chaque groupe de sujet sera soumis à une forme d'oralisation différente, la consigne intercalée ne donnant aucune précision supplémentaire sur le type de texte oralisé ; au contraire nous annonçons :

Il s'agit maintenant du même type d'exercice. Il faudra de nouveau choisir l'image qui, selon vous, correspond le mieux à ce que vous avez entendu.

Les deux IDP proposés après l'écoute sont IDP1 (à gauche dans la capture d'écran ci-après) et IDP2 (à droite dans la capture d'écran ci-après).



Dans les deux exercices, l'ordre de présentation des IDP fera l'objet d'un contre-balancement pour éviter un éventuel effet d'ordre dans le choix des sujets.

Après chaque choix, nous demandons aux sujets de l'expliquer. On s'assure ainsi pour la tâche d'entraînement que le concept d'image de page a été convenablement compris. Pour le test, cela nous permet d'établir quels sont les indices qui ont déterminé leur choix.

X.2.2.2 Résultats

Forme d'oralisation	Choix IDT test		Total
	IDP1	IDP2	
Forme contrôle	8	2	10
Forme discursive	0	10	10
Forme prosodique	0	10	10
Total	8	22	30

Dans le groupe « forme contrôle », 8 sujets sur 10 choisissent l'IDP1 seulement 2 sujets choisissent l'IDP2.

Dans les groupes « forme prosodique » et « forme discursive », la totalité des sujets choisit l'IDP2.

Ces résultats vont dans le sens de nos prédictions :

- les sujets appartenant aux groupes des formes discursives et prosodiques ont tous su « entendre » le début d'une énumération et inférer le statut de début d'IDP correspondante ;
- les sujets appartenant au groupe de la forme contrôle ont à 80% déduit de l'écoute une IDP correspondante complète représentant une phrase autonome.

Nous noterons que les résultats vont presque trop dans le sens de nos hypothèses et que cela démontre probablement la trop grande trivialité de nos hypothèses.

Cette pré-étude nous ouvre toutefois des perspectives pour l'utilisation des IDP dans un contexte d'interaction homme/machine. Ainsi, deux cadres d'intérêts nous semblent pouvoir exploiter les IDP dans des situations de présentation de l'information dégradées par l'environnement ou le type d'utilisateur.

Chapitre XI. Perspectives à court terme

XI.1 IDP, PDA et malvoyance

L'accès à l'information sur petit écran, ou sur un écran normal avec des caractères fortement grossis, souffre de difficultés qui préoccupent depuis un certain temps les chercheurs du domaine de l'IHM.

Bien que quelques études sur la présentation d'information sur petit écran montrent peu d'effets sur la compréhension, elles mettent en évidence une influence importante sur le débit de lecture, sur la présentation des graphiques, tableaux, formulaires et sur la sélection de texte et l'entrée d'information (Dillon, 1999), (Dyson, 2001). Cela dit le peu d'effet sur la compréhension n'est mis en évidence que lorsque l'information, même dans des mauvaises conditions de lecture, a été entièrement captée. Il nous semble pourtant que la perte d'information utile dans le cas de présentations sur des écrans de type PDA n'est pas négligeable. Ainsi notre objectif est ici de permettre l'accès à toute l'information, et en particulier à celle véhiculée par le *first glance*, en améliorant le rapport quantité d'information/espace d'affichage en même temps que l'accès aux structures visuelles complexes.

Alors que le *Comscore Networks* (2002) recense 9,9 millions d'adultes américains qui utilisent des *Personal Digital Assistant* (PDA) pour accéder au *Web*, et en particulier aux sites d'informations, la présentation de texte dans ces conditions ne remplit pas sa fonction de manière satisfaisante (Ericsson, 2001) en raison des problèmes de lisibilité dus au mauvais rapport entre la quantité d'information à présenter et l'espace d'affichage disponible.

Beaucoup de sites proposent une version « petits écrans » de leurs pages (images enlevées, version texte seulement, réduction des polices, ...). La mise en page, l'affichage et même le contenu est ainsi parfois modifié. Les navigateurs proposent également d'opérer leurs propres changements (Windows CE IE permet par exemple de changer la police). De manière générale, que se soit au niveau du site ou du navigateur, différentes techniques sont actuellement proposées pour améliorer la présentation d'information sur un petit écran. Nous proposons, à travers le schéma de la Figure 18, de synthétiser les différents principes qui les sous-tendent, le plus exhaustivement possible.

PRÉSENTATIONS POUR UNE LECTURE TRADITIONNELLE	
<p><i>Paging</i></p> <p>Séries de petites pages</p> <p>Navigation dans des résumés</p> <p>Navigation dans des vignettes</p>	<p><i>Scrolling</i></p> <p>Migration directe</p> <p>Recadrage</p> <p><i>Clipping</i></p>
PRÉSENTATIONS POUR UNE LECTURE DYNAMIQUE	
<p><i>Leading</i></p>	<p>RSVP</p>

Figure 18 : Synthèse de stratégies de présentation sur petit écran

De manière générale nous pouvons distinguer deux grandes classes de stratégies (Mills, 1987), (Mutter, 1996) :

- un premier format autorise une lecture traditionnelle du texte ; la disposition des différents éléments du texte est préservée et celui-ci peut être présenté :
 - o en divisant le texte en pages adaptées à la taille de l'écran nécessitant l'activation d'une touche pour transiter d'une page à l'autre (*paging*),
 - o ou comme une longue et/ou large page continue impliquant le défilement du texte petit à petit (*scrolling*) ;
- et un second format qui permet un accès dynamique au texte par la transformation d'une présentation spatiale en une présentation « temporelle » :
 - o en faisant défiler le texte sur une ligne horizontale (*leading* ou *Time Square Format*),
 - o ou en faisant apparaître des segments de mots du texte en série à un endroit localisé de l'écran (réduisant ainsi la nécessité de mouvements oculaires). Cette dernière technique (*Rapid Serial Visual Presentation – RSVP*) a été démontrée plus efficace que la précédente et aussi rapide que la lecture d'un livre papier (Rahman, 1999), (Öquist, 2003).

Si le second format apparaît *a priori* intéressant pour améliorer le rapport quantité d'information/ espace d'affichage, il entraîne la perte irrémédiable du *first glance* (la technique habituellement utilisée d'une barre de progression pour signifier les quantités lues et restantes semble limitée de ce point de vue). De plus, le gain en diminution de la charge cognitive espéré par la réduction des mouvements oculaires avec la technique RSVP n'est pas au rendez-vous (Öquist, 2004).

De la même manière la lecture traditionnelle de type *scrolling* consiste, au mieux à une transformation linéaire du document à présenter, et au pire à une migration directe du contenu prévu pour un écran normal vers un écran de petite taille. Cette dernière solution pose des problèmes de navigation à travers l'utilisation de deux « ascenseurs » (horizontal et vertical) et ne permet d'accéder d'un regard qu'à une petite partie de la page *Web* à la fois (Albers, 2000), (Jones, 1999). Quant aux transformations linéaires elles proposent un recadrage de la page pour éliminer la contrainte de navigation en largeur (Pocket IE¹⁵, Opera's Small Screen Rendering¹⁶) le problème est alors celui de la perte de la disposition originale des différents éléments qui peut entraîner des difficultés importantes de navigation (Hinz, 2004), (MacKay, 2004). Une amélioration de cette dernière technique, le *clipping*¹⁷, tente d'analyser la structure de la page et d'éliminer l'information dite inutile (certaines images, *frame*, ...) par des algorithmes intelligents de classification des fragments (Gomes, 2001), (Bickmore, 1997) ou la définition manuelle de priorités (Hori, 2000). L'évaluation de cette solution montre encore de sérieuses difficultés pour prédire le résultat d'une telle transformation qui entraîne un grand nombre d'erreurs (Chen, 2003).

Enfin, la solution du *paging* implique la restructuration du document en petites pages qui sont présentées :

- soit en série ; cette solution est bien adaptée à la lecture de document n'ayant que peu de mise en page et pour lesquels le lieu de la césure entre deux pages n'a pas de réelle importance. Cette technique est traditionnellement utilisée pour la lecture de livres électroniques (Microsoft Reader¹⁸) pour lesquels la lecture sur petit écran ne va différer de la lecture sur écran normal que par le nombre de pages présentées ;
- soit en conservant la vue d'ensemble du document ; cette technique provient des recherches autour de la présentation de grandes masses de données complexes sur grand écran (Spence, 2001), (Furnas, 1986) qui ont été adaptées à la présentation sur petits écrans (Björk, 1999), (Wobbrok, 2002), (Chen, 2003).

¹⁵ Pocket IE : <http://www.microsoft.com/windowsmobile/products/pocketpc/default.msp>

¹⁶ Opera's Small Screen Rendering : <http://www.opera.com/products/smartphone/smallscreen>

¹⁷ Palm Web Clipping: <http://www.palmos.com/dev/tech/webclipping/resources.html>

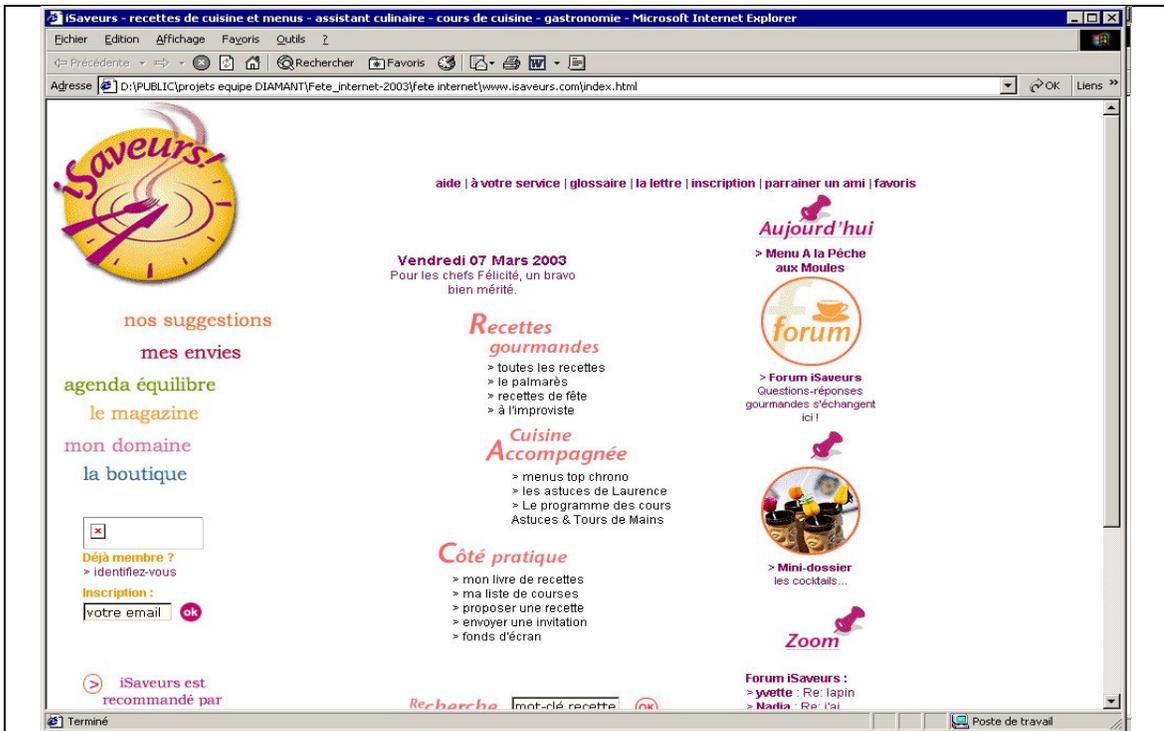
¹⁸ <http://www.microsoft.com/reader/>

L'intérêt de la solution exprimée par ce second point réside dans la conservation de la disposition des éléments qui constituent la page *Web*, mais elle implique en même temps un certain nombre de problèmes de lisibilité des informations. Ainsi, des améliorations ont pu être apportées par une restructuration de la page *web* en zones, qui sont ensuite résumées ou mises sous la forme de vignettes (Buyukkokten, 2001), (González-Castano, 2002), (Milic-Frayling, 2002). Si la technique par résumé est encore difficile à mettre en œuvre, le principe des vignettes (*thumbnails*) se rapproche quelque peu de notre notion de présentation par Image De Page.

Finalement les solutions monomodales peuvent impliquer :

- l'utilisation de stratégies de navigation plus ou moins efficace dont la pire est l'utilisation d'un système de double ascenseur (Figure 19a) ;
- la réduction scalaire de la page au prix d'une importante difficulté
 - o perceptive à déchiffrer certains éléments ainsi réduits (Figure 19b),
 - o ou linguistique pour faire sauter les verrous impliqués (résumés, classification de fragments selon leur importance).

La solution que nous proposons part du principe que les IDP permettent de se départir du contenu textuel de l'information pour ne conserver que l'architecture globale. Un intérêt est ainsi de pouvoir jouer sur la granularité de la présentation en ne fixant que les paramètres formels des structures visuelles qui paraissent importants pour le rôle dédié à l'IDP. Dans notre cas, nous pouvons ainsi procéder à la fois à la réduction scalaire de la page présentée sans pour autant réduire la taille des caractères utilisés pour exprimer l'inscription graphologique. Ceci est rendu possible par le fait de pouvoir jouer par exemple sur la largeur ou la profondeur des OT. L'accès au contenu textuel est alors permis grâce à une oralisation des OT au passage du stylet.



Exemple d'une page Web d'un site de recettes de cuisine

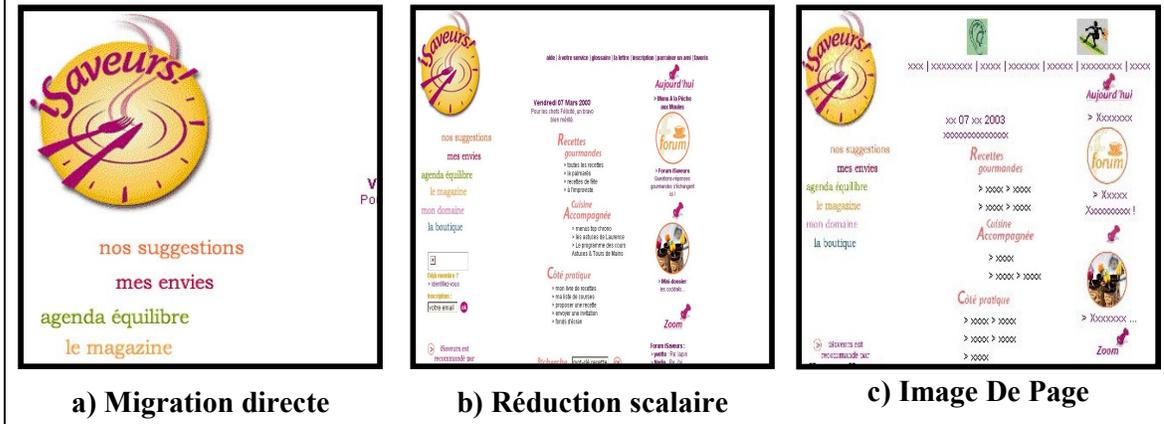


Figure 19 : Trois présentations d'une page Web

Une première estimation empirique de ce procédé a été faite à l'occasion d'une démonstration pour la fête de l'Internet à La Vilette¹⁹ : la mise en IDP d'un site de recettes de cuisine a été constituée manuellement et a été testée empiriquement dans trois situations pour lesquelles il était demandé de trouver la recette d'un plat particulier ; les deux premières situations, qui n'utilisaient pas la mise en IDP du site

¹⁹ <http://www.fete-internet.fr/> - Cité des sciences et de l'industrie, 17-23 Mars 2003 ;

(migration directe, réduction scalaire sans restructuration), semblaient moins performantes que la version qui permettait un accès oral aux informations à travers une navigation grâce aux images de page adaptées à la taille de l'écran (Figure 19c).

Cet exemple montre l'intérêt potentiel du principe des IDP pour une navigation orale dans une page *Web* sur petit écran : les différents éléments « clickables » du menu principal ne sont directement accessibles que dans cette version ; de la même manière que les différents items des énumération sont également plus faciles à appréhender.

Nous travaillons actuellement sur la restructuration automatique des pages *Web* en une série d'IDP cohérentes adaptées au support de présentation.

XI.2 IDP, navigation et non-voyance

Dans un contexte d'accès à l'information sur le web par les non voyants, les outils usuels intègrent de la synthèse de la parole à partir de texte et une plage Braille. Cet environnement pose plusieurs problèmes dont celui de la navigation textuelle intra et inter documents : ces outils ne permettent pas d'anticiper d'un premier regard la structure visuelle de la page ou de l'extrait du document affiché à l'écran ; une navigation « ciblée », plus directe, c'est-à-dire qui s'affranchisse de la présentation exhaustive et linéaire de l'ensemble des éléments de la page ou de son sommaire, s'avère ainsi limitée.

Les différentes solutions logicielles spécifiques utilisées par les non-voyants pour piloter leur systèmes de synthèse de la parole ne proposent aucune stratégie qui soit satisfaisante à tout point de vue pour aider cette population d'utilisateurs à accéder au sens véhiculé par la mise en forme des documents oralisés : les options proposées par ces logiciels ne permettent qu'une oralisation proche de notre forme discursive descriptive mais sans discernement de la pertinence des marques et donc sans reconnaissance en amont de configurations particulières intentionnelles et signifiantes. Le remède est donc souvent pire que le mal puisque la volonté d'exhaustivité dans la description de la mise en forme se fait alors au détriment d'une intelligibilité correcte. D'ailleurs, cette possibilité offerte aux utilisateurs de ce type de logiciel spécialisé n'est généralement utilisée que dans de rares cas particuliers.

Aussi, la navigation dans une page *Web* ne permet qu'un nombre réduit de stratégies. Par exemple, la recherche d'un lien hypertexte pourra se faire :

- par un parcours linéaire de toute la page jusqu'à reconnaître le lien grâce à une oralisation particulière du segment de texte ou une indication braille spécifique ;

- ou par un parcours linéaire dans une liste composée de l'ensemble des liens hypertextes de la page.

Pour la recherche dans des structures plus complexes, telles que les tableaux, seule la première possibilité est donnée à l'utilisateur. Parfois l'ébauche d'un dialogue est malgré tout envisagée pour permettre à l'utilisateur d'appliquer différentes stratégies de parcours de l'information. Par exemple, un peu à la manière des audioguides qui permettent la présentation orale d'un espace (en particulier dans les musées), un certain nombre d'informations sont fournies avant de proposer l'accès à ce type de structure : le nombre de lignes et de colonnes, le titre du tableau, ... Pourtant, l'intérêt d'une présentation tabulaire réside dans les différentes stratégies qu'elle offre pour (re)chercher de l'information : les règles de lecture de Waller (1987) énoncées dans notre mise en perspective, indiquent que le lecteur devrait pouvoir sélectionner un titre dans chaque axe de la matrice et regarder la cellule à l'intersection, ou vice et versa ; ou bien comparer toutes les cellules d'une ligne ou d'une colonne particulière ; ou encore comparer les contenus de toute les lignes et les colonnes.

Nous constatons que la recherche scientifique ne s'est que très peu intéressée à cette problématique : un non voyant peut-il avoir accès à une forme de sensation visuelle de lecture si le texte ne lui est pas fourni par les yeux, mais de manière tactile ?

Les prémisses de ce type de recherche sont toutefois observables pour réaliser la transformation d'une image visuelle en une image tactile (Bach-y-Rita, 1969). Le type de dispositif mis en œuvre consiste en une caméra vidéo reliée à une plaque de stimulation tactile hérissée de picots mobiles ou portant des électrodes qui envoient des impulsions électriques sur une partie du corps (bas du dos, index, abdomen, langue, ...). Les expérimentations indiquent que le système n'est efficace que si le sujet participe activement à l'exploration de l'environnement afin d'établir la relation entre ses actions et ses perceptions (*ibid.*). Des résultats prometteurs ont montré qu'avec un entraînement, il est possible pour un aveugle de frapper une balle roulant vers lui avec une batte de manière très précise, avec pour information la seule information tactile des mouvements correspondant à la trajectoire de la balle. Sur ce même principe, un seul système commercialisé permet de parcourir un texte grâce à un stylet muni d'une caméra afin de l'enregistrer et le restituer sous la forme de vibrations tactiles sur un doigt de la main libre. Au niveau de la recherche, nous pouvons principalement nous référer aux travaux de Lenay (2003) qui ont débouché sur un prototype d'interfaces baptisées « stylet tactile » qui, comme le dispositif décrit précédemment, envoient sur la main libre des stimulations dépendantes des formes rencontrées par le stylet.

Ce type d'interface tactile permettrait ainsi la perception des graphiques, des icônes et des formes simples bi-dimensionnelles. Dans le cadre de notre étude, la finesse des éléments à percevoir nous dirige avec le même esprit vers une autre solution, multimodale, qui est d'exploiter la notion d'Image De Page pour permettre une présentation en relief de l'architecture du texte sur une page A4 de type Braille, à picots mobiles. L'objectif est de permettre l'accès direct à l'information d'une page *Web* par une utilisation active tactile de la page A4 afin, à la fois de percevoir les différentes structures de la page, et d'agir dessus pour piloter les outils classiques de synthèse et de Braille.

Nous travaillons actuellement au prototypage d'une présentation tactilo-orale de l'information textuelle de la page *Web* prenant en compte les capacités sensorielles théoriques des usagers et à évaluer expérimentalement l'utilisabilité de ce type de dispositif pour les non-voyants.

« *Élever l'homme, c'est faire naître en lui des perspectives d'après lesquelles s'organiseront ses joies de demain.* »
Anton Makarenko, *Le chemin de la vie*

Point sur notre contribution et perspectives à moyen terme

Le travail accompli partait de l'idée que l'étude de la transmodalité permettrait de fournir un cadre argumenté pour une multimodalité utile et efficace, à partir de l'observation des caractéristiques inhérentes à la modalité source que nous ne pouvons, malgré nos efforts, transposer vers la modalité cible. L'étude s'est consacrée à cette approche à travers la problématique de la transposition orale automatique de textes conçus par le rédacteur pour une lecture silencieuse. Bien que la trame principale de ce travail soit du domaine de l'informatique, une étude à caractère pluridisciplinaire s'est avérée incontournable pour essayer d'obtenir une équivalence intermodale la plus aboutie possible, que se soit (1) en terme de contenu informationnel, (2) en terme de traitement cognitif de l'information oralisée, ou (3) en terme de conception, de représentation et d'automatisation de stratégies de reformulation des textes.

Enfin, cette démarche nous a conduit à mettre en évidence une limite de la transmodalité écrit→oral, le *first glance*, et à proposer deux prototypes exploitant un nouveau concept, les images de page, pour la réalisation d'interfaces multimodales originales en terme de présentation de l'information et d'interaction homme/machine. Ainsi les différents apports scientifiques que propose cette recherche se situent principalement dans le domaine de l'ingénierie de la représentation des documents et de l'IHM mais ouvrent deux ponts vers les champs disciplinaires de la linguistique du texte et de la psycholinguistique.

Nous avons tout d'abord justifié, à travers le point de vue de typographes, de psychologues, de psycholinguistes et de linguistes ainsi qu'à partir de notre propre réflexion autour de la projection des concepts issus de la linguistique classique sur la typographie, l'intérêt de ces disciplines pour participer à répondre à notre problématique informatique.

Nous avons pu ainsi proposer une taxonomie adaptée à notre cadre d'étude. En effet, la notion d'*inscription graphologique du texte* a été adoptée et raffinée selon cinq types : *imposée, stylistique, démarcative, distinctive thématique* et *distinctive scalaire*. L'intérêt de ce découpage est de fournir une méthode pour réduire le champ d'investigation ouvert lors de la recherche de modèles prosodiques adaptés à la transposition orale d'objets textuels particuliers. Par exemple, certaines marques, de type distinctif thématique, nous ont dirigés vers les modèles de l'intonation pragmatique du français, pour l'oralisation d'une partie de l'inscription graphologique des structures énumératives.

L'approche proposée par le Modèle d'Architecture Textuelle, pose un principe d'équivalence fonctionnelle entre différents type de marques de Mise en Forme Matérielle (typographique, dispositionnelle et lexico-syntaxique) dont nous nous inspirons pour faire l'hypothèse d'une équivalence fonctionnelle entre des marques lexico-syntaxiques et des configurations prosodiques. Une analyse linguistique des objets textuels, mis en évidence par cette Mise en Forme Matérielle, permet d'envisager l'opérationnalisation d'un premier jeu de stratégies de reformulation discursive de l'inscription graphologique démarcative et distinctive.

Cette première partie nous a ainsi montré que, dans notre perspective d'automatisation de la transposition à l'oral de textes écrits pour être lus silencieusement, nous devons nous appuyer sur :

- une architecture différente des systèmes automatiques actuels puisque une unité textuelle de traitement telle que la phrase ponctuée est insuffisante pour le traitement de l'inscription graphologique ; nous proposons une architecture dont l'unité textuelle de traitement serait l'objet textuel ; celui-ci subirait une série de transformations à travers différents modules (*segmentation, extraction, reformulation*) afin de fournir une version du texte dans une représentation interprétable à peu de coût par les logiciels de synthèse actuel ;
- des stratégies de reformulation reposant sur des transformations successives du texte ; nous proposons un Modèle d'Oralisation par Reformulation des Textes Écrits pour être Lus Silencieusement qui a pour vocation de donner une vue

exhaustive des stratégies envisageables et des relations qu'elles entretiennent entre elles ; ce modèle permet d'envisager la mise en œuvre du module de reformulation de notre architecture en fonction du type d'objet textuel susceptible d'être pris en compte et du rapport que l'on s'impose entre d'une part la complexité de mise en œuvre et d'autre part le naturel du résultat ;

- des typologies d'objets textuels pour définir quel type d'inscription graphologique les constitue et quels paramètres formels prendre en compte lors de l'oralisation.

Sur ces bases, nous avons opérationnalisé le module de reformulation pour un objet textuel particulièrement complexe : les structures énumératives. Se dégage ainsi une méthodologie générale, intéressante dans le cadre de l'ingénierie informatique, puisqu'elle nous permet de proposer :

- une formalisation de cet objet textuel sous la forme d'une DTD que nous avons entièrement écrite ;
- des recommandations pour améliorer le langage de représentation SSML, chargé de représenter les aspects prosodiques d'un texte à oraliser, et ainsi obtenir une formalisation de la version prosodique sous la forme d'une DTD de type SSML ; cette dernière version de l'objet textuel est directement exploitable par un logiciel de synthèse de la parole à partir de texte à condition que nos spécifications prosodiques pour l'interprétation du marquage soit respecté (ces spécifications ont été implémentées pour la voix de synthèse Proverbe d'Élan Speech) ;
- l'utilisation de la norme XSLT pour dériver les différentes versions de l'objet textuel selon la stratégie utilisée (discursive ou prosodique).

Enfin, les limites de la transmodalité écrit→oral nous ont finalement conduits à argumenter en faveur d'une potentielle utilité de la multimodalité pour paralléliser le traitement de la lecture du texte afin de répondre aux difficultés d'accessibilité dues à certaines contraintes situationnelles telle que l'absence ou la dégradation de la modalité visuelle.

La mesure de l'utilisabilité de la multimodalité pour répondre à cette nécessité est envisagée à travers l'étude de deux prototypes IHM impliquant l'utilisation du concept d'Images De Page (dont la plausibilité cognitive a été démontrée) pour proposer des interfaces multimodales reposant sur une version orale du contenu alphabétique du texte, accompagnée d'un accès à son architecture, visuel dans le premier prototype et tactile dans le second.

Ainsi, quatre objectifs dans la continuité de nos travaux à court, moyen et plus long terme, sous-tendent nos perspectives de recherche.

Ces travaux sont centrés sur (1) la modélisation, le prototypage, la mise en œuvre et l'évaluation de systèmes qui exploitent la notion d'Image De Page ; (2) la constitution/analyse linguistique de corpus de documents écrits pour l'étude de configurations morpho-dispositionnelles des textes. Ces analyses permettront de formaliser des connaissances pour développer un analyseur architectural des documents en vue de leur génération dans des modalités non strictement visuelles (orale, tactile ou multimodale).

(1) IHM : vers une présentation adaptable

Le travail accompli permet d'envisager de prendre en compte des situations handicapantes pour la présentation de documents électroniques écrits pour être lus silencieusement. L'adaptation de la présentation pourra être envisagée à plusieurs niveaux :

Au niveau du choix de la modalité de présentation

Dès lors que la modalité visuelle n'est pas disponible, la solution orale présentée pourra être envisagée. La non disponibilité de la modalité visuelle n'est pas « réservée » aux seuls non-voyants. En effet d'autres handicaps, comme les agnosies visuelles et les alexies, nécessitent de surmonter la même contrainte pour accéder aux documents écrits ; ce handicap peut également être provoqué pour chacun de nous par une situation d'interaction sans support visuel (serveur téléphonique) ou lorsque cette modalité est utilisée pour une autre activité (conduite de véhicule).

Au niveau des stratégies d'oralisation utilisées

L'accès aux documents peut être encore compliqué par des facteurs nécessitant une adaptation des stratégies d'oralisation en fonction des caractéristiques spécifiques à une population d'utilisateurs ; c'est, nous semble-t-il, particulièrement le cas pour les personnes âgées en raison de potentielles difficultés perceptives de la prosodie et/ou de déficiences mnésiques (Maurel, 2003e). Comme en attestent les résultats partiels du premier protocole expérimental présenté, la forme discursive interprétative et prosodique en contexte semblent être tendanciellement meilleures que les deux autres en terme de mémorisation selon les critères que nous choisissons pour évaluer leur apport. La suite de ce protocole devrait, nous l'espérons, nourrir l'hypothèse de supériorité de cette forme de manière significative d'un point de vue statistique pour les personnes âgées.

Au niveau de la combinaison des modalités

Le handicap physique de malvoyance et celui, situationnel, d'accès à l'information sur petit écran (e.g. : PDA) présentent un avantage : bien que limitée, la modalité visuelle peut malgré tout être exploitée grâce au concept d'Images De Page dont nous avons démontré la plausibilité cognitive.

De plus, dans le cas des personnes âgées, un handicap supplémentaire de malvoyance est fréquent mais d'une grande variabilité d'un usager à l'autre. Nous serons alors à même d'utiliser, lorsque la situation le permet, la notion de granularité des IDP (la structure informationnelle peut être représentée plus ou moins finement en fonction de la force du handicap visuel) en complément de la présentation orale.

Il est envisageable de proposer ce moyen d'interaction pour une navigation orale sur les documents électroniques par simple parcours de l'IDP. L'exploitation de ce procédé bimodal est actuellement en cours d'évaluation du point de vue informatique et psycholinguistique. Il s'agit de concevoir et d'évaluer des outils qui exploitent la capacité des IDP à « multimodaliser » de manière non redondante le document, et leur souplesse pour « filtrer » la structure visuelle des textes avec une plus ou moins grande granularité. La possibilité d'utiliser ce concept pour concevoir des nouvelles interfaces est conditionnée par la résolution de nombreux problèmes sémiologiques de formalisation, des problèmes computationnels de construction d'une représentation et des problèmes de traitement cognitif nécessitant l'évaluation du pouvoir d'expression de ce mode de présentation.

En plus des prototypes présentés, l'IDP permettrait de nouvelles solutions pour la visualisation interactive dans les grandes bases documentaires (Jacquemin, 2004) : dans le cas de la recherche d'information par mots-clefs sur le *Web*, la présentation simultanée des IDP des documents proposés en résultats des requêtes, avec une mise en saillance des mots-clefs, favoriserait l'évaluation rapide de la pertinence des réponses par la prise en compte des mises en saillance selon leur place dans le document (e.g. : titre versus note de bas de page) et leur nombre (e.g. : une seule versus de nombreuses apparitions).

(2) Vers un système de « transmodalisation » de l'architecture textuelle

Bien qu'il soit nécessaire de continuer le travail présenté « en profondeur » sur les objets textuels déjà considérés, nous travaillerons également sur une extension « en largeur ». Dans une première étape la constitution d'un recueil de documents écrits pour

être lus silencieusement nous permettra de proposer une liste, la plus exhaustive possible, des OT qui sont susceptibles de les composer. Un travail d'annotation des procédés lexico-syntaxiques et morpho-dispositionnels utilisés pour « architecturer » chacun de ces OT permettra de faire émerger leurs paramètres formels et les relations « métaphrastiques » qui existent entre eux.

Une seconde étape de modélisation prosodique s'attachera à déterminer les marques prosodiques qui ont la même fonction que les métaphrases de la version discursive. Nous appliquerons la même démarche pluri-fondée que pour les OT précédemment étudiés :

- écoute qualitative (attentive et critique) d'un recueil de discours oraux ;
- introspection soumise à discussion avec des spécialistes du domaine ;
- utilisation de cette immersion pour imaginer une projection des règles dictées par les modèles prosodiques du français sur la relation entre des marques lexico-syntaxiques d'un OT et sa version prosodique.

Les deux étapes précédentes permettront de procéder à une analyse critique des normes de représentation émis par le World Wide Web Consortium (W3C), que se soit au niveau des standards de formatage du texte (e.g. : XHTML, Cascading Style Sheet) ou au niveau de celles établies pour représenter les spécificités prosodiques de son oralisation (e.g. : Speech Synthesis Markup Language, Aural Cascading Style Sheet). Le cas échéant, nous proposerons des nouvelles fonctionnalités, nécessaires à une représentation formelle des OT qui permette d'envisager l'automatisation de l'étape de reformulation.

L'objectif d'une analyse architecturale est l'obtention, par une analyse automatique, du découpage des documents électroniques en OT dans leur représentation formelle définie par les étapes précédentes. La réalisation d'un tel analyseur architectural implique l'automatisation d'une analyse qui tienne compte de l'inscription graphologique des textes et de certains éléments syntaxiques et sémantiques spécifiques. Cette opérationnalisation pourra parfois être facilitée en s'appuyant partiellement sur les représentations existantes du document original. Nous pourrions également considérer dans cet objectif les nombreux travaux du domaine (Mourad, 2001), (Friburger, 2004), (Crucianu, 2002).

(3) Cognition

Celle-ci est au cœur de chacune des problématiques rapidement développées ci-dessus à travers l'**étude** de l'impact sur le traitement cognitif de différentes stratégies

d'oralisation et de présentation de l'information ainsi que l'**évaluation** de prototypes en vue de leur mise en œuvre.

La plupart des aspects de ces problématiques, peu abordés dans la littérature de la psycholinguistique ou de la psychologie de la cognition, conduiront certainement à développer des **méthodologies expérimentales originales**.

BIBLIOGRAPHIE

- ALBERS (2000)** Albers, M., Kim, L., User Web Characteristics Using Palm Handhelds for Information Retrieval, Dans actes de *IPCC/SIGDOC Technology and Teamwork*. Cambridge, pp. 125-135, 2000.
- ALESSANDRO (2001)** D'Alessandro, C., Tzoukermann, E., *Traitement automatique des langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, Hermes Science Publications, Paris, 2001.
- ARABYAN (2001)** Arabyan, M., Du nouveau sur le mythe des origines de la lecture silencieuse, dans Monique de Mattia et André Joly édés., *De la syntaxe à la narratologie énonciative*, Gap-Paris, Ophrys.
- BACCINO (1999)** Baccino, T., Schmid, S., Variations temporelles et oculométriques dans la lecture de consignes de jeux, in Virbel, J., Cellier, J.-M. & Nespoulous, J.-L. (eds), *Cognition, Discours procédural, Action*, Prescot, Toulouse, pp. 230-247, 1999.
- BACH-Y-RITA (1969)** Bach-Y-Rita, P., Collins, C.C., W., Saunders, F.A., White, B, L. .Scadden, Visual substitution by tactile image projection, *Nature*, 221, pp. 963-964, 1969.
- BADDELEY (1986)** Baddeley, A.D., *Working Memory*, Oxford University Press, New-York, 1986.
- BECHET (2001)** Béchet, P., Un système complet de phonétisation des textes, *Traitement automatique des langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, pp. 47-67, 2001.
- BELLIK (1995)** Bellick, Y., *Interfaces multimodales : concepts, modèles et architextures*, Thèse de doctorat, Université Paris-Sud, Orsay.
- BERLINER (1920)** Berliner, A., Atmosphärenwert von Drucktypen, *Ztschf. angewandte Psychol.*, 17, 165-172, 1950.
- BERNHART (1985)** Bernhardt SA (1985) Text structure and graphic design, in JD Benson & WS Greaves (Eds), *Systemic perspectives on discourse, volume 2*, Norwood, NJ: Ablex, 1985.

- BERNSEN (2002)** Bernsen, N.O., Multimodality in Language and Speech Systems – from theory to design support tool, in Björn Granström, David House, et Inger Karlsson (Eds.), *Multimodality in Language and Speech Systems*, collection “Text, Speech and Language Technology”, pp.93-148 , 2002.
- BICKMORE (1997)** Bickmore, T., Schilit, B., Digestor: Device-Independent Acces to the WWW, Dans actes de 6th *International WWW Conference*, 1997.
- BJÖRK (1999)** Björk, S., Holmquist, L.E., Redström, J., Bretan, I., Danielsson, R., Karlgren, J., Franzén, K., West: A Web Browser for Small Terminals, Dans actes de *ACM Synposium on User Interface Software and Technology*, Asheville, NC, pp. 187-196, 1999.
- BLOOMFIELD (1935)** Bloomfield, L., *Language*, revised edition, London: George Alen & Unwin, 1935.
- BOULA DE MAREUIL (2001)** Boula de Mareuil, P., Célérier, P., Cesses, T., Fabre, S., Jobin, C., Le Meur, P.-Y., Obadia, D., Soulage, B., Toën, J., Elan Text-To-Speech : un système multilingue de synthèse de la parole à partir du texte, *Traitement automatique des langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, pp. 223-252, 2001.
- BOURAOUI (2000)** Bouraoui, J.-L., *Les structures énumératives : caractérisation linguistique et reconnaissance automatique*. Mémoire de DEA, Université de Toulouse Le Mirail, France.
- BOUVERESSE (1976)** Bouveresse, R., Le nombre d'or, dans *Psychologie*, 77, 1976.
- BUCKINGAM (1931)** Buckingham, B., New data on the typography of textbooks, *Yearbook of the National Society for the Study of Education*, 30, 93–125, 1931.
- BUSH (2000)** Bush, C., Analyse des déclencheurs des énumérations d’entités nommées sur le web, *Rapport technique*, 2000-02, LIMSI.
- BUXTON (1983)** Buxton, W., Lexical and pragmatic considerations of input structures, *Computer Graphics* 17,1, pp. 31-37, 1983.

- BUYUKKOKTEN (2001)** Buyukkokten, O., Garcia-Molina, H., Paepcke, A., Seeing the Whole in Parts: Text Summarization for Web Browsing on Handheld Devices, 10th *International WWW Conference*, pp. 652-662, Hong Kong, 2001.
- CAËLEN (2000)** Caelen, J., 10 ans de recherches ... en multimodalité(s), Colloque sur les interfaces multimodales, <http://www.irit.fr/M3/CM10ans/Talks/Caelen.pdf>, 2000.
- CARBONELL (1994)** Carbonell, N., Valot, Cl., Mignot, Ch., Dauchy, P., Étude empirique : usage du geste et de la parole en situation de communication homme-machine, *ERGO'IA*, pp. 96-107, 1994.
- CARD (1990)** Card, S., Mackinlay, J., Robertson, G., The Design Space of Input Devices, Dans actes de *CHI'90*, pp. 117-124, 1990.
- CATACH (1994),** Catach, N., La ponctuation, *Que sais-je ?*, presses universitaires de France, 1994.
- CHAROLLES (1978)** Charolles, M., Introduction aux problèmes de la cohérence des textes, *Langue française*, 38, p. 7-42.
- CHAROLLES (1994)** Charolle, M. Cohésion, cohérence et pertinence du discours, *Travaux de linguistique* 29, Bruxelles & Louvain-la-Neuve, Duculot, pp. 125-151, 1994.
- CHAROLLES (1997)** Charolle, M., Les cadres du discours. Univers, champs, domaines et espaces, *Cahiers de Recherche Linguistique* 6, Université Nancy 2, UFR des Sciences du Langage & LanDisCo-CNRS, 1997.
- CHAYTOR (1945)** Chaytor, H., *From script to print: an introduction to medieval literature*, Cambridge University Press, 1945.
- CHEN (2003)** Chen, Y., Ma, W., Zhang, H., Adapting Content to Mobile Devices: Detecting Web Page Structure for Adaptative Viewing on Small Form Factor Devices, Dans actes de 12th *International WWW Conference*, pp. 225-233, Budapest, 2003.
- CHERBONNEL (1996)** Cherbonnel, B., Un module de pré-traitement multilingue de documents intégré au système de synthèse de parole à partir du texte du CENT,

Document technique DT/267/LAA/TSS/RCP, CENT, Lannion, France, 1996.

- CHOMSKY (1995)** Chomsky, N., Lasnik, H., The Theory of Principles and Parameters, in Chomsky, N., *The Minimalist Program*, MIT Press, Cambridge, pp. 13-127, 1995.
- COLES (1975)** Coles, P., Foster, J., Typographic cueing as an aid to learning from typewritten text, *Programmed Learning & Educational Technology*, 12, pp.102-108, 1975.
- COMSCORE NETWORKS (2002)** Ten Million Internet Users go Online Via a Cell Phone or PDA (août 2002) from <http://www.comscore.com/press/release.asp?id=87>
- CONTENT (1990)** Content, A., Mousty, P., Radeau, M., BRULEX Une base de données lexicales informatisée pour le français écrit et parlé, *L'Année Psychologique*, 90, pp. 551-566, 1990.
- COOK-GUMPERZ (1981)** Cook-Gumperz, J., Gumperz, J., From oral to written culture: the transition to literacy, in MF Whiteman (ed), *Writing: The nature, development and teaching of written communication. Volume 1: Variation in writing*, Hillsdale, NJ: Lawrence Erlbaum Associates, 89-109, 1981.
- COTTO (1992)** Cotto, D., *Traitement automatique des textes en vue de la synthèse vocale*, Thèse de doctorat - Université Paul Sabatier, 1992.
- COULTHARD (1985)** Coulthard, M., *An introduction to discourse analysis*, 2nd edition, London, Longman, 1985.
- COUTAZ (1995)** Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., Young, R., Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. Dans actes de *Interact'95*, pp. 115-120, 1995.
- CROSNIER (2003)** Crosnier, J., Les deux voies de communication des émotions, in Colletta, J.-M. & Tcherkassof, A. (eds.), *Perspectives actuelles sur les émotions. Cognition, langage et développement*, Hayen, Mardaga, 2003.
- CRUCIANU (2002)** Crucianu, M., Faure, C., Tappa, F., Venet, M., Vincent, N., Détection et usage des composantes typographiques dans des pages de documents, dans

- Actes de la *Conf. Fédérative sur le Document, CIFED'02*, pp. 251-260, 2002.
- CRYSTAL (1969)** Crystal, D., Davy, D., *Investigating English style*, London, Longman, 1969.
- CRYSTAL (1980)** Crystal, D., *A first dictionary of linguistics and phonetics*, London, Deutsch, 1980.
- DAGONNET (1998)** Dagonnet, F., *Les dossiers de l'audiovisuel*, 1998.
- DAUSSE (2004)** Dausse, F., NFCE et schémas de communication de l'écrit et de l'oral. *Journée d'Étude de l'ATALA – Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.)*, Paris, 2004.
- DERRIDA (1967)** Derrida, J., *De la grammatologie*, les éditions de minuit – collection critique, Paris, 1967.
- DI CRISTO (2001)** Di Cristo, A. Di Cristo, P., Syntaxe : une approche métrique-autosegmentale de la prosodie, *Traitement automatique des langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, pp. 69-114, 2001.
- DILLON (1999)** Dillon, A., Richardson, J., McKnight, C., The Effect of Display Size and Text Splitting on Reading lengthy Text from the Screen, *Behavior and Information Technology*, pp. 215-227, 1999.
- DRILLON (1991)** Drillon, J., *traité de la ponctuation française*, Gallimard, 1991.
- DUCROT (1995)** Ducrot, O., Schaeffer, J.-M., *Nouveau dictionnaire encyclopédique des sciences du langage*, Édition du seuil, 1995.
- DUPLAN (1986)** Duplan, P., Jauneau, R., *Maquette et mise en page*, Éditions de l'Usine Nouvelle, Paris, 1986.
- DUTOIT (1997)** Dutoit T., *An Introduction to Text-To-Speech Synthesis*, Kluwer Academic Publishers, 1997.
- DYSON (2001)** Dyson, M., Haselgrove, M., The influence of Reading Speed and Line Length on the Effectiveness of Reading from Screen, *International Journal Human-Computer Studies*, pp. 585-612, 2001.
- ECO (1981)** Eco, U., *The role of the reader*, London, Hutchinson, 1981.

- EDWARDS (2002)** Edwards, A.D.N., Multimodal Interaction and People with Disabilities, in Björn Granström, David House, et Inger Karlsson (Eds.), *Multimodality in Language and Speech Systems*, collection “Text, Speech and Language Technology”, pp.73-92 , 2002.
- ELLIS (1938)** ELLIS, W. D., *A Source Book of Gestalt Psychology*, Réédition dans The Gestalt Journal Press, New-York, 1997.
- ERICSSON (2001)** Ericsson, T., Chincholle, D., Goldstein, M., Both the device and the service influence WAP usability, Actes de *IHM-HCI2001, usability in practice*, J. Vanderdonckt, A. Blandford and A. Derycke (Eds.), Lille, France, pp. 79-85, 2001.
- FOLEY (1984)** Foley, J. D., Wallace, V. L., Chan, P., The Human Factors of Graphic Interaction Techniques, *IEEE Computer Graphics and Application 4.11*, pp. 13-48, (1984).
- FRIBURGER (2004)** Friburger, N., Maurel, D., Finite-state transducer cascades to extract named entities in texts, *Theor. Comput. Sci. 313(1)*, pp. 93-104, 2004.
- FROHLICH (1991)** Frohlich, The design Space of Interfaces, Multimedia Systems, Interaction and Applications, Dans actes de *1st Eurographics Workshop*, pp. 53-69, 1991.
- FURNAS (1986)** Furnas, George W., Generalized Fisheye Views, Dans actes de *Conference on Human Factors in Computing Systems*, pp. 16-23, Boston, MA, 1986.
- GADET (1996)** Gadet, F., Niveaux de langue et variation intrinsèque, Actes du Colloque de Paris III, 1994, *Palimpsestes 10*, 17-40, 1996.
- GAONAC'H (1990)** Gaonach, D., Passerault, J. M., Importance markers and Treatment of Textual Elements: Immediate and Delayed Effects, *European Journal of Psychology of Education, 5(1)*, pp. 55-64, 1990.
- GARCIA-DEBANC (2001)** Garcia-Debanc, C., Grandaty, M., Incidence des variations de la mise en forme textuelle sur la compréhension et la mémorisation de textes procéduraux (règles de jeux) par des enfants de 8 à 12 ans, *Langages, 141*, 2001.

- GARDES-TAMINE (1998)** Gardes-Tamine, J., Pellizza, M.-A., *La construction du texte : De la grammaire au style*, Armand Collin – Cursus, Paris, 1998.
- GARDES-TAMINE (2004)** Gardes Tamine, J., *Pour une grammaire de l'écrit*, Belin, 2004.
- GELB (1963)** Gelb, I., *A study of writing*, 2nd edition, Chicago, University of Chicago Press, 1963.
- GENETTE (1987)** Genette, G., *Seuils*, Editions du Seuil, 1987.
- GESTNER (1959)** Gerstner, K., *Designing programmes*, Zurich: Arthur Niggli, 1959.
- GESTNER (1974)** Gerstner, K., *Compendium for literates: a system of writing*, Cambridge, Mass, MIT Press, 1974.
- GIBSON (1977)** Gibson, J. J., The theory of affordances, in Shaw, R. E., Bransford, J. (eds.), *Perceiving, Acting, and Knowing*, Lawrence Erlbaum Associates, Hillsdale, 1977.
- GLYNN (1979)** Glynn, S. M., Divesta, F. J., Control of Prose Processing via Instructional and Typographical Cues, *Journal of Educational Psychology*, 71, pp. 595-603, 1979.
- GONZÁLEZ-CASTANO (2002)** Gonzáles-Castano, F., Anido Riffón, L., Bosta Montene-gro, E., A., New Transcoding Technique for PDA Browsers, Based on Content Hierarchy, Dans actes de la 4th *International Symposium on Mobile HCI*, 2002.
- GOODMAN (1976)** Goodman, N., *Languages of art: an approach to a theory of symbols*, 2nd edition, Indianapolis, Indiana, Hackett, 1976.
- GOODMAN (1990)** Goodman, N., *Langages de l'art* (traduction : J. Morizot), édition Jacqueline Chambon, 1990.
- GOODY (1979)** Goody, J., *La raison graphique*, Les éditions de minuit, Paris, 1979.
- GUIMIER DE NEEF (2004)** Guimier de Neef, E., Véronis, J., 1 pw1 sr la kestion ;-), *Journée d'Étude de l'ATALA – Le traitement automatique des nouvelles formes de communication écrite (e-mails, forums, chats, SMS, etc.)*, Paris, 2004.

- GULLICK (1986)** Gulick, M., Old books for new, *Fine Print*, 12, 204-208, 1986.
- GUMPERZ (1982)** Gumperz, J., *Discourse strategies*, Cambridge University Press, 1982.
- HALLIDAY (1976)** Halliday, M.A.K., Hasan, R., *Cohesion in English*, London, Longman, 1976.
- HARRIS (1971)** Harris, Z., *Structures Mathématiques du Langage*, Dunod, 1971.
- HARRIS (1990)** Harris, Z., La genèse de l'analyse des transformations et de la métalangue, *Langages*, (99), pp. 9-20, Larousse, 1990.
- HARRIS (1993)** Harris, R., *La sémiologie de l'écriture*, Paris, Editions du CNRS, 1993.
- HARTLEY (1979)** Hartley, J., Trueman, M., Burnhil, P., The role of spatial and typographic cues in the layout of journal references, *Applied Ergonomics*, 10, 165-169, 1979.
- HARTLEY (1981)** Hartley, J., Burnhil, P., The spatial arrangement of text, special issue of *Visible Language*, vol 15, no 1, 1981.
- HARTMAN (1972)** Hartmann, R.R.K., Stork, F., *A dictionary of language and linguistics*, London, Applied Science Publishers.
- HERSHBERGER (1965)** Hershberger, A., Terry, F. (1965) 'Typographical cuing in conventional and programmed texts', *Journal of Applied Psychology*, 40, 55-60, 1965.
- HINZ (2004)** Hinz, M., Fiala, Z., Wehner, F., Personalization-Based Optimization of Web Interfaces for Mobile Devices, *Mobile Human-Computer Interaction – MobileHCI 2004*, pp. 204-215, Springer-Verlag, Berlin, 2004.
- HO-DAC (2004)** Ho-Dac, L.-M., Jacques, M.-P., Rebeyrolle, J., Sur la fonction discursive des titres, in S. Porhiel & D. Klinger (eds), *Regards croisés sur l'unité texte*, Nicosie – Chypre, 2004.
- HOEK (1981)** Hoek, L. H., *La marque du titre*, La Haye, Mouton, 1981.
- HOEY (1983)** Hoey, M., *On the surface of discourse*, London, George, 1983.

- HORI (2000)** Hori, M., Kondoh, G., Ono, K., Hirose, S., Singhai, S., Annotation-based Web Content Transcoding, Dans actes de 9th *International WWW Conference*, Amsterdam, 2000.
- JACQUEMIN (2004)** Jacquemin C., Folch, H., Garcia, K., Nugier, S., Visualisation interactive d'espaces documentaires, *Revue I3*, (soumis).
- JAVAL (1878)** Javal, E., Hygiène de la lecture, *Buletin de la Société de Médecine Publique*, p. 569, 1878.
- JEWETT (1981)** Jewett, L., Multi-level writing in theory and practice, *Visible Language*, 15, 32-40, 1981.
- JONES (1999)** Jones, M., Marsden, G., Mohd-Nasir, N., Boone, K., Buchanan, G., Improving Web Interactions on Small Displays, Dans actes de 8th *International WWW Conference*, Toronto, 1999.
- KAMBA (1996)** Kamba, T., Elson, S.A., Harpold, T., Stamper, T., Sukaviriya, P.N., Using small screen space more efficiently, *Proceedings of CHI'96*, 383-389, 1996..
- KINTSCH (1978)** Kintsch, W., Van Dijk, T. A., Toward a Model of Text Comprehension and Production, *Psychological Review*, 85(5), pp. 363-394, 1978.
- KINTSCH (1988)** Kintsch, W., The Role of Knowledge in Discourse Comprehension: a Construction-Integration Model, *Psychological Review*, 95(2), pp. 163-182, 1988.
- KINTSCH (1993)** Kintsch, W., Text Comprehension, Memory and Learning, *American Psychologist*, 49(4), pp. 294-303, 1994.
- KINTSCH (1999)** Kintsch, W., Patel, V., Ericsson, K. A. The role of Long-term working memory in text comprehension, *Psychologia*, 42, pp. 186-198, 1999.
- KOSTELANETZ (1970)** Kostelanetz, R., *Moholy-Nagy*, London: Alen Lane/The Penguin Press, 1970.
- LAUFER (1979)** Laufer, R., Guillemets et marques de discours direct, in *La ponctuation*, Recherches historiques et actuelles, Actes de la table ronde internationale CNRS, Paris, Publications CNRS-HESO, pp. 235-251, Paris, 1979.
- LEGROS (1922)** Legros, A., *A note on the legibility of printed matter*, London, HMSO, 1922.

- LEMARIÉ (2004)** Lemarié, J., Eyrolle, H., Cellier, J.-M., Visual signals in text comprehension: how to restore them when oralizing a text via a speech synthesis?, *Computers in Human Behavior*, soumis.
- LENAY (2003)** Lenay, C., Prothèses perceptives et constitution spatiale du sens , *Noir sur Blanc*, n° 7 , pp. 7 et 12, 2003.
- LOMAN (1983)** Loman, N. L., Mayer, R. E., Signaling Techniques that increase the Understandability of Expository Prose, *Journal of Educational Psychology*, 75, pp. 402-412, 1983.
- LONGRACE (1979)** Longrace, R.E., the paragraph as a grammatical unit, in *Syntax and Semantics, Discourse and syntax*, Givon, Talamy , vol **12**, pages 115-134, 1979.
- LORCH (1986)** Lorch, R. F., Chen, A. H., Effects of Number Signals on Reading and Recall, *Journal of Educational Psychology*, 78(4), pp. 263-270, 1986.
- LORCH (1989)** Lorch, R.F., Text Signaling Devices and their Effects on Reading and Memory Processes, *Educational Psychology Review*, 1, pp. 209-234, 1989.
- LORCH (1995)** Lorch, R.F., Lorch, E.P., Klusewitz, M. A., Effects of typographical Cues on Reading and Recall of Text, *Contemporary Educational Psychology*, 20, pp. 51-64, 1995.
- LORENCEAU (1980)** Lorenceau, A., La ponctuation au XIXe siècle, *Langue Française*, 45: 50-9, 1980.
- LUC (2000)** Luc, C., *Représentation et composition des structures visuelles et rhétoriques du texte*, Thèse de doctorat - Université Paul Sabatier, 2000.
- LUC (2001)** Luc, C., Mojahid, M., Virbel, J., Système notationnel de l'architecture textuelle par image de page, Dans Mojahid M., Virbel J. (eds.), *Document Electronique, - Méthodes, démarches et techniques cognitives*, Europia, pp. 233-245, 2001.
- LYONS (1977)** Lyons, J., *Semantics*, 2 volumes, Cambridge University Press, 1977.
- MACKAY (2004)** MacKay, B., Watters, C., Duffy, J., Web Page Transformation When Switching Devices, *Mobile*

Human-Computer Interaction – MobileHCI 2004, pp. 228-239, Springer-Verlag, Berlin, 2004.

- MACLUHAN (1962)** McLuhan, M., *The Gutenberg galaxy*, London, Routledge & Kegan Paul, 1962.
- MACNAMARA (1996A)** Macnamara, D. S., Kintsch, W., Learning from Texts : Effects of Prior Knowledge and Text Coherence, *Discourse Processes*, 22, pp. 247-288, 1996.
- MACNAMARA (1996B)** Macnamara, D. S., Kintsch, E., Songer, N. B. & Kintsch, W., Are good Texts always better? Interactions of Text Coherence, Background Knowledge, and Level of Understanding in Learning from Text, *Cognition and Instruction*, 14 (1), pp. 1-43, 1996.
- MANN (1988)** Mann, W.C., Thompson, S.A., Rhetorical Structure Theory : Toward a functional theory of text organization. *Text*, 8(3), pp. 243-281, 1988.
- MARTIN (1994)** Martin, J., Cadre d'étude de la multimodalité sur les types et buts de coopération entre modalités, dans actes de *Interface to Real&Virtual Worlds*, EC2 Publ., pp. 97-106, 1994.
- MAUREL (2002A)** Maurel, F., Luc, C., Mojahid, M., Vigouroux, M., Virbel, J., Nespoulous, J.-L., De l'influence de la variabilité des textes sur leur transposition automatique à l'oral, Dans actes de *5e Colloque International sur le Document Électronique (CIDE 2002)*, INRIA, Rocquencourt, France, pp. 211-225, 2002.
- MAUREL (2002B)** Maurel, F., Luc, C., Mojahid, M., Vigouroux, N., Virbel, J., Nespoulous, J.-L., Transposition à l'oral des structures énumératives à partir de leurs paramètres formels, Dans actes de *Inscription Spatiale du Langage structure et processus (ISLsp 2002)*, IRIT, pp. 179-189, 2002.
- MAUREL (2002C)** Maurel, F., Luc, C., Mojahid, M., Vigouroux, N., Virbel, J., Nespoulous, J.-L., Problématique du traitement des structures visuelles dans la présentation oralisée des textes, Dans actes de *Documents Virtuels Personnalisables 2002 (DVP 2002)*, ENST Brest, pp. 11-24, 2002.

- MAUREL (2003A)** Maurel, F., Lemarié, J., Vigouroux, N., Évaluation cognitive d'une représentation du texte pour sa présentation multimodale, Dans actes de *Document Électronique Dynamique - Conférence Internationale du Document Électronique (CIDE 2003)*, Europia, pp. 25-41, 2003.
- MAUREL (2003B)** Maurel, F., Valoriser l'« architecture » des textes : - parce qu'il FAUT oraliser ; - par CE qu'il faut oraliser, Journée d'étude de l'ATALA : Le rôle de la typographie et de la ponctuation dans le traitement automatique des langues, Paris, 23 novembre 2003.
- MAUREL (2003C)** Maurel, F., Raynal, M., Oriola, B., Vigouroux, N., Extension of W3C Standards for Real Web Accessibility, Dans actes de *Universal Access in Human - Computer Interaction 2003 (UAHCI 2003)*, New Jersey - London , vol 4, pp. 995-999, 2003.
- MAUREL (2003D)** Maurel, F., Lemarié, J., Vigouroux, N., Oralisation de structures visuelles : de la lexico-syntaxe à la prosodie, Dans actes de *Interface Prosodique 2003 (IP2003)*, A. Mettouchi & G. Ferré (eds), pp. 137-142, 2003.
- MAUREL (2003E)** Maurel, F., Vigouroux, N., Raynal, M., Oriola, B., Contribution of the Transmodality Concept to Improve Web Accessibility, Dans M. Mokhtari (eds). *Assistive Technology Series Vol.12 (actes de ICOST'03 – Independent living for persons with disabilities and elderly people)*. Chapter 4 – computer accessibility, IOS Press, Amsterdam, pp. 186-193, 2003.
- MAUREL (2004A)** Maurel, F., De l'écrit à l'oral : analyses et générations. Dans actes de *11ème Conf. sur le Traitement Automatique des Langues Naturelles (TALN'2004)*, LPL-ENS/Fes-ATALA, ISBN 2-9518233-4-7, pp. 289-298, 2004.
- MAUREL (2004B)** Maurel, F., Lemarié, J., Vigouroux, N., Virbel, J., Mojahid, M., Nespoulous, J.-L., De l'adaptation de la présentation oralisée des textes aux difficultés perceptives et mnésiques du langage, *Revue Parole*, 2004.
- MERTENS (2001)** Mertens, P., Goldman, J.-P., Wehrli, E., Gaudinat, A., La synthèse de l'intonation à partir de structures syntaxiques riches, *Traitement automatique des*

- langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, pp. 145-192, 2001.
- MILIC-FRAYLING (2002)** Milic-Frayling, N., Sommerer, R., SmartView: Flexible viewing of web page contents, Dans actes de *11th International WWW Conference*, 2002.
- MITTERRAND (1985)** Mitterrand, H., *La notion de paragraphe*, Éditions du CNRS, 1985.
- MILLS (1987)** Mills, C.B., Weldon, L.J., Reading text from computer screens, *ACM Computing Surveys*, 19(4), pp.329-358, 1987.
- MOESCHLER (1994)** Moeschler, J., Reboul, A., *Dictionnaire Encyclopédique de Pragmatique*, Editions du seuil, 1994.
- MOJAHID (1990)** Mojahid, M., *Une architecture multi-agents en PAO pour le formatage de documents structurés*, Thèse de doctorat – université Paul Sabatier, 1990.
- MOREL (2001)** Morel, M., Lacheret-Dujour, A., « Kali », synthèse vocale à partir du texte, *Traitement automatique des langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, pp. 193-221, 2001.
- MOURAD (2001)** Mourad, G., *Analyse informatique des signes typographiques pour la segmentation de textes et l'extraction automatique de citations. Réalisation des Applications informatiques : SegATex et CitaRE*, Thèse de doctorat, Université Paris-Sorbonne, 2001.
- MOXON (1683)** Moxon, J., *Mechanick exercises on the whole art of printing*, edited by H Davis & H Carter, 2nd edition, London: Oxford University Press, 1962.
- MUTTER (1996)** Mutter, P., Interface design and optimization of reading of continuous text, *Cognitive Aspects of Electronic Text Processing*, pp.161-180, H. van Oostendorp and S. de Mul (Eds.), Ablex, 1987.
- NASH (1980)** Nash, W., *Designs in prose*, London, Longman, 1980.
- NESPOULOUS (1997)** Nespoulous, J.-L., Invariance et variabilité dans la symptomatologie linguistique des aphasiques agrammatiques : le retour du comparatisme ?, in C. Fuchs & S. Robert (Eds.), *Diversité des langues et représentations cognitives*, Paris, Ophrys, 1997.

- NESPOULOUS (2002)** Nespoulous, J.-L., De Mattos Pimenta, M. A., Virbel, J., Des objets textuels à leur traitement par le cerveau/esprit humain, in Virbel, J., Nespoulous, J.-L. (eds), *Inscription Spatiale du langage : structures et processus*, Prescot, Toulouse, pp. 77-87, 2002.
- NIGAY (1993)** Nigay, L., Coutaz, J., A Design Space for Multimodal System: Concurrent Processing and Data Fusion, Dans actes de *INTERCHI'93* , S. Ashlung, K. Mullet, A. Handerson, E. Hollnagel, T. White Eds., ACM New-York Publ., pp. 172-178, 1993.
- NIGAY (1994)** Nigay, L., *Conception et modélisation logicielles des systèmes interactifs*, Thèse de doctorat de l'Université J., Fourier, Grenoble, 1994.
- NIGAY (1996)** Nigay, L., Coutaz, J., Espaces conceptuels pour l'interaction multimédia et multimodale, *TSI, Spécial Multimédia et Collecticiel*, AFCET & Hermès, Publ., Vol. 15 (9), pp.1195-1225, 1996.
- NUNBERG (1990)** Nunberg, G., The Linguistics of Punctuation. *Number 18 in CSLI Lecture Notes*. CSLI Publications, Stanford, CA, 1990.
- NYSTRAND (1986)** Nystrand, M., Doyle, A., Himley, M., A critical examination of the doctrine of autonomous texts, in M Nystrand, *The structure of written communication*, Orlando, Florida: Academic Press, 81-107, 1986.
- OGILVY (1977)** Ogilvy, D., *Les confessions d'un publicitaire*. Dunod, 1977.
- ONG (1982)** Ong, J., *Orality and literacy*, London, Methuen, 1982.
- ÖQUIST (2003)** Öquist, G., Goldstein, M., Towards an improved readability on mobile devices: evaluating adaptative rapid serial visual presentation, *Interacting with Computers*, 15(4), pp. 539-558, 2003.
- ÖQUIST (2004)** Öquist, G., Sågval Hein, A., Ygge, J., Goldstein, M., Eye Movement Study of Reading on a Mobile Device Using the Page and RSVP Text Presentation Formats, *Mobile Human-Computer Interaction – MobileHCI 2004*, pp. 108-119, Springer-Verlag, Berlin, 2004.
- OVINK (1938)** Ovink, W., *Legibility, atmosphere-value, and forms of printing types*, Leyden, AW Sijthof, 1938.

- PASCUAL (1991)** Pascual, E., *Représentation de l'architecture textuelle et génération de texte*, Thèse de doctorat – Université Paul Sabatier, 1991.
- PASCUAL (1996)** Pascual, E., Integrating Text Formatting and Text Generation. In : Giovanni Adorni, Michael Zock (Eds.) *Trends in Natural Language Generation, An Artificial Intelligence Perspective*, 1993, Selected Papers. *Lecture Notes in Computer Science 1036*, Springer, 205-221, 1996.
- PASSERAULT (1991)** Passerault, J. M., Chesnet, D., Le marquage des paragraphes : son rôle dans la gestion des traitements pendant la lecture, *Psychologie Française*, 36(2), pp. 159-165, 1991.
- PEI (1954)** Pei, A., Gaynor, F., *A dictionary of linguistics*, New York, Philosophical Library, 1954.
- POWER (2003A)** Power, R., Scott, D., Bouayad-Agha, N., Document Structure, *Computational Linguistics* 29(2), pp. 211-260, 2003.
- POWER (2003B)** Power, R., Scott, D., Bouayad-Agha, N., Generating Texts with Style, *CICLing 2003*, pp. 444-452, 2003.
- RAHMAN (1999)** Rahman, T., Muter, P., Designing an interface to optimize reading with small display windows, *HUMAN FACTORS*, 41, No. 1, pp. 106-117, 1999.
- RICHAUDEAU (1984)** Richaudeau, F., *Recherches actuelles sur la lisibilité*, Retz, Paris, 1984
- RICHAUDEAU (1989)** Richaudeau, F., *Manuel de typographie et de mise en page*, Retz, Paris, 1989.
- ROSSI (1999)** Rossi, M., *L'intonation, le système du français : description et modélisation*, Paris, Ophrys, 1999.
- RUSSELL (2002)** Russell, M., James, M., Cohlma, A., Reading from a Palm Pilot™ Using RSVP, *Usability News 4.1*. [Online]. http://psychology.wichita.edu/surl/usabilitynews/41/rsvp_palm.htm
- SAENGER (1982)** Saenger, P., *Silent reading: its impact on late medieval script and society*, *Viator*, 13, 367-414, 1982.

- SAUSSURE (1964)** Saussure, F. de, *Cours de linguistique générale* (1915), Paris, Payot.
- SCHMID (2001A)** Schmid, S., Baccino, T., Stratégies de lecture dans les textes à consignes, *Langages*, 141, pp. 105-124, 2001.
- SCHMID (2001B)** Schmid, S., *Représentation organisationnelle et compréhension : rôle de la mise en forme matérielle dans la lecture*, Thèse de doctorat de l'Université de Nice Sophia Antipolis, 2001.
- SCOTT (2003)** *Layout in NLP: the case for Document Structure*, Invited presentation to ACL'03 Sapporo - <http://www.itri.brighton.ac.uk/~Donia.Scott/slides.htm> 1, Japan, July 2003.
- SHEBILSKE (1981)** Shebilske, L., Rotondo, A., Typographical and spatial cues that facilitate learning from textbooks, *Visible Language*, 15, pp. 41-54, 1981.
- SPENCE (2001)** Spence, R., *Information Visualization*, ACM Press, New-York, 2001.
- SPENCER (1975)** Spencer, H., Reynolds, L., Coe, B., Spatial and typographic coding with bibliographic entries, *Programmed Learning & Educational Technology*, 12, pp. 95-101, 1975.
- SPROAT (1998)** Sproat, R., *Multilingual Text-To-Speech Synthesis : the Bell Labs Approach*, Kluwer Academic Publishers, 1998.
- SULLET-NYLANDER (2001)** Sullet-Nylander, F., Défigements, noms propres et titres de presse, in Kronning, H., Noreen, C., Noven, B., Ransbo, G., Sundell, L.-G., Svane, B., (éds.) : *Langage et référence*, Mélanges offerts à Kerstin Jonasson à l'occasion de ses soixante ans, Uppsala, Acta Universitatis Upsaliensis : 619-627., 2001.
- TERRIER (2000)** Terrier, P., Aissani, A., Amiel, V., Cellier, J.-M. (2000). Rôle de la mise en forme dispositionnelle dans le traitement cognitif d'un texte narratif : analyse en rappel différé, *International Journal of Psychology*, 35 (3/4), pp. 58-59, Stockholm, 2000..
- THEVENIN (2001)** Thevenin, D., *Adaptation en Interaction Homme-Machine : le cas de la Plasticité*, Thèse de doctorat, Université Joseph Fourier, Grenoble 1, 2001.

- TINKER (1955)** Tinker, M.A., Perceptual and oculomotor efficiency in reading materials in vertical and horizontal arrangements, *American Journal of Psychology*, 68, pp. 444-449, 1955.
- TINKER (1963)** Tinker, M.A., *Legibility of Print*, Ames, Iowa State University Press, 1963.
- TINKER (1965).** Tinker, M.A., *Bases for effective reading*, Minneapolis, University of Minnesota Press, 1965.
- TORZEC (2001)** Torzec, N., Moudenc, T., Emerard, F., Prétraitement et analyse linguistique dans le système de synthèse TTS CVOX, *Traitement automatique des langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, pp. 17-46, 2001.
- TRUILLET (1999)** Truillet, Ph., *Modélisation de coopérations intermodales : application à l'interaction non visuelle*, Thèse de doctorat, Université Paul Sabatier, Toulouse III, 1999.
- TWYMAN (1979)** Twyman, M.L., A schema for the study of graphic language, in P Kolars, M Wrolstad, H Bouma (eds) *Processing of visible language, vol 1*, New York, Plenum, pp. 117-150, 1979.
- TWYMAN (1982)** Twyman, M.L., The graphic presentation of language, *Information Design Journal*, 3, pp. 2-22, 1982.
- TZOUKERMANN (2001)** Tzoukermann, E., La synthèse du français parlé : le système des laboratoires Bell, *Traitement automatique des langues ; Vol. 42(1) - Synthèse de la parole à partir du texte*, pp. 69-114, 2001.
- VACHEK (1948/1967)** Vachek, J., Written language and printed language, in *A Prague School reader in linguistics*, Bloomington, Indiana University Press, pp. 453-460, 1967 (originally published in *Recueil Linguistique de Bratislavia, I*, pp. 67-75, 1948)
- VACHEK (1959)** Vachek, J., Two chapters on written English, *Brno Studies in English, 1*, pp. 7-38, 1959.
- VACHEK (1973)** Vachek, J., *Written language: general problems and problems of English*, The Hague, Mouton, 1973.
- VACHEK (1979)** Vachek, J., Some remarks on the stylistics of written language, in DJ Alerton, E Carney & D Holdcroft

- (eds), *Function and context in linguistic analysis*, Cambridge University Press, pp. 206-221, 1979.
- VAN DIJK (1977)** Van Dijk, T., Semantic Macro-structures and Knowledge Frames in Discourse Comprehension, in JUST, M. A., Carpenter, P. A. (eds), *Cognitive Processes in Comprehension*, Lawrence Erlbaum, Hillsdale, pp. 3-32, 1977.
- VEDENINA (1989)** Védénina, L.G., *Pertinence linguistique de la typographie*, Peeters/Selaf, Paris, 1989.
- VIGOUROUX (2002)** Vigouroux, N., Maurel, F., Oriola, O., Truillet, P., Privat, R., Ducos, D., Virbel, V., Luc, C., Mojahid, M., Nespoulous, J.-L., Toën, J., Problématique, enjeux et perspectives de la présentation orale de documents : une approche pluridisciplinaire, Dans actes de *Inscription Spatiale du Langage structure et processus (ISLsp)*, IRIT , pp. 139-150, 2002.
- VIRBEL (1987)** Virbel, J., L'apport des connaissances linguistiques à l'interprétation des structures textuelles, *Bigre+Globule*, 53, pp. 77-97, 1987.
- VIRBEL (1989)** Virbel J., The Contribution of linguistic knowledge to the interpretation of text Structure, Dans J. André, V. Quint et R. Furuta (eds.), *Structured Documents*, Cambridge University Press, pp. 161-181, 1989.
- VIRBEL (2002)** Virbel J., Éléments d'analyse du titre, Dans *Inscription Spatiale du langage : structures et processus*, 123-134, 2002.
- WALLER (1987)** Waller, R., *The Typographic Contribution to Language*. Unpublished PhD Dissertation, Department of Typography and Graphic Communication, University of Reading, 1987.
- WARDE (1955)** Warde, B.L., *The crystal goblet: sixteen essays on typography*, London, Sylvan Press, 1955.
- WOBROCK (2002)** Wobbrock, J.O., Forlizzi, J., Hudson, S.E., Myers, B.A., WebThumb: Interaction Techniques for Small-Screen Devices. Dans actes de *ACM Symposium on User Interface Software and Technology*, pp. 205-208, Paris, France, 2002.
- WRIGHT (1978)** Wright, P., Feeding the information eaters: suggestions for integrating pure and applied research

in language comprehension, *Instructional Science*, 7,
pp. 249-312, 1978.

ZACHRISSON (1965)

Zachrisson, B. *Studies in the legibility of text*,
Stockholm, Almqvist & Wiksel, 1965.

Table des figures

FIGURE 1 : TRANSMODALITE (ECRIT→ORAL) POUR UNE MULTIMODALITE UTILE ET EFFICACE	18
FIGURE 2 : ARCHITECTURE « CLASSIQUE » D'UN SYSTEME DE TEXT-TO-SPEECH.....	33
FIGURE 3 : MODELE DU GENRE TEXTUEL (WALLER, 1987).....	66
FIGURE 4 : STRUCTURE INDENTEE D'UN DOCUMENT	72
FIGURE 5 : REPRESENTATION D'UN TEXTE PAR UNE IDP	77
FIGURE 6 : ÉNUMERATION EXTRAITE D'UNE BROCHURE DE MAISON DE RETRAITE	81
FIGURE 7 : TYPOLOGIE DES STRUCTURES ENUMERATIVES ET DE LEURS MARQUES	84
FIGURE 8 : SYSTEME SYNOPTIQUE D'ORALISATION DES STRUCTURES VISUELLES DES TEXTES	95
FIGURE 9 : EXEMPLE DE TEXTE MIS EN PAGE, INSPIRE DE DUPLAN (1986 :127).....	96
FIGURE 10 : TYPES FINAUX DES OT ISSUS DES PHASES DE SEGMENTATION/EXTRACTION.....	98
FIGURE 11 : MORTELS	102
FIGURE 12 : CHANGEMENTS DE REGISTRE A CHAQUE DECLINAISON INTONATIVE	130
FIGURE 13 : IMAGE DE PAGE DE LA STRUCTURE VISUELLE DU TEXTE SOURCE SYNTHETISE	139
FIGURE 14 : EXEMPLE DE FICHIER PROSODIQUE GENERE A PARTIR DE « MAX EST VENU. »	140
FIGURE 15 : D'UNE SE TOUT-VENANT A SA FDI ET A SA FPC : PROCESSUS ET RESSOURCES	144
FIGURE 16 : SCHEMA RECAPITULATIF DES DIFFERENTS NIVEAUX D'ANALYSE	158
FIGURE 17 ; IMAGES DE PAGE SE, BLOC, AUTRE.....	184
FIGURE 18 : SYNTHESE DE STRATEGIES DE PRESENTATION SUR PETIT ECRAN	194
FIGURE 19 : TROIS PRESENTATIONS D'UNE PAGE WEB	197

Table des tableaux

TABLEAU 1: PROPRIETES CARE	17
TABLEAU 2 : LINEARITE ET REGLES DE LECTURE SELON TWYMAN ET WALLER	26
TABLEAU 3 : LES 6 STRUCTURES TYPOGRAPHIQUES SELON RICHAUDEAU (1989)	27
TABLEAU 4 : VARIATIONS DES PARAMETRES FORMELS DES STRUCTURES ENUMERATIVES.....	83
TABLEAU 5 : DE LA FDI A LA FPC.....	135
TABLEAU 6 : EFFECTIFS POUR CHAQUE COUPLE (TYPE D'IMAGE DE PAGE ; POSITION)	185
TABLEAU 7 : EFFECTIFS POUR CHAQUE ORDRE SELON LA FORME D'ORALISATION.....	186

Table des matières

PROBLEMATIQUE ET MISE EN PERSPECTIVE.....	11
TRANSMODALITE ET MULTIMODALITE	13
LA LECTURE SILENCIEUSE.....	21
RELATIONS ENTRE L'ECRIT ET L'ORAL.....	23
PARTIE A VALEUR ET PRISE EN COMPTE DE LA MORPHO-DISPOSITION	29
INTRODUCTION	31
CHAPITRE I. SYNTHESE DE PAROLE ET TYPOGRAPHIE	33
CHAPITRE II. DU POINT DE VUE DES PSYCHOLINGUISTES.....	39
II.1 <i>La Gestalt-Théorie</i>	39
II.2 <i>Psychologie appliquée et typographie</i>	40
II.2.1 Rôle de la lisibilité comme qualité « physique » du texte	40
II.2.2 Les propriétés expressives de la typographie	42
II.2.3 Influence de la typographie sur le traitement cognitif	44
II.2.3.1 La place de la mise en forme des textes sur le plan empirique.....	45
II.2.3.2 Typo-disposition et modèle propositionnel de la compréhension du texte.....	49
CHAPITRE III. DU POINT DE VUE DES LINGUISTES	53
III.1 <i>Positionnement terminologique et taxonomie</i>	53
III.2 <i>Justifications préliminaires</i>	56
III.3 <i>Les précurseurs</i>	63
III.4 <i>L'approche de Nina Catach</i>	67
III.5 <i>Les travaux de Nunberg, Power et Scott</i>	70
III.6 <i>Le Modèle d'Architecture Textuelle</i>	73
III.7 <i>Observation de 3 Objets Textuels</i>	78
III.7.1 Le paragraphe	78
III.7.2 Les structures énumératives	80
III.7.3 Le titre.....	85
CONCLUSION	87
PARTIE B ARCHITECTURE, MODELE ET OPERATIONNALISATION	93
CHAPITRE IV. UNE ARCHITECTURE ADAPTEE A L'ORALISATION DES TEXTES PAR « OT »	95
CHAPITRE V. MODELE DES STRATEGIES D'ORALISATION	101
CHAPITRE VI. OPERATIONNALISATION DES STRATEGIES	105
VI.1 <i>Les structures énumératives (SE)</i>	106
VI.1.1 Type d'inscription graphologique et situation d'énonciation	106
VI.1.2 Le langage de représentation.....	107
VI.1.3 De l'OT normalisé et ses paramètres formels à la FDI	109
VI.1.4 De la FDI à la forme prosodique en contexte (FPC).....	115
VI.1.4.1 De la difficulté d'utiliser une méthode par analyse de corpus	115
VI.1.4.2 Recherche d'un modèle prosodique adapté	118
VI.1.4.3 Intonation et constructions discursives	119
VI.1.4.4 Projection des intonèmes de Rossi aux structures énumératives	124
VI.1.4.5 Le langage de représentation	132
VI.1.4.6 Intégration de nos spécifications prosodiques dans un système de TTS.....	136

VI.2	<i>Oralisation d'autres Objets Textuels</i>	145
PARTIE C	ÉVALUATION COGNITIVE	149
CHAPITRE VII.	HYPOTHESES ET VARIABLES	151
VII.1	<i>Variable indépendante</i>	151
VII.1.1	F : forme d'oralisation du texte	151
VII.1.1.1	Deux hypothèses spécifiques	151
VII.1.1.2	Conception de la forme contrôle	152
VII.1.2	A : âge des sujets	153
VII.1.3	T : tâche demandée	153
VII.1.4	Statuts des variables indépendantes	153
VII.2	<i>Variables de contrôle</i>	154
VII.3	<i>Variables dépendantes</i>	155
VII.3.1	Quantité d'informations justes rappelées et du degré de littéralité du rappel	157
VII.3.1.1	Choix du référent	157
VII.3.1.2	Niveaux d'analyse des données	157
VII.3.2	Restitution des informations architecturales	159
VII.3.3	Modalité de restitution du nombre d'items :	159
VII.3.4	Mesure du respect de l'ordre	159
VII.3.4.1	Codage de l'ordre lors du processus de rappel	160
VII.3.4.2	Codage de l'ordre du rappel final	160
CHAPITRE VIII.	<i>DESIGN FINAL</i>	161
VIII.1	<i>Constitution des stimuli</i>	161
VIII.2	<i>Procédure de présentation</i>	164
CHAPITRE IX.	ANALYSE DES PREMIERS RESULTATS	167
IX.1	<i>Variables contrôles en fonction de la forme d'oralisation</i>	167
IX.1.1	Effectif des sujets en fonction du sexe	167
IX.1.2	Age des sujets selon les groupes expérimentaux	167
IX.1.3	Niveau d'étude en fonction des groupes expérimentaux	167
IX.1.4	Activité impliquant l'utilisation de la mémoire	168
IX.1.5	Familiarité des sujets avec le thème abordé dans les SE : l'APA	168
IX.2	<i>Analyse d'une variable dépendante</i>	169
IX.2.1	Pourcentage de MP rappelés par rapport aux MP présentés	169
IX.2.2	Rappel faux, absent ou juste en fonction de la forme d'oralisation :	170
IX.2.3	Pourcentage de MP justes par rapport à l'ensemble des MP rappelés :	172
PARTIE D	IMAGES DE PAGE ET MULTIMODALITE	177
CHAPITRE X.	<i>FIRST GLANCE</i> ET IMAGES DE PAGE	179
X.1	<i>Principes fondateurs</i>	180
X.2	<i>Évaluation de la plausibilité cognitive des IDP</i>	181
X.2.1	Premier protocole	182
X.2.1.1	Hypothèses, méthode et prédictions	182
X.2.1.2	Résultats	185
X.2.1.3	Discussion des résultats	186
X.2.2	Second protocole	188
X.2.2.1	Hypothèses, méthode et prédictions	188
X.2.2.2	Résultats	191
CHAPITRE XI.	PERSPECTIVES A COURT TERME	193

<i>XI.1</i>	<i>IDP, PDA et malvoyance</i>	193
<i>XI.2</i>	<i>IDP, navigation et non-voyance</i>	198
POINT SUR NOTRE CONTRIBUTION ET PERSPECTIVES A MOYEN TERME		201
(1)	IHM : VERS UNE PRESENTATION ADAPTABLE	205
	Au niveau du choix de la modalité de présentation	205
	Au niveau des stratégies d’oralisation utilisées	205
	Au niveau de la combinaison des modalités	206
(2)	VERS UN SYSTEME DE « TRANSMODALISATION » DE L’ARCHITECTURE TEXTUELLE	206
(3)	COGNITION	207
TABLE DES FIGURES		229
TABLE DES TABLEAUX		229
TABLE DES MATIERES		231
ANNEXE		235
1.	REPRESENTATION DES STRUCTURES ENUMERATIVES	237
2.	STRATEGIES D’ORALISATION DES STRUCTURES ENUMERATIVES	239
2.1	<i>Stratégie discursive interprétative</i>	239
2.2	<i>Stratégie prosodique en contexte</i>	243
TITRE ET RESUME EN ANGLAIS		247

Annexe

1. Représentation des structures énumératives

Nous proposons une représentation des structures énumératives sous un format de type *eXtended Markup Language* (XML). Ci-après est décrit le *XML Schema Definition* (XSD) qui permet de valider la conformité à notre format d'une représentation de structure énumérative.

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="texte" type="TexteData"/>
  <xsd:complexType name="TexteData">
    <xsd:sequence minOccurs="1" maxOccurs="unbounded">
      <xsd:choice>
        <xsd:element name="se" type="SeData"/>
      </xsd:choice>
    </xsd:sequence>
  </xsd:complexType>
  <xsd:complexType name="SeData">
    <xsd:sequence>
      <xsd:element name="amorces" type="AmorceData"/>
      <xsd:element name="enumeration" type="EnumerationData"/>
    </xsd:sequence>
    <xsd:attribute name="type" type="parallelisme"/>
    <xsd:attribute name="exhaus" type="reponse"/>
    <xsd:attribute name="ordon" type="reponse"/>
  </xsd:complexType>
  <xsd:simpleType name="parallelisme">
    <xsd:restriction base="xsd:string">
      <xsd:enumeration value="parallele"/>
      <xsd:enumeration value="non parallele"/>
    </xsd:restriction>
  </xsd:simpleType>
</xsd:schema>
```

```

        </xsd:restriction>
</xsd:simpleType>
<xsd:simpleType name="reponse">
    <xsd:restriction base="xsd:string">
        <xsd:enumeration value="oui"/>
        <xsd:enumeration value="non"/>
        <xsd:enumeration value="neutre"/>
    </xsd:restriction>
</xsd:simpleType>
<xsd:simpleType name="genre">
    <xsd:restriction base="xsd:string">
        <xsd:enumeration value="fem"/>
        <xsd:enumeration value="mas"/>
    </xsd:restriction>
</xsd:simpleType>
<xsd:simpleType name="catsynt">
    <xsd:restriction base="xsd:string">
        <xsd:enumeration value="nom_sing"/>
        <xsd:enumeration value="nom_plur"/>
        <xsd:enumeration value="verbe"/>
    </xsd:restriction>
</xsd:simpleType>
<xsd:complexType name="AmorceData">
    <xsd:simpleContent>
        <xsd:extension base="xsd:string">
            <xsd:attribute name="genre_classifieur" type="genre"/>
            <xsd:attribute name="nom_classifieur"
                type="xsd:string"/>
        </xsd:extension>
    </xsd:simpleContent>
</xsd:complexType>

```

```

        </xsd:extension>
    </xsd:simpleContent>
</xsd:complexType>
<xsd:complexType name="EnumerationData">
    <xsd:sequence>
        <xsd:element name="item" type="ItemData" minOccurs="2"
            maxOccurs="unbounded"/>
    </xsd:sequence>
</xsd:complexType>
<xsd:complexType name="ItemData" mixed="true">
    <xsd:sequence>
        <xsd:element name="noyau" type="xsd:string"/>
    </xsd:sequence>
    <xsd:attribute name="cs" type="catsynt"/>
</xsd:complexType>
</xsd:schema>

```

2. Stratégies d'oralisation des structures énumératives

L'automatisation des stratégies d'oralisation est abordée en associant à chacune d'elles une référence pour l'utilisation d'une feuille de transformation appropriée écrite en *eXtended Stylesheet Language Transformation* (XSLT). Sont proposées ci-après les deux feuilles XSLT associées à la forme discursive interprétative et la forme prosodique en contexte des structures énumératives parallèles.

2.1 Stratégie discursive interprétative

Le format de représentation cible de la transformation est, dans ce cas, du texte brut composé de phrases ponctuées interprétables par les systèmes de synthèse de la parole à partir de texte. Si une représentation source est validée par le XSD présentée en 1. et

que celle-ci fait référence à la feuille XSLT décrite ci-après, alors le texte sera généré selon une des neuf dérivations du *pattern* discursif proposé.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<xsl:transform version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:variable name="exhaustivite">
    <xsl:value-of select="texte/se/@exhaus"/>
  </xsl:variable>
  <xsl:variable name="ordonnancement">
    <xsl:value-of select="texte/se/@ordon"/>
  </xsl:variable>
  <xsl:variable name="genreclassifieur">
    <xsl:value-of select="texte/se/amorce/@genre_classifieur"/>
  </xsl:variable>
  <xsl:variable name="nomclassifieur">
    <xsl:value-of select="texte/se/amorce/@nom_classifieur"/>
  </xsl:variable>
  <xsl:template match="/">
    <xsl:if test="$exhaustivite = 'oui'">
      <xsl:text> Les </xsl:text>
    </xsl:if>
    <xsl:value-of select="count(//item)"/>
    <xsl:value-of select="texte/se/amorce"/>
    <xsl:if test="$exhaustivite = 'non'">
      <xsl:text> parmi d'autres </xsl:text>
    </xsl:if>
    <xsl:if test="$genreclassifieur = 'fem'">
      <xsl:text> sont énumérées ci-après</xsl:text>
    </xsl:if>
  </xsl:template>
</xsl:transform>
```

```

</xsl:if>
<xsl:if test="$genreclassifieur = 'mas'">
    <xsl:text> sont enumeres ci-apres</xsl:text>
</xsl:if>
<xsl:if test="$sordonnement = 'oui'">
    <xsl:text> dans l'ordre. </xsl:text>
</xsl:if>
<xsl:if test="$sordonnement = 'non'">
    <xsl:text> dans le désordre. </xsl:text>
</xsl:if>
<xsl:if test="$sordonnement = 'neutre'">
    <xsl:text>. </xsl:text>
</xsl:if>
<xsl:for-each select="//item">
    <xsl:variable name="catsynt">
        <xsl:value-of select="./@cs"/>
    </xsl:variable>
    <xsl:choose>
        <xsl:when test="position() = 1">
            <xsl:if test="$genreclassifieur = 'fem' and
                $sordonnement = 'oui'">
                <xsl:text> La 1ère </xsl:text>
            </xsl:if>
            <xsl:if test="$genreclassifieur = 'mas' and
                $sordonnement = 'oui'">
                <xsl:text> Le 1er </xsl:text>
            </xsl:if>
            <xsl:if test="$genreclassifieur = 'fem' and
                $sordonnement != 'oui'">

```

```

        <xsl:text> Une 1ère </xsl:text>
    </xsl:if>
    <xsl:if test="$genreclassifieur = 'mas' and
    $ordonnement != 'oui'">
        <xsl:text> Un 1er </xsl:text>
    </xsl:if>
</xsl:when>
<xsl:otherwise>
    <xsl:if test="$genreclassifieur = 'fem' and
    $ordonnement = 'oui'">
        <xsl:text> La </xsl:text>
    </xsl:if>
    <xsl:if test="$genreclassifieur = 'mas' and
    $ordonnement = 'oui'">
        <xsl:text> Le </xsl:text>
    </xsl:if>
    <xsl:if test="$genreclassifieur = 'fem' and
    $ordonnement != 'oui'">
        <xsl:text> Une </xsl:text>
    </xsl:if>
    <xsl:if test="$genreclassifieur = 'mas'and
    $ordonnement != 'oui'">
        <xsl:text> Un </xsl:text>
    </xsl:if>
    <xsl:value-of select="position()"/>
    <xsl:text>ème </xsl:text>
</xsl:otherwise>
</xsl:choose>
<xsl:if test="position()=last() and $genreclassifieur='fem'">

```

```

        <xsl:text> et dernière </xsl:text>
    </xsl:if>
    <xsl:if test="position()=last() and $genreclassifieur='mas'">
        <xsl:text> et dernier </xsl:text>
    </xsl:if>
    <xsl:value-of select="$nomclassifieur"/>
    <xsl:if test="$catsynt = 'nom_sing'">
        <xsl:text> est </xsl:text>
    </xsl:if>
    <xsl:if test="$catsynt = 'nom_plur'">
        <xsl:text> comprend </xsl:text>
    </xsl:if>
    <xsl:if test="$catsynt = 'verbe'">
        <xsl:text> est de </xsl:text>
    </xsl:if>
    <xsl:value-of select="."/>
    <xsl:text>. </xsl:text>
</xsl:for-each>
</xsl:template>
</xsl:transform>

```

2.2 Stratégie prosodique en contexte

Le format de représentation cible de la transformation est, dans ce cas, une nouvelle représentation au format défini par le langage *Speech Synthesize Markup Language* (SSML) à l'exception de la balise FOCUS que nous proposons de rajouter. Si une représentation source est validée par le XSD présentée en 1. et que celle-ci fait référence à la feuille XSLT décrite ci-après, alors le texte sera généré de manière à respecter le format SSML proposé et interprétable par les systèmes de synthèse de la parole à partir de texte.

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<xsl:transform version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:variable name="exhaustivite">
    <xsl:value-of select="texte/se/@exhaus"/>
  </xsl:variable>
  <xsl:variable name="ordonancement">
    <xsl:value-of select="texte/se/@ordon"/>
  </xsl:variable>
  <xsl:template match="/">
    <xsl:element name="se">
      <xsl:if test="$exhaustivite = 'oui'">
        <xsl:text> Les </xsl:text>
      </xsl:if>
      <xsl:element name="focus">
        <xsl:attribute name="type">
          <xsl:text>etendu</xsl:text>
        </xsl:attribute>
        <xsl:value-of select="count(//item)"/>
        <xsl:value-of select="texte/se/amorce"/>
      </xsl:element>
      <xsl:if test="$exhaustivite = 'non'">
        <xsl:text>, parmi d'autres, </xsl:text>
      </xsl:if>
      <xsl:text> sont </xsl:text>
      <xsl:if test="$ordonancement = 'oui'">
        <xsl:text> dans l'ordre </xsl:text>
      </xsl:if>
      <xsl:if test="$ordonancement = 'non'">

```

```

        <xsl:text> dans le désordre </xsl:text>
</xsl:if>
<xsl:element name="break">
    <xsl:attribute name="size">
        <xsl:text>large</xsl:text>
    </xsl:attribute>
</xsl:element>
<xsl:for-each select="texte/se/enumeration/item">
    <xsl:variable name="arbreitem">
        <xsl:value-of select="."/>
    </xsl:variable>
    <xsl:variable name="arbrenoyau">
        <xsl:value-of select="./noyau"/>
    </xsl:variable>
    <xsl:variable name="pitch">
        <xsl:value-of select="4*((count(//item)-position()) div
(count(//item)-1))"/>
    </xsl:variable>
    <xsl:element name="prosody">
        <xsl:if test="position()=last()">
            <xsl:attribute name="pitch">
                <xsl:value-of select="$pitch"/>
            </xsl:attribute>
        </xsl:if>
        <xsl:if test="position()=last()">
            <xsl:attribute name="range">
                <xsl:text>+80Hz</xsl:text>
            </xsl:attribute>

```

```

</xsl:if>
<xsl:element name="focus">
  <xsl:attribute name="type">
    <xsl:text>rhematique</xsl:text>
  </xsl:attribute>
  <xsl:value-of select="$arbrenoyau"/>
</xsl:element>
<xsl:value-of select="substring-
after($arbreitem,$arbrenoyau)"/>
</xsl:element>
<xsl:if test="position()=last()">
  <xsl:element name="break">
    <xsl:attribute name="size">
      <xsl:text>medium</xsl:text>
    </xsl:attribute>
  </xsl:element>
</xsl:if>
</xsl:for-each>
</xsl:element>
</xsl:template>
</xsl:transform>

```

Titre et résumé en anglais

Written/oral transmodality and multimodality:
*modeling, automatic treatment and evaluation of strategies
for the presentation of texts “visuo-architectural” structures*

Considered for a long time as ornamental, the informational structure of written documents carried by texts morpho-disposition becomes a full object of investigation in various disciplines such as linguistic, psycholinguistic or computer sciences. In Particular, we are interested in the utility of these aspects of documents and, if the need arises, their usability, within the framework of their oral transposition. In the objective to provide solutions which make it possible to react effectively to this “morphological inscription”, we propose the synoptic of an oralisation system for “ordinary documents” who leads to a text representation directly interpretable by Text-To-Speech systems. We modelled and partially realized the module specific to the oralisation strategies, in order to render some signifying parts of the text often “forgotten” by synthesis systems. The first results of this study led to specifications in the course of integration by an industrial partner. Predictive hypothesis, related to the impact on memorizing/understanding of two strategies (discursive and prosodic) coming from our Reformulation-based Oralisation Model for Texts Written to be Silently Read (MORTELS), have been formulated and tested. This work shows that cognitive functions was lost. Prototypes, exploiting the “Page Reflection” notion (IDP), have been conceived through interfaces in which multimodality is used to fill this gaps.